MOLECULAR CHARACTERIZATION OF THE DROSOPHILA GENOME[1]

CHARLES D. LAIRD AND BRIAN J. McCARTHY

Department of Zoology, The University of Texas at Austin
and
Departments of Genetics and Microbiology, University of Washington, Seattle 98105

Received February 14, 1969

A remarkable feature of evolution in the genus Drosophila is the conservation of the basic karyotype. Postulation of Robertsonian centric fusions (ROBERTSON 1916) is sufficient to explain most of the karyotypic rearrangements found in the various species. The fixation of large duplications of genetic material during the evolution of this genus appears to have been rare (STONE 1955). On the other hand, amino acid sequence data for proteins of other organisms indicate that many cistrons have very similar nucleotide sequences, and are presumably related by prior gene duplication (JUKES 1966). Physical studies of DNA of various eucaryotes also suggest that these genomes contain very large families of related cistrons (BRITTEN and KOHNE 1968). The size of such families is often in excess of that inferred from existing protein data (BRITTEN and KOHNE 1968; McCARTHY 1967; WETMUR and DAVIDSON 1968). This may in some cases reflect the presence of cistrons with similar nucleotide sequences which specify proteins of overtly different functions. It is also likely that some of these families of genes have functions other than the determination of amino acid sequences.

In keeping with the cytological studies, the reassociation kinetics of denatured *Drosophila melanogaster* DNA imply that the genome of this organism is much less complex, in terms of intragenome nucleotide sequence homologies, than those of vertebrates (LAIRD and McCARTHY 1968b). Estimates made from these measurements suggest that perhaps 90% of *D. melanogaster* nucleotide sequences are not closely related to others elsewhere in the genome.

In light of the extensive diversity within the genus Drosophila (LAIRD and McCARTHY 1968a), a detailed analysis of the intragenome homologies of representative species is expected to provide information about the evolution of genomes. In particular, our results indicate that genome simplicity extends to other Drosophila species and to a representative of another Dipteran family. In addition, there are variations among Diptera in genome size as inferred from the informational content of nucleotide sequences.

MATERIALS AND METHODS

Details of labeling and isolation of DNA and RNA from Drosophila, preparation of filter-bound DNA, hybridization and duplex formation procedures, and analytical CsCl pycnography

[1] Supported by NSF grant No. GB6099, NIH postdoctoral funds and NSF grant GU-1598.

have been previously described (LAIRD and MCCARTHY 1968a). Additional and modified procedures are listed below.

A) *Preparation of isotopically-labeled DNA:* ^{32}P-labeled T2 DNA—*Escherichia coli* (2×10^8 cells/ml), growing in low phosphate minimal medium containing glucose, were infected with T2 (strain H, kindly provided by Dr. R. CROUCH) at a multiplicity of 0.1. Five minutes after infection, 10 μc/ml ^{32}P (as Na_2HPO_4) was added, and aeration was continued at 37°C for 4 hr. After removal of debris from the lysate by low-speed centrifugation, T2 bacteriophage and DNA were purified by the method of FREIFELDER (1965).

^{3}H-labeled bacterial DNA's—*E. coli* B and *Bacillus subtilis* 168 were grown in Davis minimal medium containing glucose and 5 μc/ml thymidine-methyl-^{3}H (17 c/m mole, New England Nuclear). Bacteria were collected during exponential growth phase and the labeled DNA was isolated by the method of MARMUR (1961).

B) *Unlabeled DNA's:* T2 DNA was isolated from purified phage by the $NaClO_4$ method of FREIFELDER (1965). Bacterial DNA's from *E. coli* B, *Bacillus subtilis* 168, *B. globigii* (Department of Genetics, Stanford University) and *Cytophaga succinicans* were purified as described by MARMUR (1961). A variety of procedures was used to isolate DNA from insects. The procedure detailed previously (LAIRD and MCCARTHY 1968a), involving a crude preparation of cell nuclei, was used unless otherwise specified. Nuclei of various extents of purity were also used in order to examine the relative contributions of nuclear and cytoplasmic DNA's to the total DNA. A crude nuclear pellet from pupae or adults was obtained by homogenization in Tris buffer (0.05 M, pH 7.6) containing KCl (0.025 M), magnesium acetate (0.005 M) and sucrose (0.35 M), using a mortar and pestle. The homogenate was filtered through cheese cloth, and the filtrate was further homogenized with a motor-driven teflon pestle in a glass tube. Nuclei were pelleted by centrifugation at $500 \times g$ for 10 min. DNA was prepared from crude nuclear pellets from *D. melanogaster* (Oregon-R) pupae and adults, *D. funebris* (Austin, Texas) pupae, and fleshfly (*Sarcophaga bullata*) pupae obtained from Carolina Biological Supply House. Nuclei from *D. melanogaster* adults were also purified by centrifugation through 2.0 M sucrose containing 0.003 M $CaCl_2$ and 0.01 M Tris, pH 7.4 and then used as a source of nuclear DNA.

C) *Renaturation of DNA in solution:* Denaturation and renaturation of DNA were studied using optical methods. The change in absorbance at 260 mμ which accompanies denaturation of DNA (hyperchromicity) was measured with an automatic recording Gilford-Beckman spectrophotometer. Rates of reassociation of denatured DNA were measured in $1 \times$ SSC (SSC is 0.15 M NaCl, 0.015 M trisodium citrate) with DNA sheared at about 680 atmospheres (10,000 p.s.i.) yielding a single-strand molecular weight of about 400,000 daltons. Renaturation at low DNA concentrations (< 50 μg/ml) was followed in 1 cm light-path cuvettes. To determine DNA concentrations and purity by the total hyperchromic shift during denaturation, OD_{260} was followed as the temperature was raised in the cuvettes from 60° to 100°C. These solutions were rapidly cooled to the optimum renaturation temperature of 25° below the midpoint of the thermal transition (T_m, MARMUR and DOTY 1961); the optical density was monitored continuously. For highly purified DNA's, the percent hyperchromicity was 38%; this value dropped to 33% during the cooling to renaturation temperatures as the single strands assumed a less extended configuration. Renaturation of DNA at higher concentrations was carried out under mineral oil in 1 mm light-path cuvettes with the continuously recording spectrophotometer. In this case, DNA's were denatured at 100°C for 10 min, rapidly cooled, and brought up to renaturation temperature. Hyperchromic values during denaturation of these same DNA solutions were determined in 1 cm light-path cuvettes. In some experiments, renaturation at even higher concentrations (> 2 mg/ml) was followed measuring at 62°C the absorbance of aliquots diluted into $1 \times$ SSC.

RESULTS

A) *Calibration of genome size with initial rate of DNA/DNA duplex formation:* The reaction of labeled, denatured DNA with filter-bound DNA permits quantitation of initial rates of reaction, whereas solution reactions measure more

accurately reassociation rates of the total DNA (McCarthy 1967; Britten and Kohne 1968). Measurements using filter-bound DNA to determine initial rates of duplex formation indicate that DNA from organisms with small genomes, such as bacteriophage, reassociates more rapidly than DNA from organisms with larger genomes, for example, bacteria (McCarthy 1967; Laird and McCarthy 1968b). Immobilization of one of the interacting components complicates theoretical analysis of this expected dependence of reassociation rate on sequence diversity and necessitates an empirical correlation. We have attempted to determine this correlation by mixing DNA's from unrelated organisms to provide an increased heterogeneity in nucleotide sequences. This increased heterogeneity should simulate the sequence diversity in DNA derived from an organism with a larger genome.

One experiment reported here involved dilution of T2 bacteriophage DNA (both labeled and filter-bound) with bacterial DNA from *Escherichia coli*. With increasing proportions of bacterial DNA, the labeled T2 sequences react with a smaller and smaller proportion of the total DNA. This effect is equivalent to studying reactions of a few genes in an increasingly larger genome. Six different mixtures of T2 and *E. coli* DNA's were used, varying from 100% T2 to 100% *E. coli* DNA, with the most pertinent ratio being 1 T2:20 *E. coli*. Reaction solutions contained ^{32}P-labeled T2 and/or ^{3}H-labeled *E. coli* DNA's at a total concentration of 5 μg/ml in 0.2 ml of 1 × SSC. To each of a series of such vials, a filter containing 12 μg denatured DNA was added. The ratio of T2 to *E. coli* DNA was adjusted during filter preparation to be identical to that in the solution DNA. Reactions were followed for up to 90 min at 60°C to determine initial rates of T2 and *E. coli* duplex formation. The results of measuring the rate of association of ^{32}P-labeled, sheared, denatured T2 DNA with filter-bound T2 DNA in the presence of increasing proportions of *E. coli* DNA are shown in Figure 1A. The upper abscissa indicates the percent of T2 DNA in the reaction mixtures and on the filters, while the lower abscissa indicates the genome molecular weight, taking 1.3×10^8 daltons to be the genome size of T2 (Rubenstein *et al.* 1961). The ordinate shows the observed initial rate of reaction (percent labeled DNA bound per hr) of the ^{32}P-T2 DNA (open and closed circles). For example, with no *E. coli* DNA present, T2 DNA reacted at nine percent per hour. When 10% of the DNA was T2, (and 90% was *E. coli* DNA), the rate was about four percent per hour.

At this 1:10 ratio of T2 : *E. coli* DNA, the solution can be thought of as containing DNA from a hypothetical organism, 10% of whose genes were derived from T2, and 90% from *E. coli*. Thus the minimum genome size in which each T2 sequence is represented once is ten times that of T2, or 1.3×10^9 daltons. By using different proportions of the bacterial and bacteriophage DNA's, we have estimated the expected initial reassociation rates of nucleotide sequences from genomes ranging in size from 1.3×10^8 to 10^{10} daltons (Figure 1A). As a confirmation of the validity of this estimate, we include reaction rate data for two bacterial DNA's (closed symbols, Figure 1A). *B. subtilis* and *E. coli* DNA's reassociated at rates of 2.7 and 2.3% per hr, respectively, leading to genome size

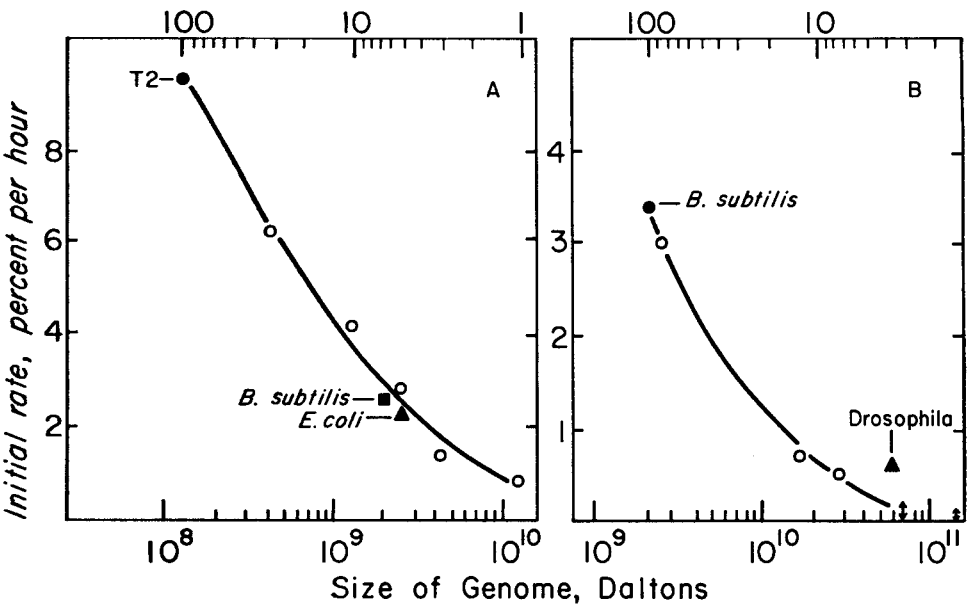

FIGURE 1.—An empirical relationship between genome size and initial rate of duplex formation as measured by the filter method.

A. ^{32}P-labeled T2 DNA (21,000 cpm/μg) at a concentration of 5 μg/ml was incubated with 15 μg filter-bound T2 DNA at 62° C, in 0.2 ml 1 × SSC. At 5, 12, 24, and 45 min, duplicate filters were removed from parallel vials and washed three times in 1 × SSC at 62°C. The filter-bound ^{32}P radioactivity was measured by liquid scintillation spectrometry. The initial rate of this duplex formation (9.6%/hr) is plotted *versus* the molecular weight of the T2 genome, 1.3×10^8 daltons (●). Similarly, rates of reaction were determined for ^{3}H-labeled *E. coli* DNA (4,300 cpm/μg) and *B. subtilis* DNA (12,000 cpm/μg) with homologous filter-bound DNAs (▲, ■); these values are also plotted against the genome sizes of 2.7×10^9 for *E. coli* and 2.0×10^9 for *B. subtilis*.

The open circles (○) represent initial rates of reaction of ^{32}P-labeled T2 DNA with homologous filter-bound DNA. In these cases, however, *E. coli* DNA was mixed with both the solution and filter-bound DNA's. Solution DNA's and filter-bound DNA's were again in concentrations of 5 μg/ml and 15 μg/filter. Each circle represents rate determinations on mixtures with different proportions of the T2 and *E. coli* DNA's; the upper abscissa scale indicates the percentage of T2 DNA in the mixture.

B. A similar experiment was carried out with ^{32}P-labeled *B. subtilis* DNA (41,000 cpm/μg) and ^{3}H-labeled *D. melanogaster* DNA (2,000 cpm/μg). Reaction rates obtained with *B. subtilis* (●) or Drosophila DNA's (▲) are plotted *versus* genome size (data for the molecular weight estimate of the Drosophila genome were taken from Figure 3). The rates of homologous reaction of *B. subtilis* DNA in the presence of Drosophila DNA are indicated by the open circles (○); the upper abscissa indicates the percentage of *B. subtilis* DNA in the mixture.

estimates of 2×10^9 and 2.7×10^9, respectively. These values are similar to estimates based on other techniques for these bacteria (DENNIS and WAKE 1966; CAIRNS 1963).

The ability to measure even more dilute mixtures of T2 DNA was limited by the ^{32}P specific radioactivity, and consequently this particular calibration curve

could not be extended much beyond genomes of 10^{10} daltons. However, more slowly reacting *B. subtilis* DNA was diluted in a similar fashion, and the results of this experiment are shown in Figure 1B. In this case ^{32}P-labeled *B. subtilis* DNA was diluted with *D. melanogaster* DNA, and various ratios of this mixture were reacted with filter-bound DNA containing a mixture of bacterial and Drosophila DNA's which reflected their relative proportions in solution. The actual reaction rate of Drosophila DNA alone, plotted at a haploid genome mol. wt. of 7×10^{10} (this value will be discussed below), falls considerably above the calibration curve at this point. This rate of 0.6% per hr would be expected for DNA from a genome of 2×10^{10} daltons. If all of the Drosophila nucleotide sequences were participating in this initial reassociation, an average multiplicity of 3 for each nucleotide sequence would be indicated. However, only about 10% of reassociated Drosophila DNA has reduced stability (LAIRD and MCCARTHY 1968b). It is this fraction, presumably, that is dominating the initial rapid reaction kinetics (BRITTEN and KOHNE 1968). Thus the observed initial rate of 0.6%/hr represents a rate of 6.0%/hr for the reacting sequences. Figure 1A shows that this rate would be observed for DNA from a genome of about 5×10^8 daltons rather than 7×10^{10}. This suggests that the rapidly reacting components represent sequences belonging to families with about 100 members.

B) *Relationships among rRNA cistrons:*

The multiplicity of rRNA cistrons in Drosophila is 100–200, representing about 0.27% of the DNA (VERMUELEN and ATWOOD 1965; RITOSSA and SPIEGELMAN 1965). The rate of hybridization of rRNA with DNA may be used to determine whether a given rRNA molecule can hybridize with any of these ribosomal RNA cistrons, or whether the sequences are sufficiently different to preclude such interactions. A comparison of this rate for Drosophila nucleic acids with that observed for rRNA/DNA hybrids with nucleic acids from an organism with the same proportion of rRNA cistrons, but fewer in number, would distinguish between these possibilities. *Bacillus subtilis* fulfills these requirements, having 5–10 ribosomal RNA cistrons for each of 16S and 23S molecules per genome (Figure 2; also see YANKOFSKY and SPIEGELMAN 1964). Figure 2A shows that the saturation values for homologous 28S or 23S rRNA with Drosophila and *B. subtilis* DNA, respectively, are similar as expected. The initial rates of hybridization of these homologous nucleic acids (Figure 2B) are 1.2% per hr and 1.7% per hr. On the assumption that bacterial rRNA sequences will cross-react with most of the members of the family of rRNA cistrons (KOHNE 1968), this similarity in rate of reaction would indicate that the majority of rRNA sequences in Drosophila are also closely related. Small numbers of very different sequences would not, however, be detected by these reaction rate measurements.

C) *Renaturation of Drosophila DNA: Optical Measurements.*

Hybridization and duplex reactions with filter-bound DNA are useful in obtaining early reaction kinetics. More complete rate measurements can be made during solution renaturation (MARMUR and DOTY 1961). The rate constants of

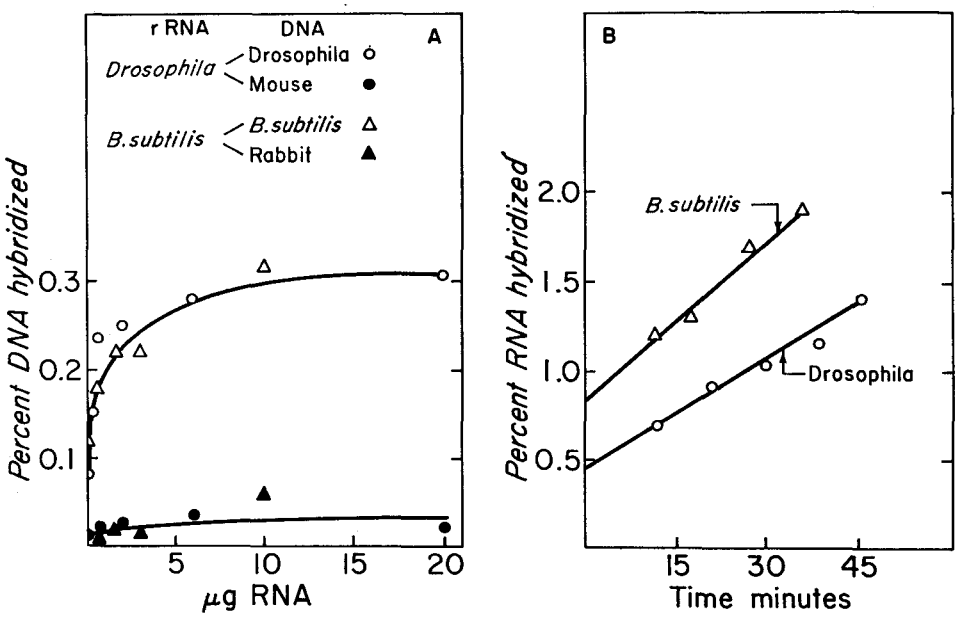

FIGURE 2.—Hybridization of ribosomal RNA.

A. Saturation of homologous and heterologous DNA's with 23S *B. subtilis* rRNA or 28S *D. melanogaster* rRNA.

Increasing amounts of ^{3}H-labeled Drosophila 28S rRNA (1500 cpm/μg) were incubated with 9 μg filter-bound Drosophila DNA or mouse DNA in 0.2 ml 2 × SSC at 60° C for 19 hr. Similarly, ^{32}P-labeled *B. subtilis* 23S rRNA (15,500 cpm/μg) was incubated with 15 μg filter-bound *B. subtilis* or rabbit DNA in 0.2 ml 1 × SSC at 67° C for 18 hr. The filter-bound hybrids were washed twice for 3 min in the buffer and at the temperature of incubation. The percent DNA hybridized was calculated from the amount of labeled RNA bound to the filters.

B. Rate of homologous rRNA hybridization.

Two series of reaction vials containing either 1 μg ^{3}H-labeled 28S Drosophila rRNA (1500 cpm/μg) and 15 μg filter-bound Drosophila DNA, or 1 μg ^{32}P-labeled *B. subtilis* 23S rRNA and 15 μg filter-bound *B. subtilis* DNA, were incubated at 62° C in 0.2 M 1 × SSC. At the indicated times, filters were removed and processed as described in part A. The rate of hybridization was 1.2% input RNA bound per hr for Drosophila and 1.7% per hr for *B. subtilis*.

the renaturation process depend on the genome size and the intragenome homology of an organism (BRITTEN and KOHNE 1968; WETMUR and DAVIDSON 1968). A useful method of describing renaturation kinetics involves plotting percent denatured DNA versus C_0t, the product of initial concentration of denatured DNA (molarity of nucleotides) and time of renaturation (in seconds). For convenience, we summarize the basis of this technique, although the reader is referred to the original work for further details (BRITTEN and KOHNE 1967, 1968). The renaturation of DNA is expected to follow second-order reaction kinetics, inasmuch as the process involves the collision of two (complementary) strands. The concentration of each is potentially rate limiting. The rate of disappearance of denatured DNA then should be:

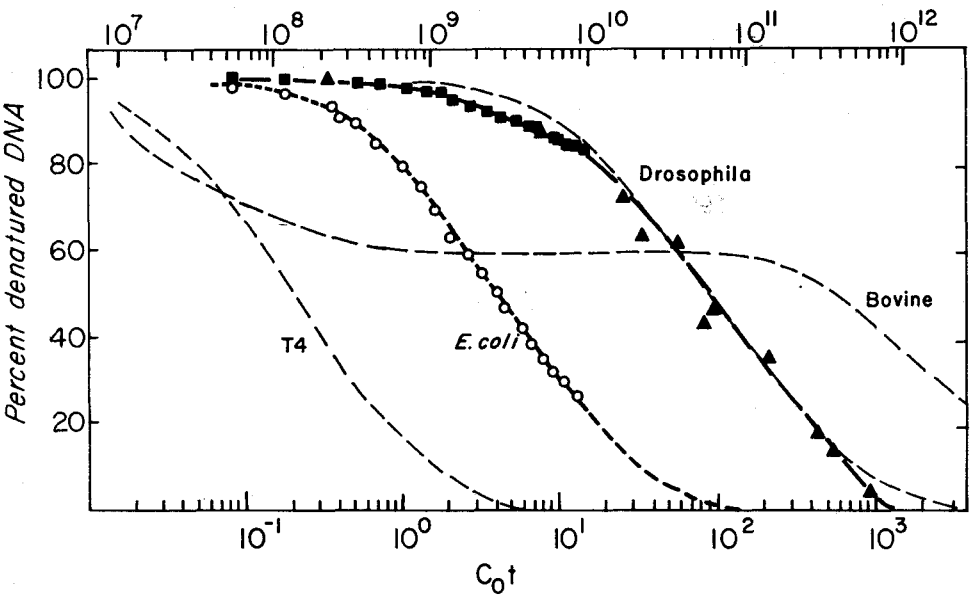

FIGURE 3.—Renaturation of *E. coli* and Drosophila DNA's.

DNA's from *D. melanogaster* and *E. coli* were sheared and heat denatured, and incubated at 60° C in 1 × SSC. The hyperchromicity at 260 mμ was used to measure the extent of renaturation. The percent denatured DNA is plotted *versus* C_0t, the product of concentration (molarity of nucleotides) and time (seconds). *E. coli* DNA (○) and Drosophila DNA (■) were incubated at 60 μg/ml in 1 cm light-path cuvettes. Drosophila DNA at 1.7 mg/ml (▲) was incubated at 62° C in 1 × SSC and diluted 50-fold for OD_{260} readings at 62° C. The dashed lines drawn through the *E. coli* and Drosophila data at 50% renaturation represent the theoretical renaturation curves for simple second-order reactions. The bovine and T4 bacteriophage data, shown for reference as examples of DNA from complex and simple genomes, respectively, are taken from BRITTEN and KOHNE (1968).

$$-dC/dt = k_2C^2$$

where C = concentration of denatured DNA and
t = time of renaturation.

Integrating, and evaluating $t = 0$ $(C = C_0)$,

$$\frac{C}{C_0} = \frac{1}{1 + k_2 (C_0t)} .$$

At half renaturation, $\frac{C}{C_0} = \frac{1}{2} = \frac{1}{1 + k_2 (C_0t)}$ or $C_0t_{1/2} = \frac{1}{k_2}$.

Since k_2 is inversely proportional to sequence diversity (i.e., the number of different cistrons, WETMUR and DAVIDSON 1968; also see BRITTEN and KOHNE 1967, 1968), $C_0t_{1/2}$ is directly proportional to the number of different genes. Genome sizes estimated in this manner correspond to the minimum amount of DNA in which each nucleotide sequence is represented at least once. Actual amounts of DNA per cell will correspond to this estimate when the cell is in fact haploid in a genetic or information sense.

Figure 3 illustrates these points for DNA's from a variety of organisms. Under the conditions of this experiment (1 × SSC, 62° C), *E. coli* DNA is half renatured when $C_0t_{1/2} = 4$. A genome size of 2.7×10^9 daltons (CAIRNS 1963) thus corresponds to a C_0t value of 4, as illustrated on the upper abscissa. The $C_0t_{1/2}$ value for T2 DNA is 0.2, indicative of a genome of 1.3×10^8 daltons (upper scale). This size estimate is in good agreement with more conventional determinations (RUBENSTEIN *et al.* 1961). Bovine DNA, on the other hand, shows a biphasic renaturation, with about 40% renaturing very rapidly and 60% at a rate 400 times more slowly than that of *E. coli* (BRITTEN and KOHNE 1968). Drosophila DNA has a $C_0t_{1/2}$ value of about 80, indicative of a genome size approximately 20 times that of *E. coli*. The dashed lines through the *E. coli* and Drosophila data represent theoretical second-order reaction kinetics obtained by adjusting the rate constant, k_2 such that the curve passes through the midpoint of the actual renaturation data. Such idealized curves are derived from the integral form of the second-order rate equation, $\frac{C}{C_0} = \frac{1}{1 + k_2 (C_0t)}$. In practice, $\frac{C}{C_0}$, the fraction renatured DNA, is solved for various values of C_0t, letting $k_2 = 1$. Lateral transposition may then be obtained by varying k_2, which is inversely proportional to numbers of different nucleotide sequences.

The *E. coli* data follow closely the second-order rate curve (Figure 3). About five to ten percent of the Drosophila DNA, however, renatures somewhat more rapidly than expected for a second-order reaction. These rapidly reacting sequences have a lower thermal stability, after renaturation, than does native DNA (LAIRD and MCCARTHY 1968b) suggesting that these nucleotide sequences are partially repetitious (BRITTEN and KOHNE 1968). Our approach to determine the proportion of DNA which behaves as unique sequences, i.e., sequences which are not represented elsewhere in the genome, is to fit the observed data to the idealized second-order reaction curves. This involves treating the data as belonging to two groups, either unique or repetitious. If the first 8% of the renaturation curve, for example, is considered to represent repetitious DNA, then the remaining 92% can be considered to reflect unique sequences. Data for each class are then renormalized to 100%, and the derived data are plotted in the usual fashion. Figure 4A illustrates this curve-fitting procedure for the Drosophila data in Figure 3. In this case, the midpoint of the second-order rate curve (dashed line) is placed through $C_0t = 100$. Curves a, b, and c are obtained by renormalizing the initially reacting 5, 8, or 15% of the Drosophila DNA to 100%. Curves A, B, and C are derived from renormalizing the remaining 95, 92, or 85% of the data. Curves A and B both provide a good fit of the data to the second-order rate curve, suggesting that 90–95% of the Drosophila DNA renatures with kinetics expected for unique sequences. The placement of the $C_0t_{1/2}$ at 100 indicates that a genome is $100 \div 4$ times that of *E. coli*, or about 7×10^{10} daltons, using CAIRNS' (1963) estimate of 2.7×10^9 daltons for *E. coli*. Two compensatory corrections permit the use of this $C_0t_{1/2}$ value for unique sequences as an estimate of total genome size. The concentration of 90% of the sequences is 0.9 that of the total DNA, thus lowering C_0t values by 0.9. However, such corrected $C_0t_{1/2}$ estimates

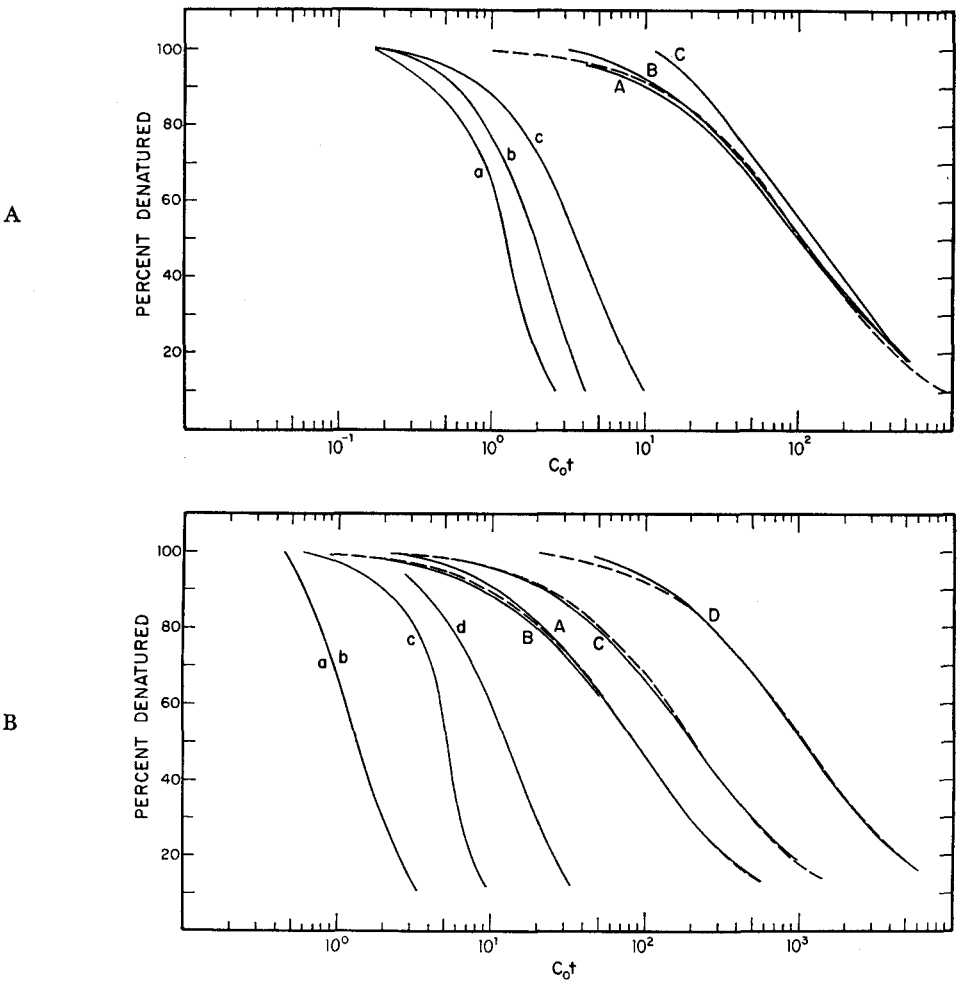

FIGURE 4.—A. Curve-fitting of *Drosophila melanogaster* DNA renaturation kinetics. The dashed line represents second-order kinetics, obtained from the equation $\frac{C}{C_0} = \frac{1}{1 + k_2 C_0 t}$. The lateral placement (i.e., the value of k_2) has been chosen to fit the Drosophila data. To illustrate how different proportions of partially related sequences would affect the kinetics of renaturation of total DNA, three distributions are shown: 5% (a), 8% (b), or 15% (c) repetitious, with the remaining 95% (A), 92% (B), or 85% (C) considered as unique. These percentages of the Drosophila data from Figure 3 are renormalized to 100% and plotted in Figure 4A. Curves A and B fit well the second-order curve (dashed line). The 70- and 50-fold displacements between A and a, and between B and b, indicate the approximate average degree of redundancy (family size) of the repetitious sequences based on the assumption that the initial 5% (a) or 8% (b) reassociation represents reaction among partially related nucleotide sequences.

FIGURE 4.—B. (lower) Curve-fitting of renaturation kinetic data from other insects. Data from Figures 6 and 7 are analyzed as described above except that only one distribution of the data into unique and partially redundant classes is shown for each DNA. Curves A and a illustrate the partitioning of *D. melanogaster* data (Figure 6) into 5% repetitious and 95% unique. The dashed line again indicates an ideal second-order kinetic curve laterally displaced to approximate the 95% curve of Drosophila data (A). Origins of other curves are: B (95%) and b (5%), *D. simulans* data, Figure 6; C (88%) and c (12%), *D. funebris* data, Figure 6; D (92%) and d (8%), *Sarcophaga bullata* data, Figure 7.

would represent only 90% of the genome, thereby necessitating an 11% increase in genome size estimate.

The displacement between curves B and b in Figure 4 indicates the extent to which the partially related sequences are renaturing more rapidly. The ratio of the $C_0t_{1/2}$ values of curves A and a, and B and b, are 50 and 80, respectively, suggesting that the repetitious sequences belong to families with about 50 to 80 members. These values are in reasonably good agreement with the estimate obtained by the filter reactions (Section A).

D) *Characterization of renatured Drosophila DNA:*

An alternative method of analyzing renaturation products depends on the buoyant density difference between native and renatured DNA in CsCl (MESELSON and STAHL 1958). Separation of the polynucleotide strands is accompanied by an increase in buoyant density of about 0.017 gm cm^{-3}. At various times during renaturation of Drosophila DNA, aliquots were diluted and cooled to prevent further reaction and analyzed by CsCl pycnography (Figure 5). Samples d, e, and f were taken at times equivalent to C_0t values of 8, 98, and 554 (Figure 3). Renaturation by measurements of hypochromicity at 260 mμ was 12%, 53%, and 86%. Renaturation, as measured by the density increment relative to an initial increase of 0.017 g cm^{-3} after denaturation, was 25%, 50%, and 65%. During renaturation the density distribution of DNA becomes more narrow, probably as a consequence of the formation of higher molecular weight complexes which have lower rates of diffusion. Such molecular weight changes result from elongation of duplexes by concatenation of the randomly sheared fragments (THOMAS 1966).

The thermal stability of renatured DNA is also useful in characterizing the reformation of base-paired structures (MCCARTHY and MCCONAUGHY 1968). Table 1 lists the T_m's of renatured *Drosophila melanogaster* ($C_0t = 554$) and *B. subtilis* ($C_0t = 20$) DNA's. Compared with the native sheared DNA, renatured Drosophila DNA exhibits about 90% of the original hyperchromicity, and a stability of 84° C, about 1° C below that of native DNA. Comparable values were

TABLE 1

Thermal stabilities of native, sheared native, and renatured DNA's

DNA	Native		Sheared, native		Renatured	
	T_m in °C	Percent hyperchromicity	T_m in °C	Percent hyperchromicity	T_m in °C	Percent hyperchromicity
D. melanogaster	85.5	38%	85.0	38%	84.0	36%
B. subtilis	86.7	32%	86.0	27%	85.5	24%

Denaturation of Drosophila and Bacillus DNA's used in the renaturation experiments illustrated in Figures 3 and 6 was followed by measuring the hyperchromic transition at 260 mμ. Native DNA, native sheared DNA (10,000 psi, MW = 8×10^5 daltons), and renatured *Drosophila melanogaster* DNA ($C_0t = 1600$) and renatured *B. subtilis* DNA ($C_0t = 20$) were denatured in $1 \times$ SSC at a concentration of 1.5 mg/ml (Drosophila, and 40 μg/ml *B. subtilis*). The midpoints of the thermal transition (T_m) and the percent hyperchromicity were determined with a Gilford spectrophotometer attached to an automatic recorder.

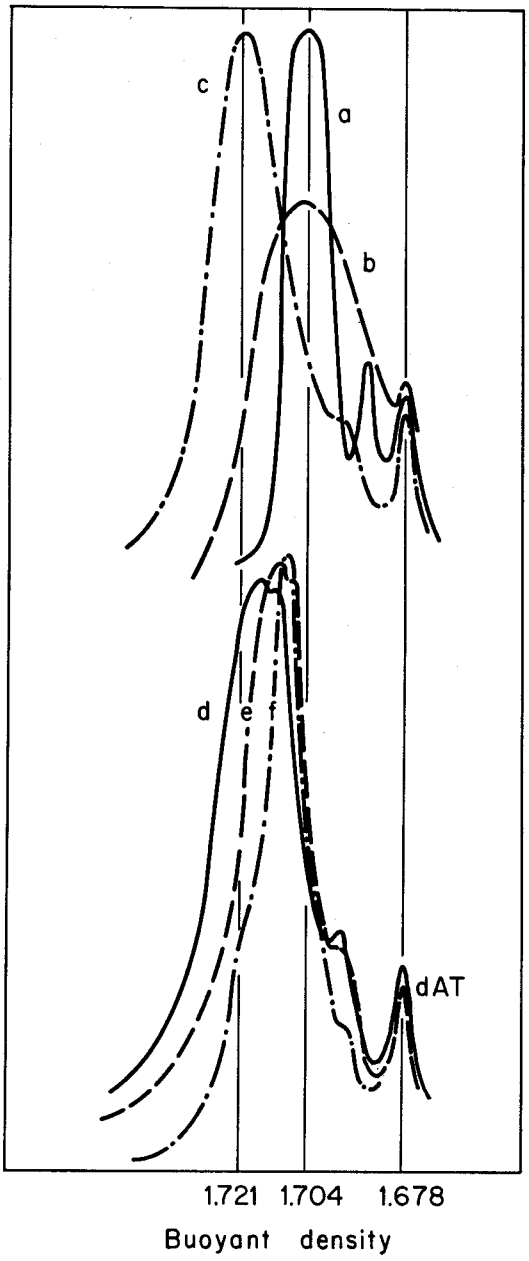

FIGURE 5.—Analytical pycnography of Drosophila DNA after various extents of renaturation. Sheared *D. melanogaster* pupal DNA was heat-denatured (100°C, 10 min) in 1 × SSC at 1.7 mg/ml and renatured at 62°C. Samples were diluted 10-fold into cold 0.1 × SSC at various times for subsequent analysis by analytical CsCl equilibrium centrifugation. The complete renaturation curve for the Drosophila DNA used in this experiemnt is shown in Figure 3. The analysis illustrates equilibrium distributions of samples taken at $C_0t = 8$ (d, 32 min, 12% renaturation); $C_0t = 98$ (e, 5 hr, 53% renaturation); and $C_0t = 554$ (f, 28 hr, 86% renaturation). Tracings a, b, and c are of native, sheared and denatured DNA. The samples, containing 5–9 μg of Drosophila DNA and 1 μg poly dAT as a density marker, were centrifuged 24 hrs at 20°C in a Spinco An F rotor.

found for the bacterial DNA which was expected to form perfectly paired duplexes after renaturation. Thus the buoyant density and thermal stability measurements show that most of the Drosophila DNA renatures homogeneously, and that the products of renaturation are well-paired structures.

E) *Renaturation Kinetics of other Dipteran DNA's:*

The size and complexity of genomes from species other than *D. melanogaster* have been analyzed by optical measurements of renaturation kinetics. Figure 6 shows that *D. simulans* DNA renatures with kinetics essentially identical to those observed for *D. melanogaster* DNA with a $C_0t_{1/2}$ of 80. *D. funebris* DNA, however, renatures with a $C_0t_{1/2}$ of 160 (Figure 6). The bacterial control, in this case *Bacillus globigii* DNA, had a $C_0t_{1/2}$ of 5. These data have been analyzed by the curve-fitting procedure discussed in Section C above. The results (Figure 4B) indicate that the data fit second-order rate curves (dashed lines) assuming that the following percentages of DNA are unique sequences: *D. melanogaster*, 95%, (A); *D. simulans*, 95% (B); *D. funebris*, 88% (C). The corrected $C_0t_{1/2}$ value for *D. funebris* is about 200, or about twice that of the *D. melanogaster* and *D. simulans*. This suggests that the minimum haploid genome of *D. funebris* contains about twice as many different genes as does that of *D. melanogaster*. *D. simulans* DNA,

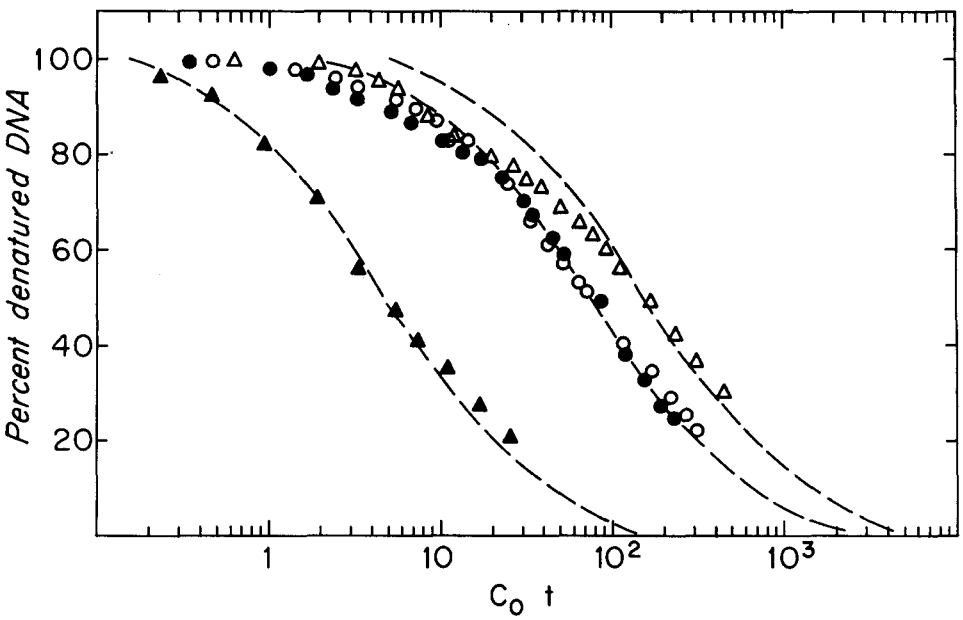

FIGURE 6.—Renaturation of Drosophila DNA's. DNA's from different Drosophila species were sheared and heat-denatured, and incubated at 60°C in 1 × SSC, 0.2% Tween 80, under mineral oil, in 1 mm light-path cuvettes (1 cm light-path cuvettes were used for bacterial DNA's). The hyperchromicity at 260 mμ was used to measure the extent of denaturation. The concentrations of DNA were 317 μg/ml (*D. melanogaster*, ○), 435 μg/ml (*D. simulans* ●), 595 μg/ml (*D. funebris* △), and 22 μg/ml (*Bacillus globigii* ▲). The percent denatured DNA is plotted *versus* C_0t. The dashed lines drawn through the 50% renaturation value for each DNA are theoretical renaturation curves for second-order reactions.

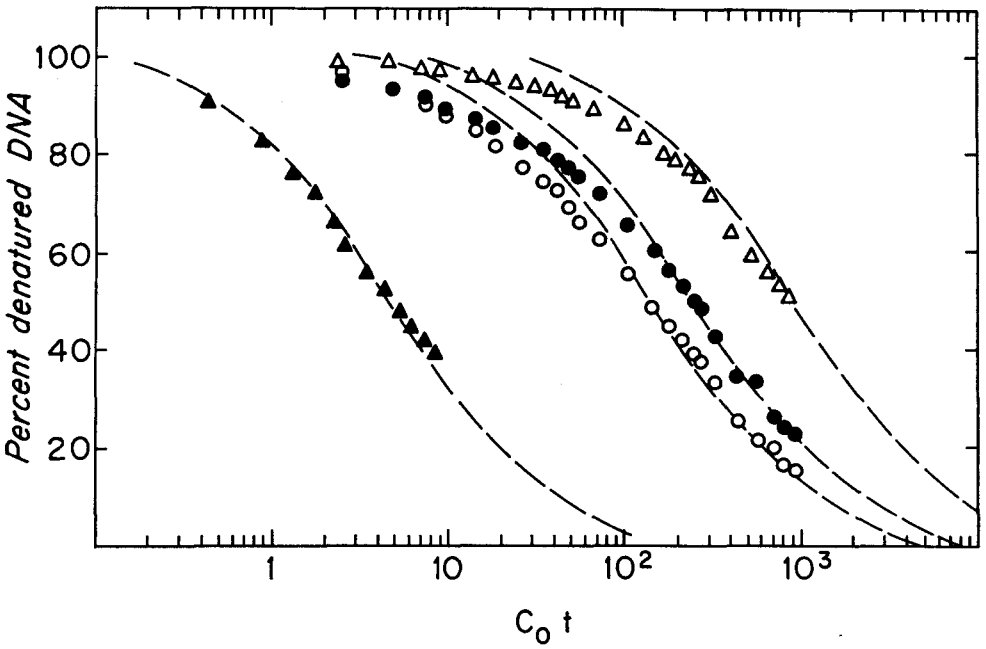

FIGURE 7.—Renaturation of nuclear Dipteran DNA's. Nuclear DNA's from *D. melanogaster*, *D. funebris*, and the flesh fly (*Sarcophaga bullata*) were sheared and heat-denatured, and renatured as described in the legend to Figure 6, at concentrations of 910 μg/ml (○), 900 μg/ml (●), and 855 μg/ml (△), respectively. The dashed lines, drawn through the 50% renaturation value, are theoretical curves for second-order reactions. The renaturation of bacterial DNA (*Cytophaga succinicans*, ▲) was carried out at 67 μg/ml under similar conditions.

on the other hand, shows the same sequence heterogeneity as does its sibling species, *D. melanogaster*. During the early stages of renaturation, all of the Drosophila DNA's reacted somewhat more rapidly than expected for a simple second-order reaction (Figures 6 and 7). The fitting of these data to idealized curves (Figure 4B) indicates that the multiplicity of these rapidly reacting sequences is about 70 for *D. simulans* and *D. melanogaster* (curves A and a; B and b) and about 40 for *D. funebris* (C and c). This relative simplicity contrasts markedly with the 1,000 to 400,000 family size range observed with mouse DNA (BRITTEN and KOHNE 1968). To determine if this simplicity is limited to the Drosophila genus, we determined renaturation kinetics of DNA from a different family of insects, Sarcophagidae (order Diptera). The data shown in Figure 7, and analyzed in Figure 4B, indicate that the flesh fly, *Sarcophaga bullata*, also has a relatively simple genome with about 8% repetitious sequences, belonging to families of about 60 members (curves D and d, Figure 4B). The C_0t at half-renaturation is approximately 6 times that of *D. melanogaster*, indicative of a minimum genome of 4×10^{10} daltons.

DISCUSSION

The precise inferences obtainable from protein data concerning the history of a

genome can be extended by the study of polynucleotide sequences. Specific examples of multiple cistrons (YANKOFSKY and SPIEGELMAN 1962), have been generalized to include large numbers of related sequences within many eukaryote genomes. In some vertebrates, the size of such gene families far exceeds that expected from the protein data (BRITTEN and KOHNE 1968).

From a consideration of cytological data, it seems unlikely that large duplications played an important role in the evolution of modern Drosophila species (PATTERSON and STONE 1952). That this conclusion extends to the molecular level was suggested by the rates of Drosophila DNA/DNA duplex formation using filter-bound DNA. Preliminary estimates suggested that only five to ten percent of *D. melanogaster* DNA reassociated with a slightly greater rate than expected for this genome. The thermal stability of these complexes was lower than that of well-paired duplexes whereas products formed after more extensive renaturation displayed native-like thermal stability (LAIRD and MCCARTHY 1968b).

The results detailed here permit more precise estimates of the size and complexity of several Dipteran genomes. Table 2 summarizes these estimates obtained from kinetics of solution renaturation of total and nuclear DNA's. Haploid genomes of *D. melanogaster* and *D. simulans* appear to have molecular weights 25 times that of *E. coli*, or about 7×10^{10} daltons. This value is at the lower end of the microspectrophotometric estimates obtained by KURNICK and HERSKOWITZ (1952). The precise relationship between information content, as measured by

TABLE 2

Characterization of Dipteran DNA's

	Buoyant density in CsCl, g cm^{-3} (% G+C)[1]	T_m, °C (% G+C)[2]	Percent repetitious sequences[3]	Average multiplicity of repetitious sequences[4]	Minimum size of haploid genome[5] (daltons).
Drosophila					
D. melanogaster	1.702 (43%)	85.6 (40%)	5–10	60	7×10^{10}
D. simulans	1.702 (43%)	85.7 (40%)	5–10	60	7×10^{10}
D. funebris	1.698 (39%)	84.0 (36%)	10–15	40	14×10^{10}
Sarcophaga					
S. bullata	1.693 (33%)	82 (31%)	5–10	80	40×10^{10}
Bacterial control					
B. subtilis	1.703 (44%)	86.6 (42%)	—	—	2×10^{9}

[1] Densities were measured using *Myxococcus xanthus* as standard (1.727 g cm^{-3}). Base compositions were calculated from SCHILDKRAUT, MARMUR and DOTY (1962).

[2] Thermal denaturations were carried out in 1 × SSC. Base compositions were calculated using the relationship established by MARMUR and DOTY (1962).

[3] The data of Figures 3, 6 and 7 have been analyzed as shown in Figure 4B to provide this estimate of percent repetitious sequences.

[4] The displacement between curves (A and a, for example) shown in Figure 4B is taken as an estimate of family size.

[5] These estimates, taken from Figure 4B, are based on ratios of the $C_0t_{1/2}$ value of *E. coli* DNA (Figure 4) and the value obtained from the curve (A, B, C, or D) which fits second-order kinetics for the various DNA's. The value for *B. subtilis* is taken from DENNIS and WAKE (1966).

renaturation rates, and the actual amount of DNA per cell is presently under investigation.

That these sibling species have genomes of identical size is predictable in light of the banding pattern similarity of the salivary gland chromosomes (HORTON 1939). It is surprising, however, that the nucleotide sequence diversity of *D. funebris* DNA is twice as great as that of *D. melanogaster*. These species have been separated for perhaps 40×10^6 years (THROCKMORTON, personal communication) and are classified in different subgenera. On the basis of nucleotide sequence homology differences between these species, the genus Drosophila is comparable in diversity to the taxon "order" among mammals (LAIRD and MCCARTHY 1968a). However, very little difference in DNA content per haploid cell exists among the entire class Mammalia, although variation among even closely related plants is common (STEBBINS 1966).

It is not clear which DNA value, if either, represents that of the ancestral Drosophila line. It is possible that *D. melanogaster* and other species in the Sophophora subgenus represent a reduction in DNA values. Alternatively, *D. funebris* may represent a species which has undergone duplication of sequences by polyploidy, with subsequent divergence so that these duplicate sequences are no longer recognizably similar. Since autopolyploidy is theoretically improbable after the establishment of differentiated sex chromosomes (MULLER 1925; OHNO 1967), it may be that such duplication occurred prior to fixation of the sex-determining *X* and *Y* homologues in Drosophila. Although banding pattern relationships are too distant between these species for salivary gland chromosome comparisons, metaphase chromosomes of the two species appear to be rearranged by centric fusion only (PATTERSON and STONE, 1952). However, it is perhaps significant that genetic markers on the *X* chromosome of *D. funebris* are separated by twice the map distance as are the homologous markers in *D. melanogaster* (PERJE 1955). Further comparison of these and other species may indicate whether nucleotide sequences have been lost in the evolution of *D. melanogaster* or added during the evolution of *D. funebris*.

It will be of interest to examine these values in light of suggestions made by STEBBINS (1966) that primitive, ancestral-like species often have a greater number of genes than do their more specialized derivatives. The argument is based on the concept that variety and versatility of gene function may be more vital to an unspecialized organism. In this respect, *D. funebris* is thought to be more similar to the ancestral line, on the basis of morphology and karyotype, than is *D. melanogaster* (PATTERSON and STONE 1952), and their genome sizes are consistent with this correlation.

The rather large genome of Sarcophaga is also of interest in comparison with mammals. A haploid mouse cell, for example, has a DNA equivalent of about 2×10^{12} daltons (VENDRELY and VENDRELY 1949), only five times that of Sarcophaga and thirty times that of *D. melanogaster*. While the difference between flies and rodents may be rationalized in terms of the sophistication of the organism, the significance of a six-fold difference between flies is not obvious.

The insect genomes, while differing considerably in size, show remarkable

similarity in the extent of intragenome homology. About 10% of the DNA base sequences appear to be partially redundant; the average size of these families of related sequences is about 40 to 80. These numbers are far lower than those obtained for mammalian DNA's when estimates are based on similar criteria of nucleotide sequence homology. Bovine DNA, for example, seems to consist of 60% unique, and 40% partially redundant sequences. The family size of these repititious sequences ranges up to 100,000 (BRITTEN and KOHNE 1968). The relative simplicity of the Drosophila genome, however, is not characteristic of all invertebrates, since sea urchin DNA, for example, shows a considerable amount of intragenome homology (BRITTEN and KOHNE 1968). Other examples of simple genomes among eukaryotes are known, the most extreme offered by yeast (BICKNELL, personal communication; BRITTEN and KOHNE 1968).

One possible explanation for these differences may be that simple genomes are correlated with higher mutation rates. Clearly, if the rate of divergence by fixation of mutations exceeds the rate of creation of new intragenome homologies by gene duplication, apparent genome simplicity would result. Some evidence for very distant sequence homologies in *D. melanogaster* has been found by reassociating DNA at reduced temperature (LAIRD and MCCARTHY 1968b). These sequences may represent members of larger groups of common ancestry which have since diverged to a considerable extent. In some organisms, recombination restrictions may also force rapid divergence of related sequences (THOMAS 1966). Alternatively, specific requirements of gene function may determine to what extent intragenome homology may be tolerated by an organism.

SUMMARY

The nucleotide sequence heterogeneity of DNA influences the kinetics of strand reassociation during renaturation. Kinetic measurements of renaturation have been used to estimate the number of different sequences in several insect DNA's. Our studies with Dipteran DNA's indicate that the minimum genome sizes of the sibling species, *Drosophila melanogaster* and *D. simulans*, are about 7×10^{10} daltons, corresponding to 100,000 cistrons of 1,000 nucleotides. In contrast to this, the genome of *D. funebris*, a more distantly related species, has about twice as many different nucleotide sequences.—Intragenome homology, such as would be expected to result from gene duplication, is limited. From five to fifteen percent of the DNA of these species renatures more rapidly than expected for a second-order reaction. The average multiplicity of similar nucleotide sequences is on the order of 50 to 100. Renaturation kinetics of *Sarcophaga bullata* DNA also indicate the presence of limited intragenome homology. These results are contrasted with rapid renaturation observed with DNA from many other eukaryotes.—One group of genes which share a common function, the ribosomal RNA cistrons, has been analyzed by RNA/DNA hybridization. The kinetics of hybridization of *D. melanogaster* ribosomal RNA to homologous DNA suggest that most of the 28S rRNA cistrons have similar nucleotide sequences.

LITERATURE CITED

BRITTEN, R. J. and D. E. KOHNE, 1967 Nucleotide sequence repetition in DNA. Carnegie Inst. Wash. Yearbook **66:** 78–106. ——, 1968 Repeated sequences in DNA. Science **161:** 529–540.

CAIRNS, J., 1963 The chromosome of *E. coli.* Cold Spring Harbor Symp. Quant. Biol. **28:** 43–46.

DENNIS, E. S. and R. G. WAKE, 1966 Autoradiography of the *Bacillus subtilis* chromosome. J. Mol. Biol. **15:** 435–439.

FREIFELDER, D., 1965 A rapid technique for the preparation of purified bacteriophage DNA or RNA from crude lysates. Biochim. Biophys. Acta **108:** 318–320.

HORTON, I. H., 1939 A comparison of the salivary gland chromosomes of *Drosophila melanogaster* and *D. simulans.* Genetics **24:** 234–243.

JUKES, T. H., 1966 *Molecules and Evolution.* Columbia University Press, New York.

KOHNE, D., 1968 Isolation and characterization of bacterial ribosomal RNA cistrons. Biophys. J. **8:** 1104–1118.

KURNICK, N. B. and I. H. HERSKOWITZ, 1952 The estimation of polyteny in Drosophila salivary gland nuclei based on determination of DNA content. J. Cell. Comp. Physiol. **39:** 281–299.

LAIRD, C. D. and B. J. MCCARTHY, 1968a Magnitude of interspecific nucleotide sequence variability in Drosophila. Genetics **60:** 303–322. ——, 1968b Nucleotide sequence homology within the genome of *Drosophila melanogaster.* Genetics **60:** 323–334.

MARMUR, J., 1961 A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3:** 208–218.

MARMUR, J. and P. DOTY, 1961 Thermal renaturation of deoxyribonucleic acids. J. Mol. Biol. **3:** 585–594. ——, 1962 Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. Mol. Biol. **5:** 109–118.

MESELSON, M. and F. STAHL, 1958 The replication of DNA in *Escherichia coli.* Proc. Natl. Acad. Sci. U. S. **44:** 671–680.

MCCARTHY, B. J., 1967 The arrangement of base sequences in DNA. Bacteriol. Rev. **31:** 208–227.

MCCARTHY, B. J. and B. L. MCCONAUGHY, 1968 Related base sequences in the DNA of simple and complex organisms. I. DNA/DNA duplex formation and the incidence of related base sequences in DNA. Biochem. Genetics **2:** 37–53.

MULLER, H. J., 1925 Why polyploidy is rarer in animals than in plants. Am. Naturalist **59:** 346–353.

OHNO, S., 1967 *Sex Chromosomes and Sex-Linked Genes.* Springer-Verlag, Berlin, New York.

PATTERSON, J. T. and W. S. STONE, 1952 *Evolution in the Genus Drosophila.* The Macmillan Company, New York.

PERJE, A. M., 1955 Genetic and cytological studies of *D. funebris;* some sex-linked mutations and their standard order. Acta. Zool. Stockholm **36:** 51–66.

RITOSSA, F. and S. SPIEGELMAN, 1965 Localization of DNA complementary to ribosomal RNA in the nucleolus organizer regions of *Drosophila melanogaster.* Proc. Natl. Acad. Sci. U. S. **53:** 737–745.

ROBERTSON, W. R. B., 1916 Chromosome studies. I. Taxonomic relationships shown in the chromosomes of Tettigidae and Acrididae. J. Morphol. **27:** 179–331.

RUBENSTEIN, I., C. A. THOMAS, JR. and A. D. HERSHEY, 1961 The molecular weights of T2 Bacteriophage DNA and its first and second breakage products. Proc. Natl. Acad. Sci. U. S. **47:** 1113–1122.

SCHILDKRAUT, C. L., J. MARMUR and P. DOTY, 1962 Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J. Mol. Biol. **4**: 430–443.

STEBBINS, G. L., 1966 Chromosomal variation and evolution. Science **152**: 1463–1469.

STONE, W. S., 1955 Genetic and chromosomal variability in Drosophila. Cold Spring Harbor Symp. Quant. Biol. **20**: 256–270.

THOMAS, C., 1966 Recombination of DNA molecules. Prog. Nucleic Acid Res. Mol. Biol. **5**: 315–348.

VENDRELY, R. and C. VENDRELY, 1949 La teneur du noyau cellulaire en acide désoxyribonucléique à travers les organes, les individus et les espèces animals. Experientia **5**: 327–329.

VERMEULEN, C. W. and K. C. ATWOOD, 1965 The proportion of DNA complementary to ribosomal RNA in *Drosophila melanogaster*. Biochem. Biophys. Res. Commun. **19**: 221–226.

WETMUR, J. G. and N. DAVIDSON, 1968 Kinetics of renaturation of DNA. J. Mol. Biol. **31**: 349–370.

YANKOFSKY, S. A. and S. SPIEGELMAN, 1962 Distinct cistrons for the two ribosomal RNA components. Proc. Natl. Acad. Sci. U. S. **49**: 538–544.