Characterization of the Sequence Complexity and Organization of the *Neurospora crassa* Genome†

Robert Krumlauf* and George A. Marzluf‡

ABSTRACT: Reassociation kinetic experiments with total DNA and isolated kinetic components are presented that characterize the complexity and organization of sequences in the genome of *Neurospora crassa*. The genome was found to have a size of 2.7×10^7 nucleotide pairs and to consist of unique (90%), repetitive (8%), and foldback (2%) elements. The arrangement of the repetitive sequences was determined from a variety of studies which examined the properties of long fragments and their reassociation. The results all indicate that *Neurospora* repetitive sequences are not organized in a short-period interspersion pattern. The repetitive DNA occurs in clusters at least 10 000 nucleotide pairs long and possibly in considerably longer stretches up to 73 000 nucleotide pairs. The extreme length of the repetitive sequences indicated they are not interspersed in regular intervals with single-copy DNA. As an upper limit, only 0.7% of the single-copy sequences are contiguous with repetitive DNA in fragments 6400 nucleotides in length. These results reveal that *Neurospora* does not follow either the long- or short-period interspersed pattern, since essentially no interspersed sequences were found. The low complexity of the repetitive component and its reiteration frequency suggest that the repetitive DNA could largely be comprised of ribosomal gene sequences. The lack of interspersed repetitive DNA strongly argues that repetitive DNA does not have a regulatory function in *Neurospora*.

The genomes of all eucaryotes contain foldback, repetitive, and single-copy DNA as common elements of their sequence composition. The amount of the repetitive and single-copy components in eucaryotic genomes varies over an extremely wide range of values. However, the organization of these sequences appears to follow two general patterns in most organisms (Davidson et al., 1975a). The predominant pattern is characterized by the interspersion of repetitive sequences, 200–400 nucleotide base pairs (Nbp)[1] in length, throughout a major fraction of the genome; these short repetitive sequences are linked to single-copy DNA at intervals averaging 1000–2000 Nbp (Davidson & Britten, 1973). This arrangement, termed the short-period interspersion pattern, is observed in most organisms examined (Goldberg et al., 1975; Graham et al., 1974; Pearson et al., 1978). However the genomes of several organisms, mainly insects, have been shown to lack short repetitive sequences but have instead a long-period interspersion pattern (Crain et al., 1976). Their repetitive DNA is organized in long stretches (4000–10 000 Nbp) that are interspersed with single-copy DNA at intervals of at least 10 000 Nbp (Manning et al., 1975; Hudspeth et al., 1977; Wells et al., 1976).

The evolutionary significance and importance of these organizational patterns are not clear. However, since repetitive DNA is organized in such highly ordered patterns in nearly all eucaryotic organisms, it may provide an important function. Britten and Davidson postulated that the interspersed repetitive sequences may play a central role in the coordinate regulation of sets of unlinked structural genes (Davidson & Britten, 1973; Davidson et al., 1977). This theory is supported by the observation that in some organisms a majority of the sequences that are expressed are adjacent to repetitive sequences (Smith et al., 1974; Davidson et al., 1975b).

Fungi as a group generally have very small genomes and low amounts of repetitive DNA. Only limited information concerning their sequence organization and kinetic components is available. Only the sequence organization in *Achlya* has been described in any detail (Hudspeth et al., 1977). The fungus *Neurospora* has been extensively studied by using biochemical genetic methods to examine eucaryotic regulatory systems (Marzluf, 1977). Several regulatory circuits have been defined in *Neurospora*, within which multiple unlinked structural genes are positively and coordinately regulated (Reinert & Marzluf, 1975; Metzenberg & Nelson, 1977; Jacobson et al., 1977). It was of considerable interest to determine whether repetitive DNA in *Neurospora* is involved in this type of coordinate control. Therefore, it was important to know how the repetitive sequences are organized in the genome.

In this paper we describe experiments that provide a detailed characterization of the components which comprise the *Neurospora* genome. We also describe the organization of the repetitive and single-copy sequences. The results clearly show that *Neurospora* does not have a short-period interspersion pattern of repetitive sequences. The repetitive DNA is not interspersed with single-copy DNA but rather is clustered in extremely long stretches.

Materials and Methods

Isolation of DNA. Unlabeled and in vivo ^{32}P-labeled DNA was prepared from *Neurospora crassa* strain 74-OR23-1a. Mycelial cultures grown in Vogel's minimal medium for 14 h were harvested by filtration, rinsed with distilled water, and

† From the Developmental Biology Program, Department of Biochemistry, The Ohio State University, Columbus, Ohio 43210. *Received March 5, 1979.* This work was supported by Grant GM-23367 from the National Institutes of Health and Grant PCM78-16021 from the National Science Foundation.

* R.K. was supported by a National Institutes of Health Predoctoral Traineeship in Developmental Biology.

‡ G.A.M. is supported by National Institutes of Health Career Development Award GM-00052.

[1] Abbreviations used: EDTA, ethylenediaminetetraacetic acid; Nbp, nucleotide base pairs; HAP, hydroxylapatite (Bio-Rad, HTP grade); SSC, saline sodium citrate buffer (0.15 M NaCl and 0.015 M sodium citrate, pH 8); Nt, nucleotides; PB, sodium phosphate buffer (pH 6.8); rms, root mean square error; C_0t, initial DNA concentration (moles per liter) times time (seconds); k, second-order reassociation rate constant (M^{-1} s^{-1}); GF, goodness of fit.

0006-2960/79/0418-3705$01.00/0

blotted dry. The cell pad was suspended in isolation buffer (1 M sorbitol, 5% Ficoll 400, 20% glycerol, 5 mM $MgCl_2$, 10 mM $CaCl_2$, and 0.5% Triton X-100), added to bottles containing 0.45–0.50-mm glass beads, and homogenized for three 30-s pulses in a Braun homogenizer cooled with CO_2. Nuclei were purified from the homogenate by the method of Hautala et al. (1977). DNA was always extracted from isolated nuclei. The nuclei were lysed in 0.15 M NaCl, 50 mM EDTA, 2.0% sodium lauroylsarcosine, and 10 mM Tris, pH 7.6, and digested with 500 μg/mL proteinase K for 6 h at 37 °C. The lysate was heated to 60 °C for 15 min and then centrifuged for 30 min at 10000*g*. The supernatant was extracted with an equal volume of phenol–chloroform–isoamyl alcohol (24:24:1), and the aqueous phase was removed after centrifugation. The interphase was reextracted with buffer and phenol–chloroform–isoamyl alcohol. The aqueous phases were then combined and reextracted until the interface was clear. Two additional extractions with chloroform were performed to remove traces of phenol. The DNA was precipitated by the addition of 2.5 volumes of 95% ethanol and removed by spooling on a glass rod. The DNA was redissolved in 1 × SSC (0.1 M NaCl and 0.015 M sodium citrate, pH 8.0) and digested with 100 μg/mL RNase A (pretreated at 100 °C for 15 min), 100 units/mL RNase T_1 (pretreated at 80 °C for 10 min), and 0.2 units/mL α-amylase for 6 h at 37 °C. The solution was then adjusted to 2.0% sodium lauroylsarcosine and digested for 3 h at 37 °C with 100 μg/mL proteinase K. The DNA solution was deproteinized by extraction 3 times with phenol–chloroform–isoamyl alcohol and twice with chloroform. The preparation was then brought to 10 mM Tris, 3 mM EDTA, pH 8.0, by dialysis, and CsCl was added to a final density of 1.71 g/mL. The DNA was separated and purified from any residual RNA and protein by banding in preparative CsCl density gradients in a Beckman Ti 60 rotor at 42000 rpm for 55 h at 20 °C. The band containing the DNA was removed, dialyzed against 1 × SSC, then treated with Chelex (Sigma), and stored at −20 °C.

DNA was prepared from *Escherichia coli* C600 cells. Unlabeled and ^3H-labeled DNA was isolated by using the method for Marmur (1961). Cells were grown in L broth or L broth containing 250 μCi of [^3H]methylthymidine (New England Nuclear) to an A_{595} = 0.8, then collected by centrifugation, and lysed in 0.15 M NaCl, 0.1 M EDTA, pH 7.8, and 500 μg/mL lysozyme. Following incubation at 37 °C for 10 min, this mixture was adjusted to 2% sodium lauroylsarcosine and incubated at 60 °C for an additional 10 min. The DNA in this lysate was extracted and purified by using the same procedures described above for lysed nuclei from *Neurospora*.

DNA Labeling. DNA labeled in vivo was prepared by incubating 500-mL cultures of *Neurospora* in Vogel's medium minus phosphate supplemented with limited amounts of KH_2PO_4 and 2.5 mCi of carrier-free $Na^{32}PO_4$ (New England Nuclear) for 16 h. The labeled DNA was extracted from isolated nuclei as described above. Under these conditions when the amount of KH_2PO_4 used was 0.05 or 0.01%, the specific activity of the isolated DNA was approximately 4000 or 25000 cpm/μg, respectively.

Total nuclear DNA from *Neurospora* was also labeled in vitro by nick translation based on the procedure of Rigby et al. (1977) and Maniatis et al. (1975). *Neurospora* DNA in a reaction buffer was incubated for 1 min at room temperature after addition of DNase I to provide nicks in the DNA. The labeling was initiated by addition of DNA polymerase I to obtain a final reaction mixture containing 50 mM sodium phosphate buffer (PB), 5 mM $MgCl_2$, 15 μM dATP, dGTP, and dCTP, 200 μCi of [^3H]TTP (New England Nuclear; 61.5 Ci/mmol), 1 μg of *Neurospora* DNA, 1.3 ng of DNase I (Worthington), and 90 units of DNA polymerase I (Boehringer-Mannheim) in a total volume of 120 μL. The reaction mixture was incubated at 14 °C for approximately 2 h. The rate of incorporation of label into DNA was monitored by removing 1-μL samples and counting the acid-precipitable counts on Millipore filters. The reaction was terminated by extracting the mixture twice with an equal volume of chloroform. The aqueous phase was passed over a Sephadex G-50 column to remove most of the unincorporated label. The fractions excluded from the column were pooled and then adjusted to 0.12 M PB. The samples were then placed in a boiling water bath for 5 min to denature the DNA and immediately passed over hydroxylapatite (HAP) to remove any foldback sequences introduced by the labeling procedure. The DNA isolated under these conditions had a specific activity of about 2×10^7 cpm/μg and an average single-stranded size of 350 Nt. Preparative alkaline sucrose gradients were used to remove any residual unincorporated label or to isolate a particular fragment size.

Preparation and Sizing of DNA Fragments. All labeled and unlabeled DNA preparations, except for in vitro labeled DNA, were routinely sheared to a fragment length of 400 Nt by sonication in 1 × SSC at 4 °C by using a Branson 350 sonifier equipped with a microprobe. Each preparation was given six 30-s pulses at a power setting of 5.0. The 400-Nt fragments labeled by nick translation were isolated from preparative alkaline sucrose gradients by pooling the fractions greater than 350 Nt. A Virtis 60 homogenizer (Britten et al., 1974) was used to shear ^{32}P-labeled DNA samples at 3000 rpm for 20 s, 10000 rpm for 1 min, or 50000 rpm for 3 min to obtain a broad distribution of fragment lengths. These DNA preparations were combined and centrifuged in preparative 5–20% alkaline sucrose gradients. DNA fragments of discrete sizes, ranging from 200 to 11000 Nt in length, were isolated as single fractions from these gradients.

The modal single-stranded length of all DNA preparations was determined several times by sedimentation through shallow alkaline sucrose gradients (Hell et al., 1972). Linear gradients containing 5–11% w/w sucrose in 0.1 M NaOH, 0.9 M NaCl, and 3 mM EDTA were prepared in SW 41 tubes and centrifuged at 30000 rpm for 14–24 h at 20 °C. The gradients were calibrated by using molecular weight markers and a computer program designed to calculate sedimentation rates from each gradient, measured by refractive index, with the precise conditions of each run. Molecular weights were calculated by using the equations of Studier (1965). The lengths of several of the large-sized fragment preparations were verified by agarose electrophoresis using λDNA restriction fragments as markers.

DNA Reassociation. All DNA preparations for reassociation were filtered through Gelman GA8 filters and passed over Chelex and a mixed-bed resin (Bio-Rad AG 501-X8) and then dialyzed against 0.012 M PB. DNA renaturation experiments were performed in 0.12 M PB at 60 °C. Samples were sealed in siliconized glass capillaries or ampules and denatured in a boiling water bath for 10 min. After incubation at 60 °C to the desired C_0t [initial DNA concentration (moles per liter) times time (seconds)], the samples were rapidly frozen in dry ice–ethanol. Each sample was diluted to 0.03 M PB and bound to hydroxylapatite (HAP) at 60 °C. The single-stranded DNA fraction was eluted with 5 mL of 0.14 M PB and 0.2% sodium lauroylsarcosine, and the duplex

Table I: Reassociation of 400-Nt Total *Neurospora* DNA[a]

component	fraction of genome	rate of whole DNA, k_w (M^{-1} s^{-1})	rate of pure DNA, k_p (M^{-1} s^{-1})	$C_0t_{1/2,w}$ (M s)	frequency repetition	complexity (Nbp)
(A) Unconstrained Fit (No Parameters Fixed): GF = 2.06%; rms = 2.00%						
foldback	0.03					
repetitive	0.0718 (±0.0324)	6.430 (±8.760)	89.51	0.1556	149	1.21×10^4
single copy	0.9016 (±0.0112)	0.04300 (±0.00198)	0.0480	23.26	1	2.25×10^7
(B) Constrained Fit (Single-Copy Rate Fixed at 0.04 M^{-1} s^{-1}): GF = 2.06%; rms = 2.02%						
foldback	0.03					
repetitive	0.0708 (±0.0123)	4.510 (±0.998)	63.7	0.2217	112	1.72×10^4
single copy	0.8984 (±0.0123)	0.0400[a]	0.0450	25.00	1	2.42×10^7
(C) Constrained Fit (Repetitive Rate Fixed at 5.6 M^{-1} s^{-1}; Single-Copy Rate Fixed at 0.04 M^{-1} s^{-1}): GF = 2.08%; rms = 2.04%						
foldback	0.02					
repetitive	0.0825 (±0.0125)	5.600[a]	70.00	0.1780	140	1.53×10^4
single copy	0.9010 (±0.0089)	0.0400[a]	0.0450	25.00	1	2.42×10^7

[a] Rate constants were fixed from the fit to the data in Figure 2A,B.

fraction was subsequently eluted with two 2-mL washes of 0.5 M PB and 0.2% sodium lauroylsarcosine (Britten et al., 1974). The 5-mL single-stranded fraction was mixed with 8 mL of Instagel (Packard), and each 2-mL duplex rinse was mixed with 3 mL of distilled water and then 8 mL of Instagel. The samples were shaken to form an emulsion and cooled at 4 °C for 30 min before counting. The fraction of DNA in duplex form for each point was calculated by dividing the counts per minute in the duplex samples by the sum of the counts per minute in the single-stranded and duplex samples. The results for unlabeled DNA were determined from A_{260} readings of the fractions in a similar manner after correcting for the difference in the absorbance of single-stranded DNA due to hyperchromicity. The values for the reassociation parameters and curve fits of the data were determined from the best nonlinear least-squares solution for the data calculated by using the program of Pearson et al. (1977).

Results

Reassociation of Short Neurospora DNA. The reassociation kinetics of short DNA fragments can be used to characterize the rate constants and the number of sequence components in total DNA. Figure 1A shows the reassociation of unlabeled 400-Nt total nuclear DNA fragments from *Neurospora* relative to similar-sized fragments from *E. coli*. The best least-squares fit of the data revealed only a single second-order kinetic component of *Neurospora* DNA reassociating with a rate (k) of 0.05 M^{-1} s^{-1}, which represents about 5 times the complexity of *E. coli*. However, there is too much scatter in data from reassociation of unlabeled DNA, especially at low C_0t values, to reliably determine whether a small repetitive component is present. Therefore, in vitro [^{3}H]DNA and in vivo [^{32}P]DNA 400-Nt fragments were reassociated and analyzed under identical conditions, as shown in parts B and C of Figure 1. The reassociation data for all three preparations yielded very similar kinetic properties, indicating that the labeling procedures did not alter the reactivity of the DNA. Table I summarizes three different computer analyses of the data. The best "free fit" or unconstrained least-squares solution for the combined data is listed in Table IA. This solution reveals the presence of two second-order components, repetitive and single-copy DNA, represented by the dashed curves in Figure 1C. The repetitive component is a minor fraction of the DNA, comprising only 7% of the total DNA bound to HAP, and has a k of 6.43 M^{-1} s^{-1}. The single-copy sequences represent 90% of the HAP-bound DNA and have a k of 0.043 M^{-1} s^{-1}. At the earliest C_0t values (10^{-4} M s) about 3% of the sequences were already bound by HAP and are considered to be foldback DNA. Despite the large number of data points and good overall fit (rms = 2.0%), the small repetitive fraction of the genome makes it difficult to obtain an accurate determination of the rates and repetition frequencies for the various sequence components. This is illustrated by the large variance in the values for the repetitive rate constant, 6.43 ± 8.76 M^{-1} s^{-1}, and fraction of the genome, 7.18 ± 3.24%.

To further characterize the repetitive and single-copy sequences, we isolated ^{32}P-labeled fractions of each component. An enriched repetitive tracer was isolated by reassociation of total ^{32}P-labeled 400-Nt DNA fragments to a C_0t of 0.75 M s and represented the duplex fraction (9.9%) bound to HAP. The single-copy tracer was prepared by reassociating the total DNA to a C_0t of 15 M s where all repetitive sequences are in duplex structures. Upon hydroxylapatite fractionation, this tracer comprised the single-stranded fraction (63%) which was not bound to HAP. To more reliably estimate the rate constants and repetition frequencies of the individual components, we reassociated both tracers, each with a large excess of unlabeled total DNA (Britten et al., 1974), as shown in Figure 2. The analysis of the reassociation of the single-copy tracer revealed a single component (98%), with a k of 0.04 M^{-1} s^{-1} in whole DNA (Figure 2B). By fixing the rate constant for the single-copy component at this value, we obtained a constrained second fit for the total DNA reassociation data, as detailed in Table IB. This fit clearly reduced the error estimates for the repetitive parameter values without affecting error estimates of the overall fit.

The best least-squares solution for the reassociation of the enriched repetitive tracer in Figure 2B indicates two second-order components, the repetitive sequences and a small contaminating amount (8%) of single-copy DNA. The repetitive component comprises 67% of the tracer with a k of 5.6 M^{-1} s^{-1}. The remaining 25% of the tracer was bound to HAP at the earliest C_0t values and represents foldback sequences also enriched by the procedure used to isolate the repetitive DNA tracer. Since the repetitive tracer represents 9.9% of total DNA and 25% of these sequences are actually foldback DNA, we calculate that 2.4% (25% × 0.0990) of total DNA contains foldback sequences. At the C_0t of 0.75 M s used to isolate the repetitive tracer, 81% of repetitve sequences (with a k of 5.6 M^{-1} s^{-1}) should be in duplex form. Since 67% of the tracer was repeated DNA, then 8% [67%/(0.81 × 0.099)] of total DNA is repetitive DNA. These values for the foldback and repetitive DNA content are in good agreement with those calculated from the reassociation of whole DNA (Figure 1). Therefore, the rate constant for the repetitive

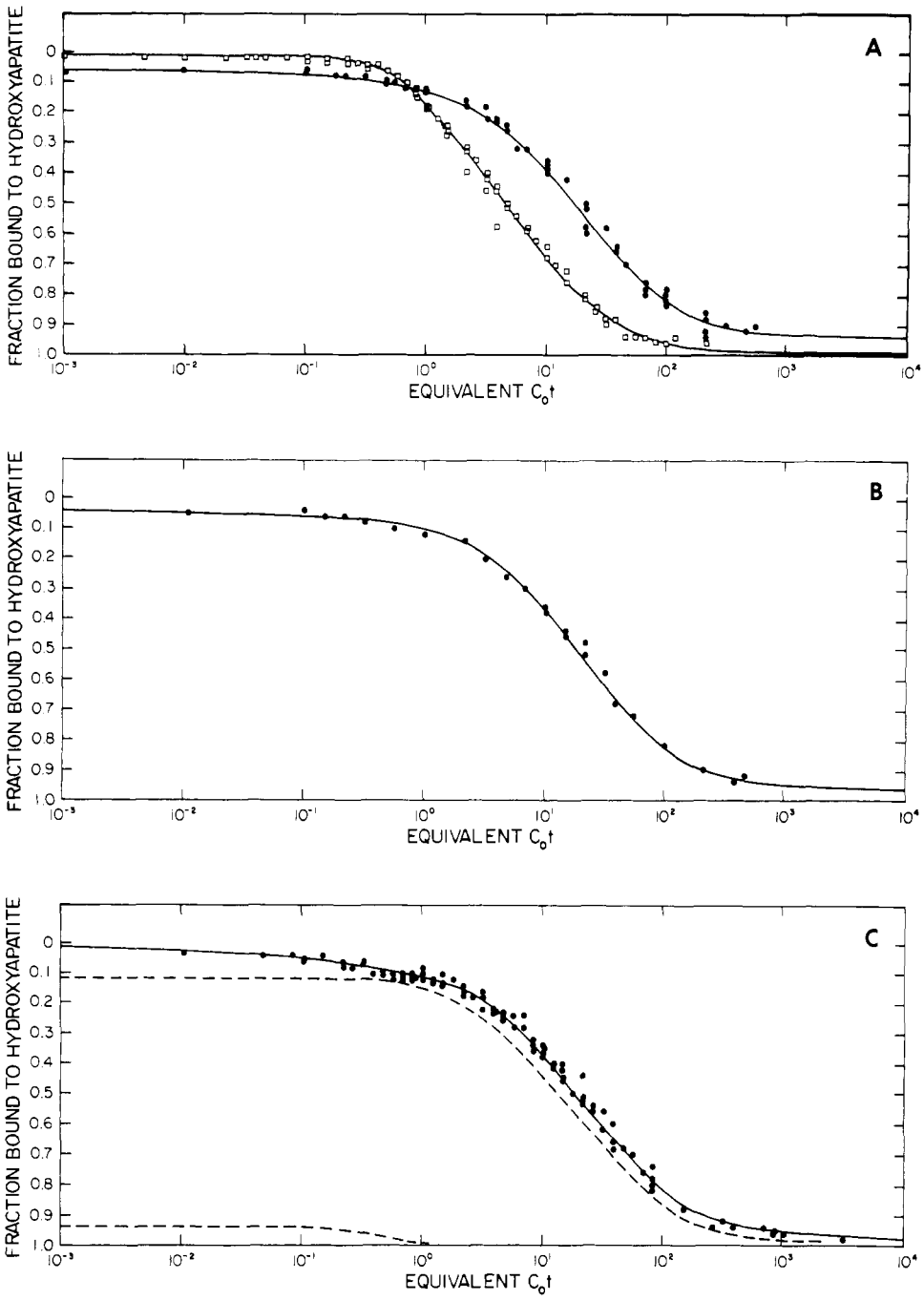

FIGURE 1: Reassociation kinetics of 400-Nt *Neurospora* DNA fragments. Samples were denatured and reassociated to various C_0t values and analyzed on hydroxylapatite as described under Materials and Methods. (A) Reassociation of unlabeled *Neurospora* DNA (●) and ^{3}H-labeled *E. coli* 400-Nt DNA fragments (□). The solid lines represent the best least-squares fit to the data. The *Neurospora* DNA reveals only a single second-order component (97%) with a k of 0.050 M^{-1} s^{-1}, and the *E. coli* is revealed as a single component (99%) with a k of 0.023 M^{-1} s^{-1}. (B) Reassociation of ^{3}H-labeled *Neurospora* DNA (●) labeled in vitro. The best least-squares solution for the data is shown by the solid line. The analysis indicates two second-order components: repetitive (6%) with a k of 6.3 M^{-1} s^{-1} and single copy (91%) with a k of 0.045 M^{-1} s^{-1}. The root mean square error of this solution is 2.6%. An additional 3% bound to HAP at the earliest times and was considered to be foldback DNA. (C) Reassociation of in vivo ^{32}P-labeled *Neurospora* DNA (●). The solid line represents the fit to the data, and the dashed lines portray the two second-order components from the best least-squares solution of the data (rms = 2.00%). The repetitive component comprises 7% of the HAP-bound DNA with a k of 6.43 M^{-1} s^{-1}, and the single-copy DNA comprises 90% with a k of 0.043 M^{-1} s^{-1}. The remaining 3% of the DNA is foldback sequences.

component was fixed at 5.6 M^{-1} s^{-1}, and a third interpretation of the data was obtained by computing another constrained least-squares solution using fixed rates for both the repetitive and single-copy components (Table IC).

There is essentially no difference in the root mean square error or goodness of fit parameters for the overall fit of the three solutions in Table I. However, since solution C is based on the most accurate estimates of the individual rate constants and minimizes the degree of variance for the other parameters, it provides the best characterization of the sequence components in the *Neurospora* genome. These values will be used in all future calculations. Relative to the rate of reassociation of *E. coli*, the size of the *Neurospora* genome was determined to be 2.7×10^7 Nbp.

The criterion used in these experiments, 60 °C and 0.12 M PB, should be able to detect duplexes of 25 Nbp or longer (Britten et al., 1974). It is possible that a component containing shorter repetitive duplexes or highly diverged repetitive

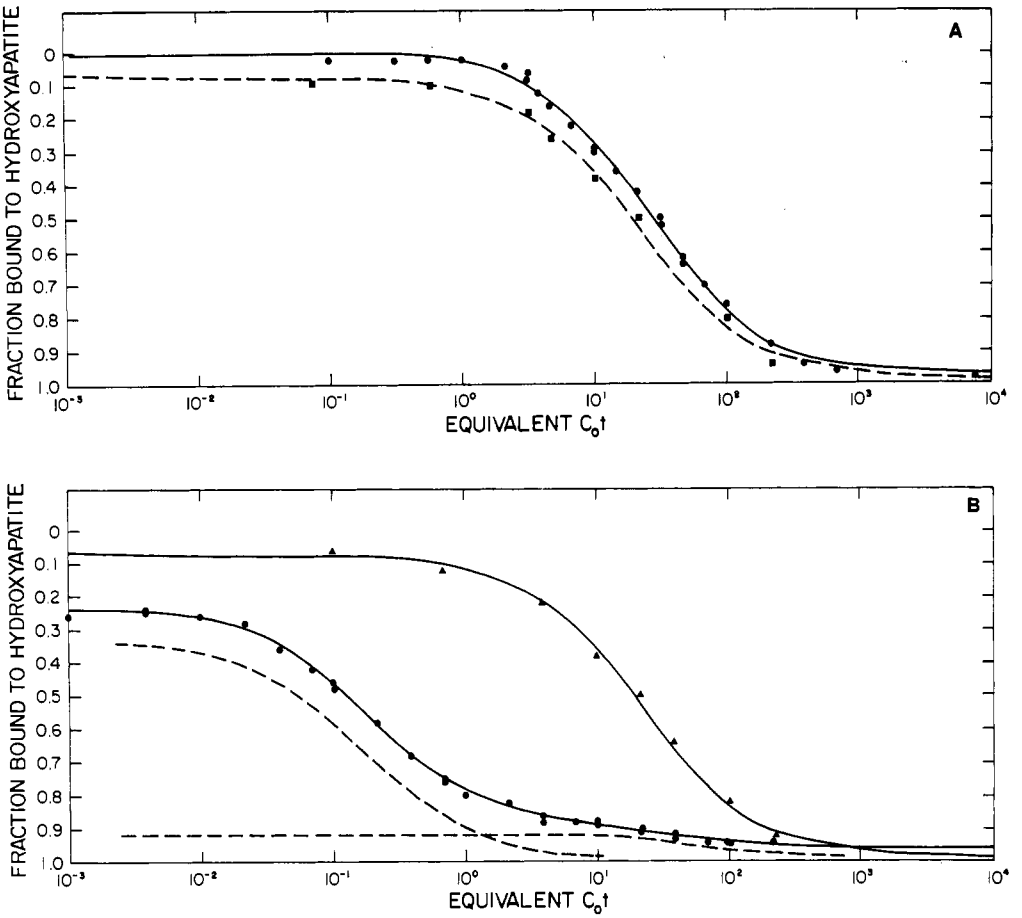

FIGURE 2: Reassociation of isolated repetitive and single-copy fractions of the *Neurospora* genome. (A) "Slave mini-C_0t" reassociation of a single-copy tracer. A single-copy component was isolated from 400-Nt ^{32}P-labeled DNA as the fraction not bound to HAP (67%) after reassociation to a C_0t of 15.0 M s. The 400-Nt ^{32}P-labeled single-copy fraction (●) was reassociated with a large excess of unlabeled 400-Nt total DNA fragment (■). The dashed line portrays the reassociation of the unlabeled DNA driver in the reaction (k = 0.05 M^{-1} s^{-1}). The solid curve represents the best least-squares solution for the reassociation of the single-copy tracer, indicating a single second-order component, single-copy DNA (98%) with a k in whole DNA of 0.039 M^{-1} s^{-1}. The root mean square error of the fit was 2.03%. (B) Reassociation of the repetitive component. An enriched repetitive tracer was isolated as the fraction of 400-Nt ^{32}P-labeled total DNA bound to HAP (9.9%) after reassociation to a C_0t of 0.75 M s. Under these conditions 80% of the repetitive sequences and 3% of the single-copy DNA form duplexes. The ^{32}P-labeled repetitive tracer (●) was reassociated with a large excess of unlabeled 400-Nt total DNA fragments. The unlabeled total DNA driver in the samples reassociated with a k of 0.051 M^{-1} s^{-1} (▲). The best least-squares solution (rms = 1.09%) shows that the repetitive tracer reassociated as two second-order components: a repetitive component (67%) with a k of 5.58 M^{-1} s^{-1} and a small amount of contaminating single-copy DNA (8.5%) with a k of 0.1 M^{-1} s^{-1}. The remaining 25% of the tracer was foldback DNA and bound to HAP at the earliest C_0t values. The single-copy and repetitive rate constants from these reactions were used as fixed values for the analyses of total DNA reassociation in parts B and C of Table I.

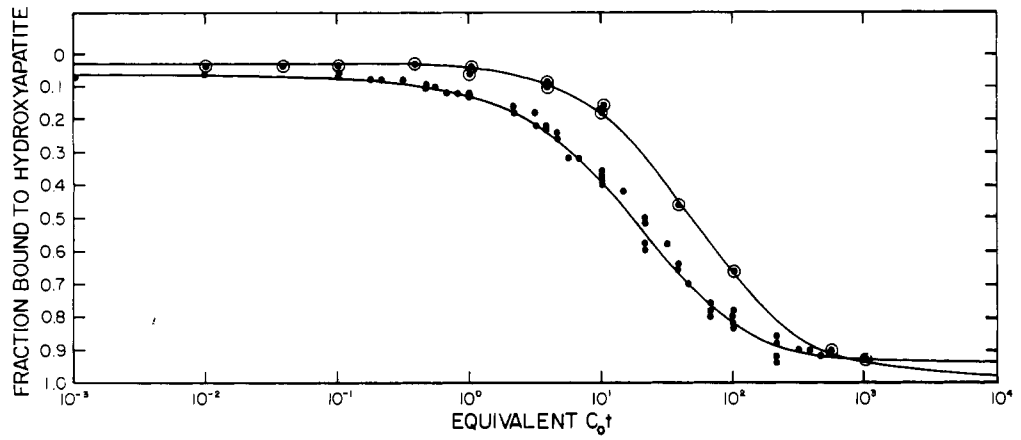

FIGURE 3: Reassociation of 400-Nt *Neurospora* DNA at two criteria. Total 400-Nt ^{32}P-labeled *Neurospora* DNA fragments were reassociated at either 60 °C in 0.12 M PB (●) or 50 °C in 0.12 M PB (⊙). The samples were analyzed at the same respective temperatures on hydroxylapatite. The solid lines represent the best least-squares computer fit to the data. The DNA incubated at 60 °C reassociated as a single component with a k of 0.05 M^{-1} s^{-1} (rms = 1.6%). The samples analyzed at 50 °C reassociated as a single component with a k of 0.019 M^{-1} s^{-1} (rms = 2.4%).

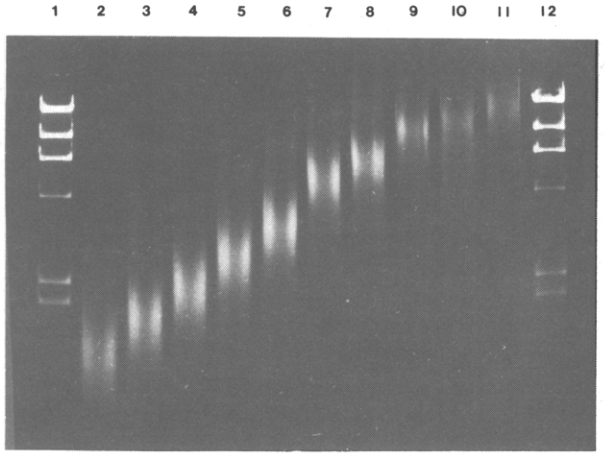

FIGURE 4: Electrophoresis of various-sized *Neurospora* DNA. The size and discrete-size profile of a range of fragments isolated as described under Materials and Methods was examined by electrophoresis. DNA samples were electrophoresed in 50 mM Tris (pH 7.9), 5 mM sodium acetate, and 1 mM EDTA for 12 h in 1% agarose at 25 V. The gel was stained in ethidium bromide (0.5 μg/mL) for 20 min, and the DNA was visualized under ultraviolet light. Lanes 1 and 12 are *Hin*dIII digests of λDNA. Lanes 2–10 represent DNA fragments of 1500, 1900, 2400, 3000, 3600, 5300, 6400, 7500, 8700, and 10 200 Nbp, respectively. The size of the fragments represented by these bands agrees with the sizes determined by sedimentation through shallow alkaline sucrose gradients using DNA markers.

sequences was unstable and not detected under these conditions. Therefore, we compared the reassociation of the same labeled DNA fragments under a less stringent condition. Figure 3 shows that DNA reassociated at the lower criterion, 50 °C and 0.12 M PB, reanneals with a k of 0.019 M^{-1} s^{-1} compared to a k of 0.05 M^{-1} s^{-1} for the original criterion. The lower rate is a consequence of reassociation of the DNA nearly 43 °C below the melting temperature, which falls below the optimum temperature plateau (Britten et al., 1974). However, no additional repetitive component was detected at this lower criterion, where 18-Nbp duplexes should be stable. We conclude that our original conditions were satisfactory to detect all the repetitive DNA. It should be noted, however, that the ability to see small repetitive duplexes may be more dependent on the size of the duplexes that will bind to HAP than on the stringency of the criteria of reassociation.

Reassociation of Long Neurospora DNA Fragments. The reassociation of short vs. long DNA fragments can be used to determine the extent of interspersion of repetitive and unique

Table II: Reassociation Rate Constants for Various DNA Fragment Lengths

fragment length (Nt)	rates from computer fit (M^{-1} s^{-1})	equivalent rates, k_{400} (M^{-1} s^{-1})[a]
400	0.051	0.051
1500	0.098	0.050
3000	0.120	0.049
6400	0.204	0.051
10200	0.252	0.049

[a] The equivalent rates were calculated by using a correction for length of $k_{400} = k_L(400/L)^{1/2}$ according to Wetmur & Davidson (1968).

sequences (Davidson et al., 1973). A series of ^{32}P-labeled DNA fragments of discrete sizes ranging from 200 to 11 000 Nt in length were isolated and characterized as shown in Figure 4. The results indicate that the fragments do have a narrow size distribution and discrete length. The self-reassociation of DNA of four fragment sizes (1500, 3000, 6400, and 10 200 Nt) is shown in Figure 5. If a significant fraction of the repetitive sequences were interspersed with single-copy DNA, then the amount of apparent duplex bound to HAP at repetitive C_0t values should increase with fragment length. This is due to the presence of single-stranded unique DNA linked to short repetitive duplexes. The rate of reassociation of the *Neurospora* DNA does increase with fragment length. However, this increase is not due to interspersed sequences but results from the expected rate enhancement due to fragment length (Wetmur & Davidson, 1968). Table II lists the best least-squares solutions for the rate constants for reassociation of DNAs of different fragment lengths as shown in Figure 5. When the rates are appropriately corrected for the increase as a function of fragment length, it is obvious that all fragments reassociated with nearly identical kinetics. These results strongly suggest that the repetitive sequences in *Neurospora* are organized in stretches at least 10 200 Nbp in length and are not significantly linked to single-copy DNA at shorter fragment lengths.

Interspersion and Spacing of Repetitive DNA. The organization of repetitive sequences in the genome was studied by another type of experiment. The fraction of total DNA containing interspersed repetitive sequences as a function of DNA fragment length can be analyzed by reassociating tracers of various lengths to a repetitive C_0t with a large excess of unlabeled short driver DNA (Davidson et al., 1973; Crain et

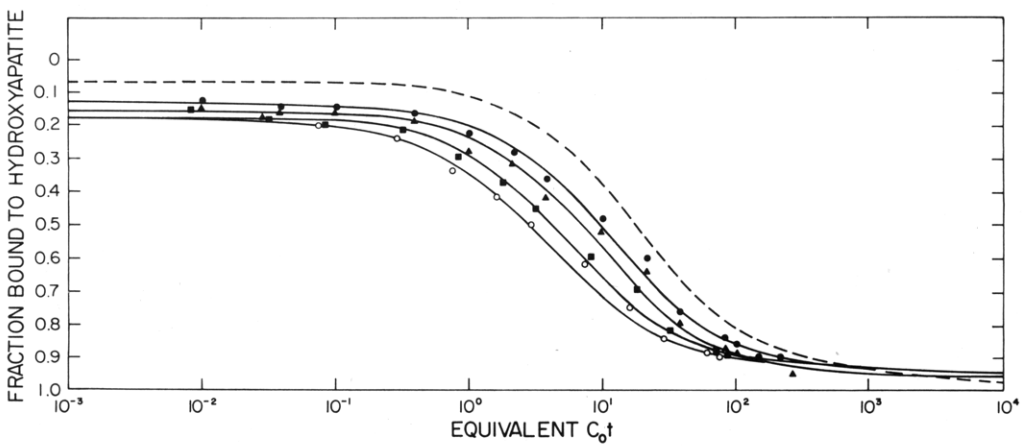

FIGURE 5: Reassociation kinetics of long *Neurospora* DNA fragments. Four fragment sizes of ^{32}P-labeled total *Neurospora* DNA were reassociated and analyzed on hydroxylapatite under conditions previously described under Materials and Methods. The solid lines through the data represent the least-squares fit for the 1500- (●), 3000- (▲), 6400- (■), and 10 200-Nt (○) fragments. The rate constants for the computer fit are shown in Table II. The dashed line represents the reassociation of 400-Nt fragments.

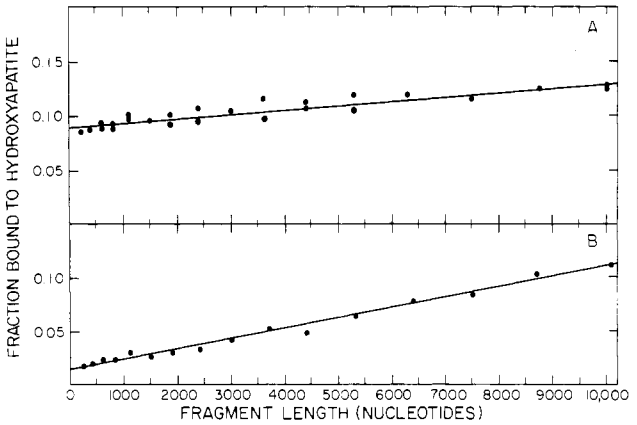

FIGURE 6: The fraction of *Neurospora* DNA containing repetitive sequences as a function of increasing fragment length. Labeled DNA fragments from 200 to 10200 Nt long were reassociated to a repetitive C_0t of 0.75 M s with a 10000-fold excess of unlabeled 400-Nt *Neurospora* DNA fragments. The fraction of DNA containing duplexes (F) was determined by fractionation on hydroxylapatite. The values plotted in (A) represent the fraction bound to HAP at a C_0t of 0.75 M s corrected for the contribution of zero-time binding DNA (foldback). The equation $F = (B - Z)/(1 - Z)$, where F is the fraction of DNA containing repetitive sequences, B is the total fraction of DNA bound at C_0t = 0.75 M s, and Z is the fraction of DNA binding at zero time, was used to calculate the data according to Davidson et al. (1973). The solid line represents the best linear least-squares fit to the data (A) and has a correlation coefficient of 0.89. The y intercept of the solution is 0.0875. The values for zero-time binding of DNA (B) were measured by denaturating the samples and immediately separating the duplex fraction on HAP. The correction for zero-time binding was determined from a linear least-squares fit to the data in (B). The fit has a correlation coefficient of 0.992 and a y intercept of 0.0175.

al., 1976). This also provides an estimation of the intervals at which single-copy and repetitive DNA are linked. A range of discrete-sized ^{32}P-labeled *Neurospora* DNA tracers was reassociated with a 10000-fold excess of unlabeled 400-Nt total DNA to a C_0t of 0.75 M s. The zero-time binding of duplex DNA on HAP observed as a function of fragment length is shown in Figure 6B. These duplex regions arise due to the formation of intrastrand duplexes in fragments containing inverted repeats (foldback DNA). The increased amounts of DNA which was bound to HAP as a function of increasing fragment length indicate that foldback sequences are interspersed throughout at least 12% of the genome in fragments of 10000 Nt. The y intercept of this curve represents the fraction of the genome comprised of foldback sequences and is 2%, which agrees with the values of 2–3% calculated by other methods (see Table I). The fraction of the DNA reassociated to the repetitive C_0t which bound to HAP, corrected for zero-time binding, is shown in Figure 6A. The value of the y intercept in this curve is a measure of the total amount of repetitive DNA in the genome. This value, 8.75%, for repetitive DNA is in good agreement with the value (8%) determined from both the data on isolation and reassociation of the repetitive tracer and that for total DNA reassociation (Table I).

The results show that even at the largest tracer lengths employed, only 12.5% of the DNA contains repetitive sequences, which represents only an increase of 3.75% over the actual amount of repetitive DNA. This increase, of about 4%, could be due to a small amount of single-copy sequences linked to repetitive DNA. However, we have calculated that because of the effect of fragment length on reassociation rates in short driver–long tracer experiments (Wetmur, 1971), this small increase represents single-copy sequences which have reannealed because of the accelerated rate of renaturation. No transition can be seen in the curve of Figure 6A in the 1000–2000 Nt length region, which is a characteristic of short-period interspersion (Pearson et al., 1978). Therefore, these results indicate that at most only a small fraction (4%) of the single-copy sequences could be linked to repetitive DNA at fragment lengths up to 10200 Nbp. The repetitive DNA must be organized in very long clusters and is not generally interspersed throughout the genome. However, the foldback sequences are interspersed in about 12% of the *Neurospora* genome.

Repeat-Contiguous Single-Copy DNA. The results of the final and most direct test for the amount of single-copy sequences which are linked to repetitive DNA are shown in Figure 7. Long ^{32}P-labeled DNA fragments (6400 Nt) were stripped of foldback sequences and then reassociated to a C_0t of 0.2 M s which represents an equivalent C_0t of 0.75 M s for 400-Nt fragments. The repetitive duplexes were isolated on HAP, and the amount bound represented 6% of the total labeled DNA and 81% of the repetitive DNA. These fragments were then mixed with a large excess of unlabeled total

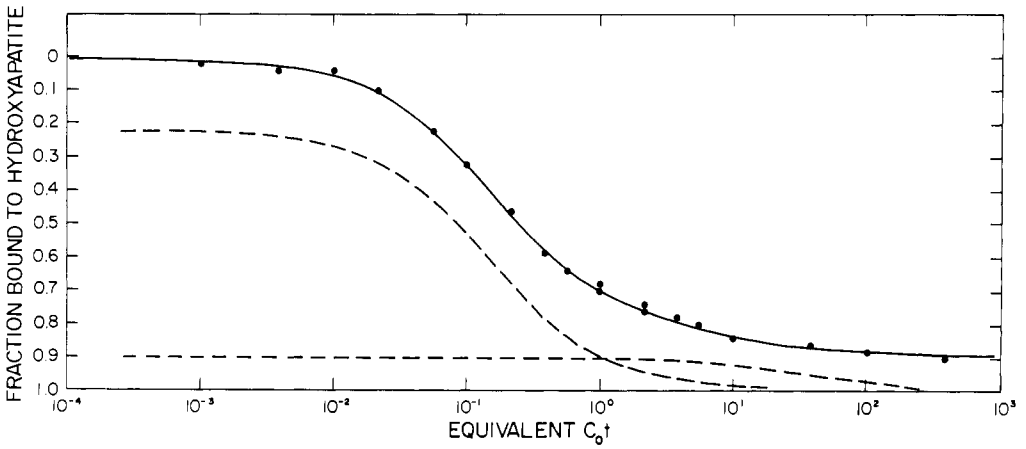

FIGURE 7: The amount of single-copy DNA adjacent to repetitive sequences. Total ^{32}P-labeled DNA fragments 6400 Nt long were denatured, and the foldback sequences were removed by passage over HAP. The fragments were then reassociated to a C_0t of 0.2 M s, and the fraction containing repetitive duplexes (6%) was isolated on HAP. The ^{32}P-labeled repetitive fraction was mixed with a 4000-fold excess of unlabeled 3200-Nbp total DNA fragments and sheared to a length of 425 Nt. This preparation was then reassociated and analyzed on hydroxylapatite. The solid curve represents the best least-squares solution for the reassociation of the labeled repetitive fraction (●). The dashed lines portray the two second-order kinetic components from this solution: a repetitive component (80%) with a k of 6.4 M^{-1} s^{-1} and a small single-copy component (10%) with a k of 0.06 M^{-1} s^{-1} (rms = 0.5%).

DNA, sheared to 425 Nt, and reassociated again to determine the amounts of repetitive and single-copy DNA. This technique is sensitive enough to detect the linkage of repetitive DNA with only a small percentage of single-copy sequences (Crain et al., 1976). These results provide an upper limit for the extent of such linkage, since at least some unique sequences could be expected to be present as a contaminant in the long repetitive tracer. The best least-squares fit for the data of Figure 7 indicated the presence of two components in the labeled repetitive fragments. The major fraction (80%) is repetitive DNA with a k of 6.4 M^{-1} s^{-1}, whereas only a small fraction (10%) is unique DNA with a k of 0.06 M^{-1} s^{-1}. Therefore, the maximum amount of single-copy DNA contiguous with repetitive sequences at a fragment length of 6400 Nt is 0.7% (10% × 0.06/0.81) of the total ^{32}P-labeled DNA and 0.8% of single-copy DNA. These results clearly show that the repetitive sequences must be organized in very long stretches uninterrupted by single-copy DNA in the *Neurospora* genome.

Discussion

In this paper the complexity and sequence components of the *N. crassa* genome are detailed. We have utilized the reassociation kinetics of total nuclear DNA and isolated kinetic components under carefully defined conditions to obtain an accurate determination of the genome composition. The *Neurospora* genome is simple and consists of foldback (2%), repetitive (8%), and single-copy (90%) sequences. The size of the genome is 2.7×10^7 Nbp, which is approximately 6.3 times larger than that of *E. coli*. The kinetic complexity of the single-copy component corresponds to a coding capacity for 18 000 average-sized structural genes. Studies examining the mRNA content in vegetative cells indicate that under these conditions about 10% of the single-copy DNA in *Neurospora* is transcribed, representing the expression of 2000 unique sequences (unpublished experiments).

Several different and sensitive experimental techniques were used to examine the sequence arrangement in *Neurospora* and conclusively showed that repetitive and single-copy DNA are not organized in a short-period interspersion pattern. The data from the reassociation of long DNA fragments indicate the repetitive sequences occur in stretches at least 10000 Nbp long. A second approach, in which various sizes of labeled DNA tracers were reassociated with a short driver, measured the interspersion interval and the fraction of the DNA containing repetitive sequences. This experiment clearly revealed that there is no regular interval, large or small, at which the repetitive sequences are interspersed with single-copy DNA. These results also imply that little, if any, unique DNA is adjacent to repetitive DNA even in fragments up to 10 000 Nbp. In contrast, the foldback sequences did show interspersion throughout 12% of the genome when examined in long fragments. A quantitative determination of the extent to which single-copy sequences are contiguous with repetitive DNA revealed that there is almost no detectable amount of interspersion in the genome. In fragments 6400 Nbp in length, a maximum estimate of 0.7% of the single-copy DNA is adjacent to repeated DNA. This is in direct contrast to organisms with short-period interspersion patterns, in which 50–80% of their unique DNA is interspersed with repetitive sequences (Davidson et al., 1975a; Goldberg et al., 1975). The conclusion derived from these experiments is not merely that *Neurospora* has a long-period interspersion pattern instead of the more common short pattern but rather that no significant interspersion occurs at all. Thus, the sequence organizaton of *Neurospora* does not comply with either of the two general patterns observed in most other eucaryotes (Crain et al., 1976).

The determination of the length of the repetitive sequences in experiments was limited by the size of the DNA fragments we could conveniently handle. Nevertheless, our results suggest that the repetitive DNA occurs in clusters much longer than 10 000 Nbp. A method for estimating the lengths of repeated sequences in the genome from reassociation data was derived by Manning et al. (1975) and Crain et al. (1976). Their derivation assumes that if all the repetitive DNA is of a uniform length, then the ratio of the amount of repeat-contiguous single-copy DNA genome (Y) to the total amount of repetitive DNA (F_r) is a function of the fragment length being analyzed (L) and the length of the repetitive sequences in the genome (R). These parameters are related by the expression $Y/F_r = L/R$, which was applied to the data obtained from our quantitative analysis of the repeat-contiguous DNA (calculated from the experiment shown in Figure 7). The maximum linkage of single-copy DNA to repetitive DNA (Y) is 0.7% at a fragment length (L) of 6400 Nbp, and the amount of repetitive DNA in the genome (F_r) is 8%. Therefore, we calculate that the repetitive sequences must have an average minimum genomic length of at least 73 000 Nbp, which implies that there could only be a maximum of about 30 clusters of repeated sequences interspersed in the *Neurospora* genome. In fact, we suspect that the repetitve sequences are even longer, possibly organized as only a single cluster.

The complexity of the repetitive component, 15 300 Nbp, is surprisingly low. Considering the extremely long stretches of repetitive DNA implied by our data and that the reiteration frequency of this component is 140, it seems likely that a small number of relatively simple repetitive sequences must be tandemly repeated a considerable number of times. This consideration immediately suggests that these sequences might represent the ribosomal genes, which includes both spacer and rRNA coding sequences. In yeast the multiple copies of ribosomal DNA are believed to be on a single chromosome (Petes & Botstein, 1977). A recent study in *Aspergillus* suggests that ribosomal sequences may comprise the major fraction of the repetitive DNA (Timberlake, 1978). If this were the case in *Neurospora*, the organization of the ribosomal genes could account for the large stretches of repetitive DNA and lack of interspersed sequences. The complexity of the major ribosomal DNA unit in *Neurospora* has been shown to be 9700 Nbp (Free et al., 1979). Therefore, if the ribosomal genes are repeated about 140 times, as suggested by the reiteration frequency of the repetitive component, they could easily represent 5–6% of the total DNA. We are presently examining this possibility.

Regardless of the exact nature of the repetitive sequences, the conclusion that they occur in very long stretches, at least 73 000 Nbp, and are interspersed very few times in the genome precludes any major regulatory role for these sequences. We simply were unable to observe any significant amount of short interspersed repetitive sequences, even under conditions that should have detected stable duplexes greater than 18 Nbp. The results indicate that the interspersed middle repetitive sequences detected in DNA reassociation experiments, proposed to act as recognition sites for the coordinate regulation of gene sets in higher eucaryotes (Davidson et al., 1973, 1977), cannot account for coordinate control of gene expression in *Neurospora*. It is not yet clear to what extent the mechanisms of genetic regulation are the same in higher and lower eucaryotes. The differences in the amount and organization of repetitive sequences between *Neurospora* and more advanced

systems do not necessarily imply that they possess different molecular mechanisms for controlling gene expression, since it has not yet been fully established that the interspersed repetitive DNA plays a role in gene regulation in higher eucaryotes. At present it is only possible to speculate on the selective value and function of these sequences.

Acknowledgments

We thank Dr. William Timberlake for helpful discussions and Dr. Robert Ross and Sid Hendrickson for help in setting up our computer programs.

References

Britten, R. J., Graham, D. E., & Neufeld, B. R. (1974) *Methods Enzymol. 29E*, 363.

Crain, W. R., Eden, F. C., Pearson, W. R., Davidson, E. H., & Britten, R. J. (1976) *Chromosoma 56*, 309.

Davidson, E. H., & Britten, R. J. (1973) *Q. Rev. Biol. 49*, 555.

Davidson, E. H., Hough, B. R., Amenson, C. S., & Britten, R. J. (1973) *J. Mol. Biol. 77*, 1.

Davidson, E. H., Galau, G. A., Angerer, R. C., & Britten, R. J. (1975a) *Chromosoma 51*, 253.

Davidson, E. H., Hough, B. R., Klein, W. R., & Britten, R. J. (1975b) *Cell 4*, 217.

Davidson, E. H., Klein, W. H., & Britten, R. J. (1977) *Dev. Biol. 55*, 69.

Free, S. J., Rice, P. W., & Metzenberg, R. L. (1979) *J. Bacteriol. 137*, 1219.

Goldberg, R. B., Crain, W. R., Ruderman, J. V., Moore, G. P., Barnett, T. R., Higgins, R. C., Gelfand, R. A., Galau, G. A., Britten, R. J., & Davidson, E. H. (1975) *Chromosoma 51*, 225.

Graham, D. E., Neufeld, B. R., Davidson, E. H., & Britten, R. J. (1974) *Cell 1*, 127.

Hautala, J. A., Conner, B. H., Jacobson, J. W., Patel, G. L., & Giles, N. H. (1977) *J. Bacteriol. 130*, 704.

Hell, A., Birnie, G. D., Slimming, T. K., & Paul, J. (1972) *Anal. Biochem. 48*, 369.

Hudspeth, M. S. E., Timberlake, W. E., & Goldberg, R. B. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 4332.

Jacobson, J. W., Hautala, J. A., Lucas, M. C., Reinert, W. R., Stroman, P., Barea, J. L., Pately, V. B., Case, M. E., & Giles, N. H. (1977) *ICN–UCLA Symp. Mol. Cell. Biol. 6*, 269.

Maniatis, T., Jeffery, A., & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 1184.

Manning, J. E., Schmid, C. W., & Davidson, N. (1975) *Cell 4*, 141.

Marmur, J. (1961) *J. Mol. Biol. 38*, 208.

Marzluf, G. A. (1977) in *Regulatory Biology* (Marzluf, G. A., & Copeland, J., Eds.) p 196, Ohio State University Press, Columbus, OH.

Metzenberg, R. L., & Nelson, R. E. (1977) *ICN–UCLA Symp. Mol. Cell. Biol. 6*, 253.

Pearson, W. R., Davidson, E. H., & Britten, R. J. (1977) *Nucleic Acids Res. 4*, 1727.

Pearson, W. R., Wu, J. R., & Bonner, J. (1978) *Biochemistry 17*, 51.

Petes, T. D., & Botstein, D. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5091.

Reinert, W. R., & Marzluf, G. A. (1975) *Mol. Gen. Genet. 139*, 39.

Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237.

Smith, M. J., Hough, B. R., Chamberlin, M. E., & Davidson, E. H. (1974) *J. Mol. Biol. 85*, 103.

Studier, F. W. (1965) *J. Mol. Biol. 11*, 373.

Timberlake, W. E. (1978) *Science 202*, 973.

Wells, R., Royer, H. D., & Hollenberg, C. P. (1976) *Mol. Gen. Genet. 147*, 45.

Wetmur, J. G. (1971) *Biopolymers 11*, 34.

Wetmur, J. G., & Davidson, N. (1968) *J. Mol. Biol. 31*, 349.