Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 2880–2884, May 1985
Genetics

Propagation of some human DNA sequences in bacteriophage λ vectors requires mutant *Escherichia coli* hosts

(recombinant DNA libraries/inverted repetition/*recB recC sbcB* genes/DNA heteroduplex)

ARLENE R. WYMAN, LINDA B. WOLFE, AND DAVID BOTSTEIN

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by David Botstein, December 31, 1984

ABSTRACT **The growth of clones of human genomic DNA fragments in a bacteriophage λ vector has been examined in a number of different *Escherichia coli* hosts. A large proportion (8.9%) of the phages carrying different fragments of the human genome fail to grow on standard *rec*$^+$ hosts but will grow on hosts carrying mutations in the *recB*, *recC*, and *sbcB* genes. Heteroduplex analysis in the electron microscope of DNA from four of these phages revealed substantial secondary structure, including snap-back regions 200–500 base pairs in length. Such structures were not found in phages from the same DNA library that grow in *rec*$^+$ hosts. These results are interpreted in the light of prior observations [Leach, D. R. F. & Stahl, F. (1983) *Nature (London)* 305, 448–451] showing that inverted repetitions cloned in phage λ can be propagated in *recB recC sbcB* hosts but not in *rec*$^+$ hosts.**

Recombinant DNA libraries in bacteriophage λ vectors are generally found to contain most of the sequences of the genomes from which they are derived (1, 2). Nevertheless, anecdotal reports of sequences that cannot be found in libraries of eukaryotic genomes are fairly common. One striking example is the hypervariable region 3′ to the α-globin locus, which has resisted strenuous efforts at cloning (S. Goodbourn and S. Orkin, personal communications). Another example is the junction fragment from a translocation between chromosomes 6 and 10 in a murine plasmacytoma (3). A related phenomenon is the observation of discrepancies between cloned and genomic sequences that are the result of deletions occurring during cloning of mammalian DNA (3–5). Only some of these deletions can be prevented by growth on a *recA* host.

For the past few years we have been studying a highly polymorphic locus in humans, now known as D14S1, which was identified by hybridization of gel transfers of genomic DNA using an adjacent single-copy fragment as probe (6–8). Efforts to obtain clones of the sequences surrounding and including the polymorphic site itself from the Maniatis human genomic library (2) and from libraries made in our laboratory, using conventional cloning conditions as well as hosts carrying mutations in the *recA* gene, were all unsuccessful. Given the rarity in general of highly polymorphic DNA sequences in the human genome and the striking difficulty in recovering the variable segments in two of the most studied ones (α globin and D14S1), we became concerned that there might be a systematic difficulty in cloning certain human DNA sequences. Further, we speculated that the difficulty might sometimes be associated with a high degree of polymorphism.

Recently, Leach and Stahl (9) reported that inverted repetitions cannot be cloned into phage λ vectors unless mutant hosts are used. We report here that such a host, which contains mutant *recB*, *recC*, and *sbcB* genes, will propagate many phage λ clones containing random human DNA fragments, including D14S1, that cannot be grown on the *Escherichia coli* hosts commonly used in the selection and propagation of recombinant DNA genome libraries.

MATERIALS AND METHODS

Bacterial Strains. With the exception of strains BNN45 and SF8, all strains used are derivatives of the *E. coli* K-12 strain AB1157; they have the following markers in common: *leu*6 *ara*14 *his*4 *thr*1 *thi*1 *lacY*1 *mtl*1 *xyl*5 *galK*2 *proA*2 *argE*3 *str*31 *tsx*33 *sup*37-amber (10). The additional mutations in each are as follows: JC5519, *recB*21 *recC*22; MO627, *sbcB*15; JC7623, *recB*21 *recC*22 *sbcB*15; MO611, *end recB*21 *recC*22 *sbcB*15 (by mutation of JC7623); and DB1170, *recB*21 *recC*22 *sbcB*15 *end hsdR* (by introduction into MO611 of the *hsdR* mutation in this laboratory). The genotype of BNN45 is *hsd*R *supE*44 *supF thi met* (11) and of SF8 is *recB recC thr*1 *leuB*6 *thi*1 *supE*44 *hsdR hsdM lop*11 *gal*96 Sm^R (11).

Library of Human DNA. Human DNA was digested to completion with *Eco*RI and *Bam*HI (New England Biolabs) and was size-fractionated on KOAc gradients (12). Fragments of ≈13–15 or 15–19 kilobases were ligated into the purified, hybridized arms of the phage λ vector Charon 30 (13) digested with *Eco*RI and *Bam*HI. Recombinant DNA was packaged into bacteriophage heads by using Packagene extract (Promega Biotec, Madison, WI).

Conditions for Plating and Propagation Tests. Plating bacteria were prepared as described (12). For spot tests to examine the plating properties of phages from the unamplified library, packaged phage were plated on DB1170, and phage from individual plaques were transferred by toothpick to fresh lawns of BNN45 and DB1170.

For efficiency of plating measurements, phage were plated by preadsorbing them for 10 min at 37°C to 0.1 ml of plating bacteria, adding 2.5 ml of phage λ medium (11) containing 0.7% agar, pouring onto plates containing LB agar (11) and, incubating overnight at 37°C.

Burst size was determined by infecting a midlogarithmic culture at a multiplicity of infection of 0.01 phage per cell, preadsorbing for 10 min at 37°C, diluting, and incubating the infected cells at 37°C. At 20-min intervals, aliquots were treated with $CHCl_3$, and the number of phage was determined. Little difference was noted between 60 and 80 min; therefore, burst size is given as the ratio of phage produced by the end of 60 min to the number of input phage.

Charon 30::Tn*10*Δ Construction. The tetracycline-resistance insertion Tn*10*Δ*16*Δ*17* was transduced into a Charon 30 clone containing a human DNA insert by growth on a host containing the deleted transposon and a plasmid that provides transposition functions as described (14). One derivative that could transduce tetracycline resistance and contained an insertion in the left arm of the phage λ genome, as judged by DNA heteroduplex analysis with vector DNA,

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.

Table 1. Efficiency of plating of Charon 30 phages containing human genome DNA inserts

Phage	DB1170 *recB recC sbcB*	BNN45 *rec*$^+$
684-1	(1.0)	0.48*
684-2	(1.0)	<0.004
684-3	(1.0)	<0.003
684-4	(1.0)	<0.0005
F1	(1.0)	2.9
F2	(1.0)	1.8
F3	(1.0)	3.3
F4	(1.0)	1.1
F5	(1.0)	1.4
F6	(1.0)	1.3

For each phage, the efficiency of plating was normalized to the titer on DB1170.
*These plaques were tiny, less than 0.1-mm diameter.

was crossed *in vivo* with Charon 30 (11). Recombinants that had obtained the Tn*10*Δ were selected by tetracycline-resistance; those that had retained the central fragment from the vector (which carries the *red* and *gam* genes) were then selected by their ability to grow on a *recA* host (15). The structure of the recombinant (Charon30::Tn*10*Δ) was checked by restriction enzyme mapping, and the location of the Tn*10*Δ insertion was confirmed by electron microscopy.

Electron Microscopy. Phage were grown as plate lysates on LB medium and were purified on step gradients of 3 M and 5 M CsCl (12). Heteroduplexing and electron microscopy were performed under standard high-stringency conditions (11, 16) so that secondary structures were not particularly favored. Electron photomicrographs were made in a JEOL 100S instrument at a nominal magnification of ×5000.

RESULTS

In the course of our attempts to clone several alleles of D14S1 (6–8), a recombinant DNA library consisting of large fragments having one *Bam*HI end and one *Eco*RI end was constructed in the standard bacteriophage λ vector Charon 30 (13). Ligated DNA was packaged *in vitro* and plated on a host strain (DB1170) that is mutant at the *recB, recC*, and *sbcB* loci.

Using a spot test, we examined the ability of 514 recombinant bacteriophage clones grown on strain DB1170 to grow subsequently on the *rec*$^+$ host (BNN45). Of these, 46 (8.9%) were unable to grow at all or grew poorly. Four independent clones from these 46 (684-1, -2, -3, and -4) were chosen for further study, along with six independent control clones (F1 to F6) from the same library that showed no difference in growth on strains DB1170 and BNN45.

In order to obtain a more quantitative measure of the difference in phage growth between the two hosts, the efficiency of plating was measured (Table 1). Two kinds of growth

Table 2. Burst size in a single-step growth experiment

Phage	Yield, phage per input phage: DB1170 *recB recC sbcB*	Yield, phage per input phage: BNN45 *rec*$^+$
684-1	42	9.7
684-2	25	9.3
F1	51	100

Exponentially growing cells were concentrated and infected with phage at a multiplicity of infection of 0.01. Infected cells were diluted into warm medium and incubated for 60 min at 37°C.

defects were observed on strain BNN45: three clones, 684-2, -3, and -4, showed a complete failure to make plaques (efficiency less than 0.003), whereas clone 684-1 made tiny plaques (less than 0.1-mm diameter) with a good efficiency (about 0.5). All of the control phage clones plated at least as well on strain BNN45 as on strain DB1170.

Another measure of growth, the burst size in a single-step growth experiment was applied to phages 684-1 (small plaques on BNN45), 684-2 (no plaque on BNN45), and F1 (control). The result (Table 2) indicates that the ability to produce phage in a single cycle was indeed impaired for both 684-1 and 684-2 when strain BNN45 was the host, although the effect was modest—about one-third of that when DB1170 was the host. We conclude that efficiency of plating and plaque size are more useful indicators of growth differences in this instance than burst size.

Since BNN45 and DB1170 are not closely related strains of *E. coli*, we examined the efficiencies of plating and plaque sizes of the 684 phages on a series of strains that are closely related to DB1170 [all are derived from strain AB1157 (10)] but differ with respect to one or more of the *rec* and *sbcB* loci. These experiments were done with lysates grown on strain DB1170, which, although lacking the *E. coli* K restriction, nevertheless still modifies phage so that they become insensitive to the active restriction system retained by the other AB1157 derivatives. For this experiment, a wild-type (λ*c*I857) control was used in addition to one of the F series of human DNA clones in Charon 30.

The results of these comparisons are shown in Table 3, from which it is clear that the best hosts were strains DB1170, MO611 (the immediate parent of DB1170), and JC7623. These three strains have in common mutations in the *recB, recC*, and *sbcB* genes. The worst hosts were the related *rec*$^+$ strains MO626 (a close ancestor of DB1170) and AB1157 (a more distant ancestor). Strain MO626 plated the 684 phage clones in exactly the same way as did strain BNN45, suggesting that it is not the general strain background but instead the *recB, recC*, and/or *sbcB* mutations that are responsible for the differences in phage growth. Strain AB1157 appeared to plate phages 684-3 and 684-4 somewhat better than did strain MO626, yielding tiny plaques at a good efficiency. This result suggests that very slight differences in growth might account for the difference

Table 3. Efficiency of plating on genetically similar hosts

	recB recC sbcB			*rec*$^+$		*recB recC*	*sbcB*
Phage	DB1170	MO611	JC7623	MO626	AB1157	JC5519	MO627
684-1	(1.0)	1.5	1.1	0.59*	0.86	0.58	0.87
684-2	(1.0)	1.6	0.98	<0.02	<0.02	0.55	0.51*
684-3	(1.0)	1.6	1.2	<0.01	0.57*	0.74	0.86*
684-4	(1.0)	0.95	1.1	<0.005	0.56*	0.39	0.71*
F6	(1.0)	1.3	0.90	1.1	1.1	0.53	0.87
λcI857	(1.0)	1.1	1.1	0.95	1.2	0.90	0.98

For each phage, the efficiency of plating was normalized to the titer on DB1170. Complete genotypes are given in *Materials and Methods*.
*These plaques were tiny, less than 0.1-mm diameter.

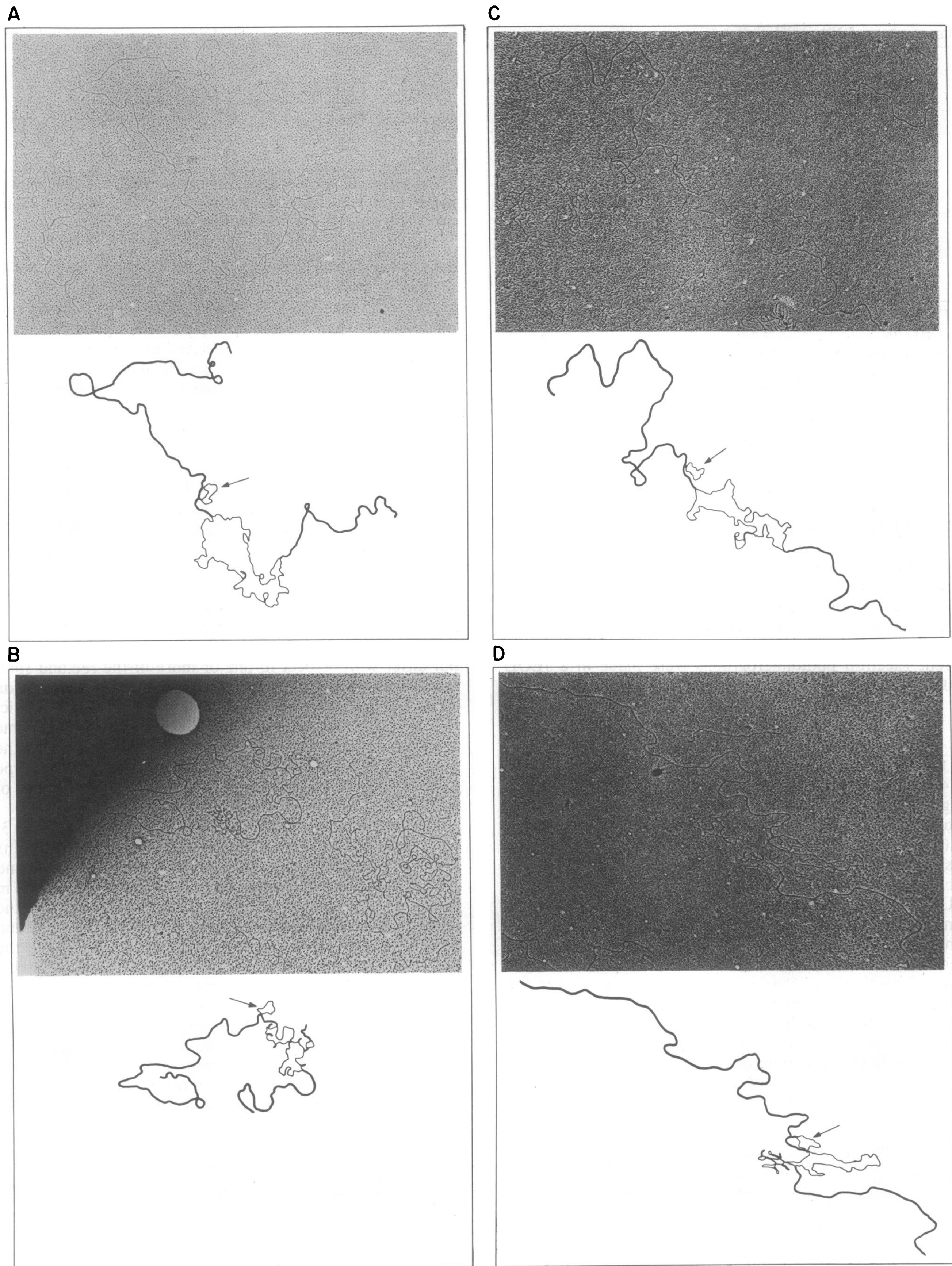
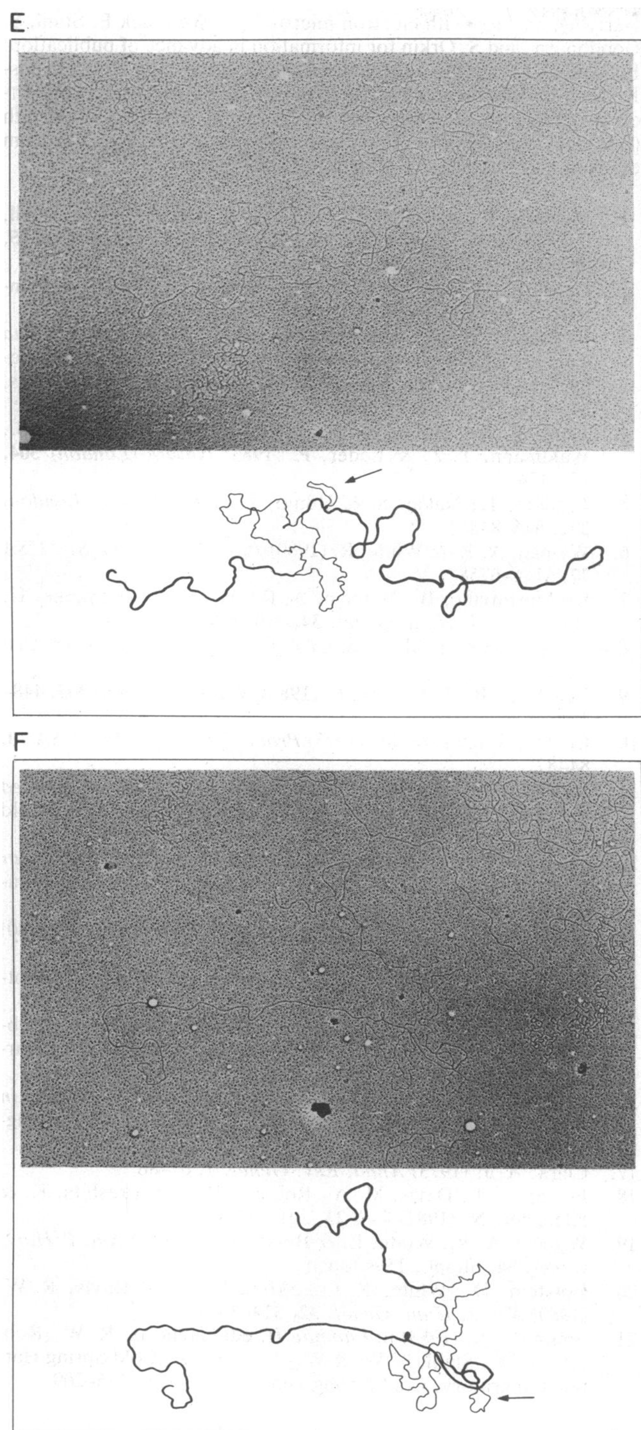


FIG. 1. Heteroduplexes of Charon 30::Tn*10*Δ with recombinant phages from Charon 30 library. Each electron micrograph is accompanied by a line drawing of the relevant molecule for the purpose of clarification. Arrows point to the 2.95-kilobase Tn*10*Δ insertion in Charon 30 left arm. The heavy lines correspond to the double-stranded left and right vector arms, while the thinner lines correspond to the single-stranded (nonhomologous) human and vector inserts. (*A*) Phage 684-1. (*B*) Phage 684-2. (*C*) Phage 684-3. (*D*) Phage 684-4. (*E*) Phage F4. (*F*) Phage F6. The small structure seen in *E* has been seen occasionally in heteroduplexes with several other library phages and is due to the presence of a small inverted repeat within the "stuffer fragment" of Charon 30.



between no visible plaques and the tiny plaques typically found with 684-1, for example. This idea is supported by the burst size measurements given above (Table 2).

Also shown in Table 3 are results of plating the phages on strains differing from DB1170 by only the *sbcB* mutation or the *recB* and *recC* mutations, so that the loss of the RecBC function (exonuclease V) and the SbcB function [exonuclease I (17)] could be assessed separately. The results show that the *sbcB* mutation alone (strain MO627) helped the growth of the 684 series, but only modestly (tiny plaques were made where none had been found, or tiny plaques became somewhat bigger). The *recB* and *recC* mutations, in contrast, had a stronger effect. All of the 684 phages plated well on strain JC5519, although the efficiency of plating in no case quite reached the levels characteristic of the hosts mutant in all three genes (*recB*, *recC*, and *sbcB*). In confirmation of this result, a relatively unrelated strain (SF8) carrying the mutations in *recB* and *recC* but not *sbcB* showed similar permissive plating properties with respect to the 684 phages (not shown).

The observations of Leach and Stahl (9) that bacteriophage λ genomes containing perfect inverted repetitions can plate on hosts mutant in *recB*, *recC*, and *sbcB* prompted us to examine the 684 phage clones for the presence of inverted repetitions and other unusual sequence organization. DNA from each phage clone was used to form a heteroduplex with Charon 30::Tn*10*Δ, a new derivative of Charon 30 that contains a 2.95-kilobase insertion (18) in the left phage arm near the cloning sites. Heteroduplexes thus formed consist of perfectly homologous phage left and right arms with an obvious stem-and-loop structure attributable to the Tn*10*Δ insertion in the left arm, just beyond which the nonhybridizing single strands corresponding to the human insertions and the vector "stuffer fragment" diverge. Such heteroduplexes allow one to examine whatever secondary structure may form within the human sequence by examining the single-stranded region in the center of the heteroduplex. The Tn*10* insertion serves as a standard for size measurements.

Fig. 1 *A–D* shows typical examples of heteroduplexes between Charon 30::Tn*10*Δ and the four phage clones that failed to grow on rec^+ host strains (i.e., the 684 phages). All four contain extensive secondary structure within the human inserts, including simple snap-back and stem-and-loop structures. Specifically, each contains at least a few simple structures consistent with a perfect or nearly perfect inverted repetition 200–500 base pairs in length. Phage 684-4 is particularly interesting, having a large number of inverted repeats that could reassociate in a variety of ways (data not shown), indicating that similar sequences several hundred base pairs in length are present many times in the human DNA insert. In contrast, the six control clones (the F series) that were able to grow equally well on BNN45 and DB1170 were also examined, and none of these gave any evidence of secondary structure in the human DNA insert (two examples are shown in Fig. 1 *E* and *F*).

These results suggest quite strongly that there is indeed a connection between the plating phenotype of human DNA clones in the phage λ vector and the presence of inverted repetitions 200–500 base pairs in length in the human DNA insert. In support of this notion is the observation that the DNA of D14S1 also shows some secondary structure and shares the growth phenotype (19).

DISCUSSION

The primary observation we have presented is that a sizable fraction of the human genome, at least 8.9%, fails to survive as bacteriophage λ clones in rec^+ bacterial hosts. This means that the standard human genomic libraries made in phage λ vectors, including the widely used libraries of Lawn *et al.* (2), must be missing many human DNA sequences since they were amplified on rec^+ hosts.

One characteristic of the sequences requiring mutant hosts for their propagation in phage λ vectors appears to be the presence of inverted repetitions large enough to be visible in the electron microscope. This result is in excellent agreement with the observations of Leach and Stahl (9), who found that deliberately constructed perfect inverted repetitions could be propagated only in *recB recC sbcB* hosts. They also found that, even in *recB recC sbcB* hosts, the perfectly palindromic sequences often suffer deletions. The deletion derivatives show increased ability to grow in *recB recC* hosts not carrying the *sbcB* mutation (D. R. F. Leach, personal communication). We do not know whether the four clones we examined in detail, which contain relatively short

(200–500 base pairs) inverted repetitions, have undergone any deletions or other rearrangements, though we do find that they grow reasonably well on *recB recC* hosts. We have observed that clones of D14S1 that were recovered on strain DB1170 have undergone extensive deletion in the hypervariable region (19); some of these deletion derivatives grow on rec^+, consistent with the results of Leach and Stahl (9). In general, the results support a relationship between the plating properties and the amount of secondary structure found; for example, phage 684-1 has the least secondary structure and grows best on all hosts.

It should be pointed out that the estimate of the fraction of the human genome missing from phage λ libraries is likely to be an underestimate, since we have no way to know that DB1170 itself can propagate every human DNA sequence. Further, the observed instability of D14S1 sequences in our hands and perfect palindromes in the hands of Leach and Stahl suggest that *recB recC sbcB* is not a perfect host. Additional mutations may improve the situation: for example, addition of a *recA* mutation might reduce instability since the ensemble of mutations *recB recC sbcB* results in a recombination-proficient phenotype (17) that might make direct repetitions unstable, producing deletions.

Our initial interest in screening for clones that fail to grow on rec^+ was to determine whether the presence of inverted repeats is associated with a high degree of polymorphism of the insertion/deletion type, such as is seen at D14S1 (6). One of four clones examined is derived from a locus in the genome that is the site of an insertion/deletion dimorphism (not shown). While this locus is not itself useful as a general genetic marker [two alleles, low minor allele frequency (20)], the finding does suggest that an extensive survey of clones that fail to grow on rec^+ might turn up other insertion/deletion polymorphisms, perhaps those with as large a number of alleles as D14S1.

Note Added in Proof. Recent results indicate that the genotype of the vector as well as of the host can influence the viability of the subset of human DNA clones identified above. Specifically, Charon 30 (the vector we used) lacks Chi sequences (21). Since the recombinant clones lack the phage *red* and *gam* genes, plaque size on rec^+ hosts should be increased by the presence of Chi sequences in the human DNA inserts. The possibility exists that growth on rec^+ hosts may be poorest in those phage clones which, in addition to containing inverted repetitions, also lack a Chi sequence. In a preliminary test, we have obtained evidence that introduction of Chi elements renders the 684-4 clone more viable on rec^+ hosts. The difference between viability on rec^+ and *recB recC sbcB* hosts of inserts containing secondary structure elements like that in 684-4 may thus be less dramatic, in terms of number and degree, in vectors containing Chi sequences in the arms.

We are indebted to Paula Grisafi for help in the construction of Charon 30::Tn*10*Δ, to Paul Riggs for strain DB1170, and Erika Hartwieg for help with electron microscopy. We thank F. Stahl, S. Goodbourn, and S. Orkin for information in advance of publication, F. Stahl and R. Bourret for valuable suggestions, and Helen Donis-Keller for critical reading of the manuscript. Research was supported by grants to A.R.W. from the National Institutes of Health (GM30467) and the Wills Foundation as well as grants to D.B. from the National Institutes of Health (GM18973 and GM21253).

1. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
2. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
3. Perlmutter, R. M., Ram, D. & Hood, L. (1984) in *Genes and Cancer*, UCLA Symposium on Molecular and Cellular Biology, New Series, eds. Bishop, J. M., Rowley, J. D. & Greaves, M. (Liss, New York), Vol. 17, pp. 489–499.
4. Taub, R. A., Hollis, G. F., Hieter, P. A., Korsmeyer, S., Waldmann, T. Z. & Leder, P. (1983) *Nature (London)* **304**, 172–174.
5. Nikaido, T., Nakai, S. & Honjo, T. (1981) *Nature (London)* **292**, 845–848.
6. Wyman, A. R. & White, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6754–6758.
7. De Martinville, B., Wyman, A. R., White, R. & Franke, U. (1982) *Am. J. Hum. Genet.* **34**, 216–226.
8. Ferguson-Smith, M. A. & Cox, D. R. (1982) *Cytogenet. Cell Genet.* **32** (1–4).
9. Leach, D. R. F. & Stahl, F. (1983) *Nature (London)* **305**, 448–451.
10. Cosloy, S. & Oishi, M. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 84–87.
11. Davis, R. W., Botstein, D. & Roth, J. R. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
12. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
13. Rimm, D. L., Horness, D., Kucera, J. & Blattner, F. R. (1980) *Gene* **12**, 301–309.
14. Maurer, R., Osmond, B. C., Shekhtman, E., Wong, A. & Botstein, D. (1984) *Genetics* **108**, 1–23.
15. Murray, N. E. (1983) in *Lambda II*, eds. Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 395–432.
16. Ferguson, J. & Davis, R. W. (1978) in *Advanced Techniques in Biological Electron Microscopy II*, ed. Koehler, J. K. (Springer, New York), pp. 123–171.
17. Clark, A. J. (1973) *Annu. Rev. Genet.* **7**, 67–86.
18. Foster, T. J., Davis, M. A., Roberts, D. E., Takeshita, K. & Kleckner, N. (1981) *Cell* **23**, 201–213.
19. Wyman, A. R., Wolfe, L. & Botstein, D. (1984) *Am. J. Hum. Genet.* **36**, Suppl., 159s (abstr.).
20. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
21. Smith, G. R. (1983) in *Lambda II*, eds. Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 175–209.