Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 5976–5980, October 1982
Genetics

Clustering of spore-specific genes in *Aspergillus nidulans*

(gene organization/gene expression/fungi/conidia)

WILLIAM C. ORR* AND WILLIAM E. TIMBERLAKE†

The Plant Growth Laboratory and Department of Bacteriology, University of California, Davis, California 95616

Communicated by Norman H. Giles, June 29, 1982

ABSTRACT **We have investigated the chromosomal organization of genes that are expressed specifically in the asexual spores (conidia) of the Ascomycete fungus *Aspergillus nidulans*, using two experimental approaches. In the first, 30 different recombinant clones, containing long nuclear DNA inserts and at least one spore-specific gene, were selected randomly. The total number of spore-specific genes present in each clone was then determined by RNA blot analysis. In the second approach, several chromosomal recombinant DNA libraries, having average insert lengths ranging from 1 to 15 kilobases, were constructed. The fraction of clones in each library having one or more spore-specific poly(A)$^+$RNA-coding regions was then determined by colony or plaque filter hybridization with radiolabeled, spore-specific, complementary DNA. The results from these experiments were compared to statistical predictions based on the assumption that the spore-specific genes are randomly distributed in the *Aspergillus* genome. In both cases, the experimental values deviated significantly from the predicted values, demonstrating that the spore-specific genes are nonrandomly arranged in the genome. Rather, they appear frequently to occur in tightly linked clusters.**

The Ascomycete fungus *Aspergillus nidulans* reproduces asexually by forming spores, called conidia, on specialized, multicellular structures termed conidiophores (1, 2). We have shown that conidiophore development and spore formation are characterized by complex alterations in gene expression (3). Somatic cells (hyphae) grown in enriched medium contain 5,600–6,000 different poly(A)$^+$RNA sequences, equivalent to 26–28% of the total genomic complexity. Conidiating cultures have virtually all of the diverse poly(A)$^+$RNAs found in hyphae plus about 1,300 stage- and cell-specific sequences. Approximately 1,000 of the developmentally regulated poly(A)$^+$RNA sequences accumulate preferentially during conidiation but cannot be detected in spores. The remainder, corresponding to 1.4% of the *Aspergillus* genome, occur specifically in spores. It is probable that spore-specific poly(A)$^+$RNAs are mRNAs specifying proteins that function in conidium differentiation or germination (3–5). We are interested in the mechanisms that control the cell-specific expression of the genes coding for these mRNAs.

In the course of selecting chromosomal recombinant DNA clones that contain genes whose expression is regulated during conidiation, we obtained two types of evidence that raised the possibility that the spore-specific genes may be nonrandomly associated with one another in the *Aspergillus* genome (4). First, many fewer recombinant phage plaques gave positive autoradiographic signals when hybridized with radiolabeled, development-specific (largely spore-specific) cDNA probes than was expected based on the complexity of the cDNA and the average length of the cloned DNA inserts. Second, a few randomly selected clones that contained at least one spore-specific gene were found to hybridize to several poly(A)$^+$RNAs having different molecular weights, each of which was much more prevalent in conidia than in hyphae. This is a highly improbable event if it is assumed that the RNAs are not closely related in sequence and that the spore-specific RNA-coding regions are randomly distributed in the genome.

Clearly, close physical association of genes that are active in only one differentiating cell type may have important implications concerning the molecular processes that control and coordinate their expression or that have been responsible for their evolution, or both. Therefore, we have investigated the chromosomal arrangement of the spore-specific genes of *A. nidulans* in more detail. The results from our experiments show that these genes are very frequently organized into clusters.

METHODS AND MATERIALS

Bacteria. *Escherichia coli* K-12 strain HB101 (6) was used to propagate plasmids and strain χ2098 (DP50·SupF; ref. 7) was used to propagate phages.

Plasmids and Phages. Phages λ Charon 4A and λ Charon 4 (8) were used to construct chromosomal recombinant DNA libraries having average insert lengths of about 15 kilobases. Plasmid pBR322 (9) was used to construct libraries having average insert lengths of 1–12 kilobases (see below).

Aspergillus. *A. nidulans* Glasgow wild type (Ve$^+$) was used throughout this study.

Nucleic Acids. Polysomes, poly(A)$^+$RNA, and high molecular weight nuclear DNA were isolated from *Aspergillus* as described (3–5, 10–14). Plasmid DNA was prepared essentially by the procedure of Clewell and Helinski (15). Phage DNA was isolated as described by Thomas and Davis (16). DNA was labeled by nick-translation (17, 18) to 3×10^7 cpm/μg. The procedures used for cDNA synthesis and preparation of spore-specific cDNA by cascade hybridization have been described in detail (3).

Recombinant DNA Libraries. The construction of an *Eco*RI random partial digestion library, formed between *Aspergillus* nuclear DNA and λ Charon 4A, has been described (4). We have revised our estimate of the average insert length in this library to 15 kilobases based on the results of this investigation. A *Hae* III/*Alu* I partial digestion library was constructed by using λ Charon 4 and the procedure of Maniatis *et al.* (19). We have determined that approximately 70% of the phages in this library are recombinants, that the average insert length is about 15 kilobases, and that 95% or more of the *Aspergillus* genome is randomly represented in the collection (unpublished results).

Plasmid libraries having average insert lengths of 1–12 kilobases were constructed as follows. Samples of *Aspergillus* nuclear DNA were treated separately with *Sau* 3A (0.05, 0.1, and 0.2 units/hr per μg of DNA), and the reactions were terminated

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.

* Present address: Cellular and Developmental Biology, The Biological Laboratories, Harvard University, Cambridge, MA 02138.
† To whom reprint requests should be addressed.

by adding an excess of EDTA and heating to 68°C. Each Sau 3A-digested DNA (6 μg) was mixed with an estimated 5-fold molar excess of pBR322 DNA that had been digested with *Bam*HI and treated with alkaline phosphatase. After ligation with T4 DNA ligase, the DNAs were used to transform *E. coli* to ampicillin resistance by using the Ca^{2+}-heat shock procedure. More than 40,000 recombinant DNA clones were obtained for each of the three primary libraries. The colonies were pooled and stored as heavy suspensions in 50% (vol/vol) glycerol at −80°C.

The primary libraries were grown separately for three doublings in L-broth containing ampicillin (50 μg/ml), and plasmid DNAs were isolated. The DNAs were pooled and centrifuged on 10–40% sucrose gradients (20) for 22 hr at 24,000 rpm and 25°C in a Beckman SW 28.1 rotor. The gradients were fractionated, and samples from each fraction were analyzed by electrophoresis on 0.8% agarose gels (21). Several covalently closed circular DNA markers were fractionated in parallel. Fractions containing plasmids having the desired insert sizes were used to transform *E. coli*. Five secondary libraries were constructed in this way.

The secondary libraries were grown for three doublings, the plasmid DNAs were isolated, and the modal insert sizes were estimated by agarose gel electrophoresis and comparison to covalently closed circular DNA markers. The characteristics of the five plasmid and two phage libraries are summarized in Table 1.

Filter Hybridizations. The procedures used were as follows: DNA blots as described by Southern (22); RNA blots, Thomas (23); plaque filter hybridizations, Benton and Davis (24); and colony hybridizations, Hanahan and Meselson (25). Filters were used to expose preflashed Kodak XAR5 film at −70°C with the aid of a Gafmed SF-3 x-ray-intensifying screen.

Materials. Restriction endonucleases, T4 DNA ligase, *E. coli* DNA polymerase I, and agarose were purchased from Bethesda Research Laboratories; BA85 nitrocellulose, from Schleicher and Schuell; [α-^{32}P]dCTP (400 Ci/mmol; 1 Ci = 3.7 × 10^{10} becquerels), from Amersham; calf intestinal mucosa alkaline phosphatase, from Boehringer Mannheim; AMV reverse transcriptase, from Life Sciences (St. Petersburg, FL). Other reagents were purchased from Sigma.

Table 1. Characteristics of the *Aspergillus* recombinant DNA libraries

Library	Pooled clones × 10^{-3}	Fraction containing inserts*	Modal insert length,† kb	Genome equiv.‡
pB	30	0.97	1	1.1
pC	15	0.87	3	1.5
pD	7	0.98	6	1.6
pE	3	0.99	9	1.0
pF	6	0.94	12	2.6
λI	250	0.95	15	140
λII	80	0.70	15	32

* Determined by replica-plating representative samples on medium containing ampicillin or tetracycline (plasmid clones) or by hybridization of plaque filter replicas with radiolabeled nuclear DNA (phage clones).

† Plasmid insert sizes were determined by electrophoresis on agarose gels and comparison to covalently closed circular DNA markers having insert lengths ranging from 0 to 13.3 kilobases (kb). Center points of the heterogeneous populations were used for these estimates. Phage insert sizes were estimated as described (4).

‡ equiv., Equivalents calculated as: [(fraction containing inserts) × (number of clones pooled) × (modal insert length)]/genome size. The genome size of *A. nidulans* is 2.6 × 10^7 base pairs (14).

RESULTS

We used two different experimental strategies to investigate the chromosomal organization of spore-specific genes in *A. nidulans*. The first was based on the fact that it is possible to estimate the probability that multiple genes from a given population will occur in a cloned DNA fragment by random chance, when the length of the fragment and the complexity of the gene population are known.

A segment of chromosomal DNA can be considered to consist of potential RNA-coding regions. The number of such regions, N, which will be partially or entirely included in the segment is equal to $(L_I + L_G)/L_G$, where L_I is the length of the DNA fragment and L_G is the length of an average RNA-coding region. The probability, p, that a potential coding region will be, in fact, complementary to an RNA sequence in the population under consideration is equal to the complexity of the population, expressed as a fraction of the genome. If the sequences coding for the population are randomly distributed in the genome, the probabilities, $P(X;p)$, that a cloned fragment containing at least one coding region will contain a total of X such sequences comprise a Poisson distribution:

$$P(X;p) = \frac{e^{-pN}(pN)^X}{X!(1 - e^{-pN})}. \quad [1]$$

If, on the other hand, the RNA-coding regions are not randomly arranged in the genome, more cloned DNA fragments having one sequence of interest will have additional ones than is predicted by this equation. Therefore, we compared the distribution of spore-specific genes in 30 chromosomal regions containing at least one such gene to the estimate made with the assumption that the genes are randomly distributed in the genome.

We have reported the selection and partial characterization of 370 phage λ clones that contain developmentally regulated poly(A)$^+$RNA-coding regions from *A. nidulans* (4). Over 200 of these cloned DNA sequences encode poly(A)$^+$RNAs that accumulate preferentially in spores. We obtained from this collection 30 clones that contained one or more spore-specific genes and were different from one another on the basis of electrophoretic patterns of their DNA restriction digests. The mean insert length of these clones is 15.1 kilobases.

To determine the number of different RNA sequences that are complementary to each of the 30 clones and their regulatory behavior, equal quantities of poly(A)$^+$RNA isolated from hyphae, sporulating cultures (containing hyphae, conidiophores, and conidia), and purified conidia were fractionated in parallel on denaturing agarose gels and blotted onto nitrocellulose filters. The filters were then individually hybridized with radiolabeled clone DNAs. Reconstruction experiments (5) showed that only RNA sequences representing 0.01% or more of the poly(A)$^+$RNA mass were detected with this procedure. Some representative results from these experiments are shown in Fig. 1. Some of the clones hybridized to only one poly(A)$^+$RNA sequence that is expressed specifically in spores (e.g., λAn Spo28). Many hybridized to several spore-specific RNA species but not detectably to any sequences that are not regulated during conidiophore development (e.g., λAn SpoC3, λAn SpoC13, and λAn SpoC1). Others hybridized to two or more poly(A)$^+$RNAs having very different regulatory properties (e.g., λAn Spo27 and λAn SpoC10).

It was possible that hybridization to multiple RNA bands was due either to transcription of repeated DNA sequences (multigene families) or to the presence of RNA precursors in the whole-cell poly(A)$^+$RNA preparations used and, therefore, might not accurately reflect the number of coding regions in

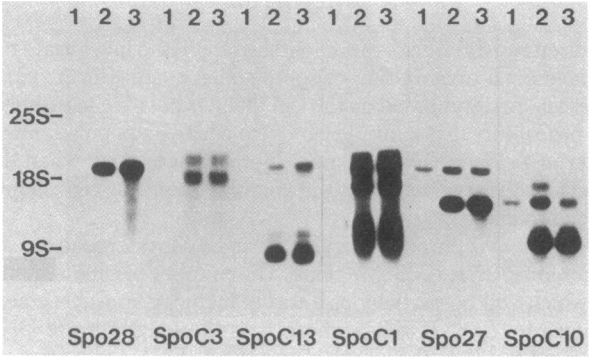

FIG. 1. RNA blot analysis of spore-specific phage clones. Equal amounts (3 μg) of poly(A)$^+$RNA from hyphae (lanes 1), unfractionated conidiating cultures (lanes 2), and spores (lanes 3) were subjected to electrophoresis on 1% agarose gels containing 2.2 M formaldehyde and were transferred to nitrocellulose. Individual filter strips were then hybridized with radiolabeled clone DNA, washed at 65°C in 0.04 M Na^+, and autoradiographed. The positions of molecular weight standards were determined by ethidium bromide staining and photography.

the clones. To test these possibilities, we performed two types of experiments. In the first, each cloned DNA fragment that reacted with two or more RNA species was hybridized to blots of *Aspergillus* nuclear DNA which had been digested with *Eco*RI or *Xba* I. In each case, the pattern of restriction fragments in genomic DNA was equivalent to that observed for the cloned DNA, demonstrating that the cloned sequences are not repeated in the genome.

In the second set of experiments, poly(A)$^+$RNA was isolated from spore polysomes (5) and subjected to RNA blot analysis as described above. The hybridization patterns were consistent with those obtained with total poly(A)$^+$RNA, showing that the sequences are not related as precursors and products.

Fig. 2 shows the distribution of spore-specific RNA-coding regions in the 30 clones we investigated, in comparison to the estimates based on the assumption that the genes are randomly dispersed in the genome. The results deviate significantly from the statistical predictions. It was expected that only two of the clones would contain two spore-specific genes and that a negligible fraction would contain three or more such sequences. In contrast to the expectation, five clones contained two spore-specific RNA-coding regions and more than one-third of the DNA fragments contained three or more such regions.

The validity of these experiments requires an accurate estimation of the complexity of the spore-specific poly(A)$^+$RNA population and random sampling of the recombinant DNA library. It is difficult to confirm experimentally that these requirements have been met. Therefore, we adopted a second experimental strategy, which does not depend on these parameters, to estimate linkage tendencies of the spore-specific genes.

By following the arguments used in the derivation of Eq. 1, the fraction, F, of chromosomal clones having a mean insert length, L_I, that are expected to contain no members of a randomly dispersed gene population can be expressed as:

$$F = e^{-p[(L_I+L_G)/L_G]}, \quad [2]$$

or

$$F = e^{-p[1 + (L_I/L_G)]}, \quad [3]$$

or

$$\ln F = -p(L_I/L_G) - p. \quad [4]$$

Therefore, ln F is a linear function of L_I, having a slope of $-p/L_G$ and an ordinate intercept of $-p$. If the genes are not randomly arranged in the genome, the experimentally determined slope of the line will be less negative than expected. However, the intercept at zero fragment length will remain unchanged, thus providing an internal control for the complexity of the gene population. Therefore, to extend our observations on the organization of the spore-specific genes, we determined values of F for the population at a variety of cloned DNA fragment lengths and compared the results to the statistical expectations based on the assumption that the genes are randomly dispersed in the genome.

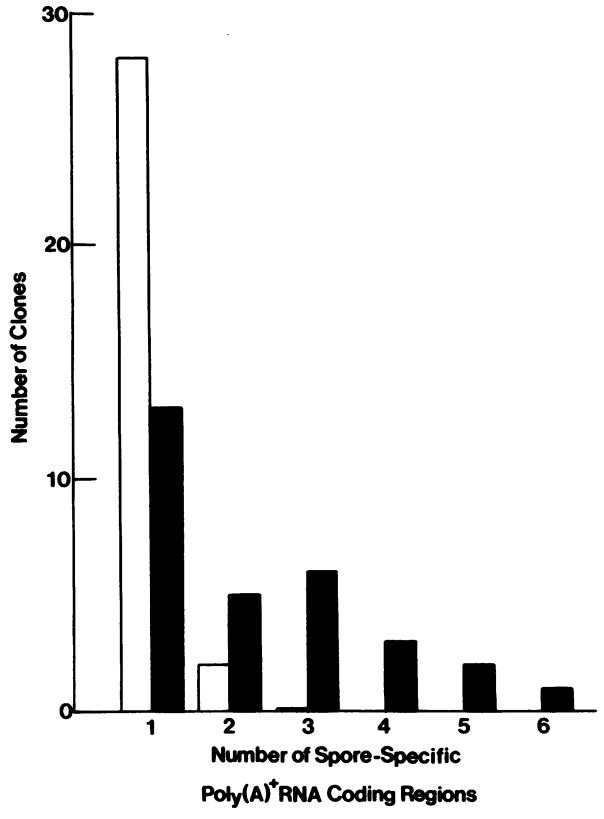

FIG. 2. Distribution of spore-specific genes in cloned *Aspergillus* DNA fragments. The number of spore-specific RNA-coding regions in each of 30 *Aspergillus* recombinant DNA clones having at least one such region was determined experimentally (■) as described in the legend to Fig. 1. Predicted values (□) were calculated by using Eq. 1 and the following parameters: L_I = 15.1 kilobases, L_G = 1,800 nucleotides, $N = (L_I + L_G)/L_G$ = 9.4, and p = 0.014. The mean cloned fragment length (L_I) was determined by agarose gel electrophoresis of *Eco*RI-digested clone DNA. The mean length of spore-specific coding regions (L_G) was obtained by averaging the lengths of the 69 spore-specific poly(A)$^+$RNAs observed. The complexity of the spore-specific poly(A)$^+$RNA population was considered to be 3.6×10^5 nucleotides (3), equivalent to 1.4% of the total genomic complexity (14). The spore-specific sequences individually represent 0.01% or more of the mass of the poly(A)$^+$RNA (3).

Seven quasi-random recombinant DNA libraries having average insert lengths ranging from 1 to 15 kilobases were constructed. The characteristics of these libraries are summarized in Table 1. To determine the fraction of cloned inserts that did not contain sequences complementary to spore-specific poly(A)$^+$RNA, samples of 400–1,000 clones were plated and subjected to colony or plaque filter hybridization with a radiolabeled, spore-specific cDNA probe (p = 0.014). As a control, similar samples were hybridized with cDNA synthesized from total hyphal poly(A)$^+$RNA. The effective complexity of this probe was p = 0.018, rather than 0.27, because the sensitivity of the filter hybridizations was such that only abundant and

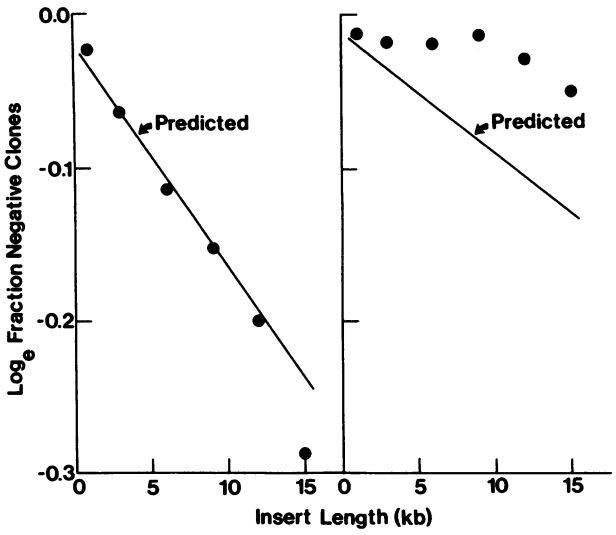

FIG. 3. Hybridization of recombinant DNA libraries with hyphal (*Left*) and spore-specific cDNA (*Right*). Samples of clones from the libraries described in Table 1 were subjected to colony or plaque filter hybridization with radiolabeled hyphal or spore-specific cDNA and, after autoradiography, the fraction of nonreactive clones was determined. The sample sizes used were 400 for the plasmid libraries and 1,000 for each of the phage libraries. The results for the two phage libraries have been averaged. The predicted lines were calculated from Eq. 4 with the following parameters: hyphal abundant and moderately abundant cDNA, $p = 0.018$ and $L_G = 1{,}200$ nucleotides (3); spore-specific cDNA, $p = 0.014$ and $L_G = 1{,}800$ nucleotides.

moderately abundant sequences would be detected (3, 4).

The results from these experiments are summarized in Fig. 3. In the case of hyphal cDNA, the experimentally determined values of F were in close agreement with those predicted from Eq. 4. These results indicate that the genes coding for abundant and moderately abundant poly(A)$^+$RNA in hyphae are randomly dispersed in the genome and that the statistical treatment used is valid. In contrast to this control experiment, the values of F for the spore-specific cDNA probe were similar to those predicted by Eq. 4 only at the shortest fragment length investigated. At other fragment lengths, many fewer clones yielded positive hybridization signals than was predicted from the assumption that the genes are randomly arranged in the genome. The ordinate intercept of the experimental line, derived from linear regression analysis, was -0.012, in agreement with our previous estimate (3) of the complexity of the spore-specific poly(A)$^+$RNA population.

DISCUSSION

We have conducted two sets of experiments designed to test the hypothesis that the spore-specific genes of *A. nidulans* are nonrandomly arranged (clustered) in the genome. In the first, recombinant DNA clones containing at least one spore-specific gene were examined for the presence of additional, similarly regulated genes. The distribution of spore-specific genes deviated significantly from that predicted with the assumption that the genes are randomly dispersed in the genome. A total of 69 spore-specific genes was detected in the 30 cloned sequences examined. Thus, an average clone in the sample contained 2.3 spore-specific genes, as compared to an expected average of 1.1. More than 50% of the clones contained two or more spore-specific genes, in contrast to the statistical prediction of only 7%. Twelve of the clones had three or more spore-specific RNA-coding regions. The probability of this occurring by chance association is negligible. Finally, more than 80% of the spore-specific genes detected (56 of 69) occurred in close association with one or more similarly regulated genes, as compared to a predicted value of 12.5% (4 of 32).

In the second set of experiments, samples of clones from differently sized, quasi-random recombinant DNA libraries were hybridized with spore-specific cDNA to determine the fraction of clones at each fragment length that did not contain spore-specific coding regions. As in the first set of experiments, the results differed substantially from the statistical expectations. Fewer clones yielded positive hybridization signals with spore-specific cDNA than predicted at all fragment lengths tested.

The results from these two experimental approaches can be compared quantitatively for the 15-kilobase size class. At this fragment length, 12.1% of the clones were expected to give positive hybridization signals if the genes were randomly distributed. On the other hand, if the distribution of genes observed in the sample of 30 spore-specific clones shown in Fig. 2 is characteristic of the entire population, only about 6% of the clones would hybridize with the spore-specific cDNA probe. The latter value is very similar to that actually observed (4.9%). This result strongly suggests that the 30 clones examined in this study represent a random sample of the entire spore-specific gene population. Therefore, it would appear that the spore-specific genes are not randomly distributed in the *Aspergillus* genome but rather frequently occur in close association with one another.

We have described the organization and regulated expression of one spore-specific gene cluster (5). This 13.3-kilobase chromosomal region, designated SpoC1, contains six discrete genes that code for polysomal poly(A)$^+$RNAs found only in conidia. These genes are all expressed at the same time in synchronously conidiating cultures. Therefore, it would appear that the SpoC1 region constitutes a functional unit of developmental gene regulation. The results presented here indicate that this type of organization is a general feature of the spore-specific genes of *Aspergillus*. The exact mechanisms that coordinate the expression of the gene clusters during spore formation and control the activities of genes within the clusters remain to be determined.

We thank Drs. Page Painter, Albert Siegel, and Robert Goldberg for helpful discussions during the preparation of the manuscript and Mr. Charles Zimmermann for assistance in selecting the spore-specific phage clones. This research was supported by grants from the National Science Foundation and the National Institutes of Health to W.E.T.

1. Thom, C. & Raper, K. (1945) *A Manual of the Aspergilli* (Williams & Wilkins, Baltimore).
2. Oliver, P. T. P. (1972) *J. Gen. Microbiol.* **73**, 45–49.
3. Timberlake, W. E. (1980) *Dev. Biol.* **78**, 497–510.
4. Zimmermann, C. R., Orr, W. C., Leclerc, R. F., Barnard, E. C. & Timberlake, W. E. (1980) *Cell* **23**, 709–715.
5. Timberlake, W. E. & Barnard, E. C. (1981) *Cell* **26**, 29–37.
6. Boyer, H. W. & Rouland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–472.
7. Leder, P., Tiemeier, D. & Enquist, L. (1977) *Science* **196**, 175–177.
8. Blattner, F. R., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L., Grunwald, D. J., Kiefer, D. A., Moore, D. O., Schumm, J. W., Sheldon, E. L. & Smithies, O. (1977) *Science* **196**, 161–169.
9. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. & Boyer, H. W. (1977) *Gene* **2**, 95–113.
10. Timberlake, W. E. (1976) *Dev. Biol.* **51**, 202–214.
11. Timberlake, W. E., Shumard, D. S. & Goldberg, R. B. (1977) *Cell* **10**, 623–632.
12. Rozek, C. E., Orr, W. C. & Timberlake, W. E. (1978) *Biochemistry* **17**, 716–722.

13. Gealt, M. A., Sheir-Neiss, G. & Morris, N. R. (1976) *J. Gen. Microbiol.* **94,** 204–210.
14. Timberlake, W. E. (1978) *Science* **202,** 973–974.
15. Clewell, D. B. & Helinski, D. R. (1969) *Proc. Natl. Acad. Sci. USA* **62,** 1159–1166.
16. Thomas, M. & Davis, R. W. (1974) *J. Mol. Biol.* **91,** 315–328.
17. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1184–1188.
18. Rigby, P. W. J., Dieckman, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113,** 237–251.
19. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15,** 687–701.
20. Neal, M. W. & Florini, J. R. (1972) *Anal. Biochem.* **45,** 271–276.
21. Johnson, P. H. & Grossman, L. I. (1977) *Biochemistry* **16,** 4217–4225.
22. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
23. Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5201–5205.
24. Benton, W. D. & Davis, R. W. (1977) *Science* **196,** 180–182.
25. Hanahan, D. & Meselson, M. (1980) *Gene* **10,** 63–67.