NUCLEAR VOLUME CONTROL BY NUCLEOSKELETAL DNA, SELECTION FOR CELL VOLUME AND CELL GROWTH RATE, AND THE SOLUTION OF THE DNA C-VALUE PARADOX

T. CAVALIER-SMITH

Department of Biophysics, University of London King's College,
26–29 Drury Lane, London WC2B 5RL, England

SUMMARY

The 40000-fold variation in eukaryote haploid DNA content is unrelated to organismic complexity or to the numbers of protein-coding genes. In eukaryote microorganisms, as well as in animals and plants, DNA content is strongly correlated with cell volume and nuclear volume, and with cell cycle length and minimum generation time. These correlations are simply explained by postulating that DNA has 2 major functions unrelated to its protein-coding capacity: (1) the control of cell volume by the number of replicon origins, and (2) the determination of nuclear volume by the overall bulk of the DNA: cell growth rates are determined by the cell volume and by the area of the nuclear envelope available for nucleocytoplasmic transport of RNA, which in turn depends on the nuclear volume and therefore on the DNA content. During evolution nuclear volume, and therefore DNA content, has to be adjusted to the cell volume to allow reasonable growth rates.

The great diversity of cell volumes and growth rates, and therefore of DNA contents, among eukaryotes results from a varying balance in different species between *r*-selection, which favours small cells and rapid growth rates and therefore low DNA C-values, and *K*-selection which favours large cells and slow growth rates and therefore high DNA C-values. In multicellular organisms cell size needs to vary in different tissues: size differences between somatic cells result from polyteny, endopolyploidy, or the synthesis of nucleoskeletal RNA. Conflict between the need for large ova and small somatic cells explains why lampbrush chromosomes, nurse cells, chromatin diminution and chromosome elimination evolved. Similar evolutionary considerations clarify the nature of polygenes, the significance of the distribution of haploidy, diploidy and dikaryosis in life cycles and of double fertilization in angiosperms, and of heteroploidy despite DNA constancy in cultured cells, and other puzzles in eukaryote chromosome biology.

Eukaryote DNA can be divided into genic DNA (G-DNA), which codes for proteins (or serves as recognition sites for proteins involved in transcription, replication and recombination), and nucleoskeletal DNA (S-DNA) which exists only because of its nucleoskeletal role in determining the nuclear volume (which it shares with G-DNA, and performs not only directly, but also indirectly by coding for nucleoskeletal RNA). Mechanistic and evolutionary implications of this are discussed.

INTRODUCTION

The immense variation in the haploid DNA content (the DNA C-value) of eukaryotes is a major problem in biology (for reviews see Callan, 1967; Holliday, 1970; Thomas, 1971; Rees & Jones, 1972; Smith, 1972; Hinegardner, 1976). Eukaryote

C-values vary from 0·005 pg in a yeast (Sokurova, 1973) to 200 pg in the dinoflagellate *Gonyaulaux* (Holm-Hansen, 1969) – a 40 000-fold range. Though some authors have suggested that this variation is related to the number of genes in an organism or its developmental complexity (e.g. the number of cell types – Kauffmann, 1971), most are now agreed that there is no significant correlation between amount of DNA and organismic or genetic complexity. This lack of correlation is referred to as the C-value paradox (Thomas, 1971; Lewin, 1974). It is most strikingly shown in unicellular eukaryotic algae, where there is a 5000-fold range in C-values (Holm-Hansen, 1969; Sparrow, Price & Underbrink, 1972), but no significant variation in developmental complexity; this range exceeds that in multicellular plants and animals, where maximum C-values are around 100 pg (in the lungfish *Lepidosiren* and the plant *Fritillaria*). These observations seem to rule out all explanations of the C-value paradox in terms of differing numbers of protein-coding genes or of differing needs for transcriptional control of such genes.

Though there are problems in estimating the numbers of protein-coding genes in eukaryotes (Bishop, 1974), traditional arguments, plus recent estimates by hybridization of messenger RNA complexity (Davidson, 1976; Hereford & Rosbash, 1977), are consistent with the idea that all eukaryotes have only between 4000 and 30 000 protein-coding genes, and that the bulk of the DNA in high C-value eukaryotes does not code for proteins. These considerations have led many authors to suggest that most eukaryote DNA is 'junk or garbage' (Ohno, 1972) which has 'no function' (Gierer, 1974), or is 'relatively useless junk' (Comings, 1972). It has been suggested that DNA C-values evolve by a mixture of random drift and an orthogenetic tendency to acquire extra DNA (Hinegardner, 1976) and that selection is relatively unimportant in determining C-values.

There is, however, considerable evidence for a strong correlation between DNA C-values and a variety of cellular and organismic properties; this strongly suggests that C-values evolve in response to selection, and therefore that the amount of DNA in a nucleus has some important but unknown function, quite separate from, and in addition to, the protein-coding function of DNA. Bennett (1971, 1972) has referred to this function as 'nucleotypic' so as to contrast it with the better understood protein-coding or 'genic' function of DNA. The nucleotype and the protein-coding genes both affect the phenotype; both are directly heritable and therefore genetic in nature. The purpose of this paper is 3-fold: (1) to review the evidence that DNA C-values are determined by strong selective forces and to discuss the nature and consequences of these forces; (2) to give a molecular explanation for the nucleotypic or non-genic function of DNA as well as for the mechanisms of evolution of different C-values; and (3) to show how these evolutionary and mechanistic considerations clarify a wide variety of hitherto puzzling aspects of chromosome behaviour ranging from chromatin elimination and supernumary chromosomes, through the significance of repetitive DNA sequences and the nature of polygenes, to the chromosome aberrations that occur in cancer cells and established cell lines.

My basic argument is that natural selection acts powerfully on organisms to determine their cell size and developmental rates (which are inversely related). The mean

cell volume of an organism is the result of an evolutionary compromise between con-
flicting selection for large cell size and for rapid developmental rates: the particular
compromise reached for a particular species will depend on its ecological niche and
organismic properties. Since larger cells require larger nuclei, selection for a particular
cell volume will secondarily select for a corresponding nuclear volume, producing
a close correlation between cell and nuclear volumes in different organisms. I suggest
that the basic nucleotypic function of DNA is to act as a nucleoskeleton which deter-
mines the nuclear volume; small C-values are therefore required by small cells with
small nuclei, and large C-values by large cells needing large nuclei. The DNA C-value
of an organism is therefore simply the secondary result of selection for a given nuclear
volume, which in turn is the secondary result of the evolutionary compromise between
selection for cell size and for developmental rates.

Correlation between DNA C-values, cell volume and cell cycle length

A strong positive correlation between cell size and DNA C-value has been estab-
lished in angiosperm plants (Martin, 1966; Price, Sparrow & Nauman, 1973) and in
vertebrates (Commoner, 1964; Szarski, 1976; Olmo & Morescalchi, 1976). In 10
species of unicellular algae, where total organic carbon content was taken as an index
of cell size, this correlation extends over a 600-fold range (Holm-Hansen, 1969). At the
other end of the scale it is noteworthy that the yeasts, which have the smallest cells
among eukaryotes, also have the lowest C-values. There are indications that this
relationship extends also to the prokaryotes, both to bacteria (Commoner, 1964) and to
blue-green algae; the larger-celled blue-green algae have the highest C-values of any
prokaryote (Stanier & Cohen-Bazire, 1977) – larger than those yeasts with the smallest
genomes (Table 1).

A highly significant inverse correlation exists between C-value and developmental
rates in eukaryotes as diverse as angiosperms (Bennett, 1972; Smith & Bennett, 1975),
amphibians (Goin, Goin & Bachmann, 1968) and insects (Bier & Müller, 1969);
species having low amounts of DNA develop rapidly and have very short minimum
generation times whereas those with high C-values have much slower development.
In herbaceous plants annual species have much lower DNA contents than perennial
species (for diploid monocotyledons the mean DNA content of perennials is over
5 times that for annuals); ephemerals have less than a quarter the DNA content of
other annuals. At the cellular level there is a strong positive correlation between
C-values and the length of the mitotic cell cycle (Van't Hof & Sparrow, 1963; Van't
Hoff, 1965; Evans, Rees, Snell & Sun, 1970) and the length of meiosis (Bennett, 1971)
in both plants and animals. Yeasts have the lowest C-values and the shortest cell
cycles and therefore most rapid rates of population growth of any eukaryotes.

Clearly low C-value organisms have small cells and rapid reproductive rates while
high C-value organisms have large cells and slow growth rates. This basic correlation
is true of all eukaryotes, including protozoa, algae and fungi as well as multicellular
animals and plants, and reveals a fundamental feature of DNA evolution and function
that is quite independent of the variations in bodily structure, number of genes, or
developmental mechanisms that occur in these diverse groups. The C-value paradox

can be solved only by finding the reasons for these correlations. The correlation between cell size, cell cycle length and DNA content means either that one of these 3 physically determines the others or that all 3 are determined by a fourth unknown factor. The increased cell size and reduced growth rate of artificial polyploids (Darlington, 1937; Swanson, 1958), and of plants with extra supernumary chromosomes (Jones, 1975, 1976), suggests that the DNA content itself is a fundamental physical determinant of cell volumes and growth rates.

Table 1. *Haploid DNA contents in various groups of organisms*

	Lowest C-value, pg	Highest C-value, pg	Ratio $\dfrac{\text{Highest}}{\text{Lowest}}$
Prokaryotes	0·0017	0·013	7·7
Mycoplasmata	0·0017	0·0037	2·2
Other bacteria	0·0033	0·01	3·0
Blue-green algae	0·0026	0·013	4·8
Eukaryotes	0·005	350	70 000
Algae	0·04	200	5 000
Protozoa	0·06	350	5 800
Bryophytes	0·64	4·3	6·7
Tracheophytes	1·0	310	310
Pteridophytes	6·0	310	52
Gymnosperms	4·2	50	12
Angiosperms	1·0	89	89
Fungi	0·005	0·19	38
Annelids	0·9	5·3	5·9
Arthropods	0·1	22·6	230
Crustaceans	0·7	22·6	32
Insects	0·1	7·5	75
Molluscs	0·43	5·4	13
Echinoderms	0·54	3·3	6·1
Chordates	0·20	100	500
Fish	0·39	9·8	25
Amphibians	1·2	100	83
Urodeles	19	100	5·3
Anurans	1·2	7·9	6·6
Reptiles	1·5	3·5	2·3
Birds	1·7	2·3	1·4
Mammals	3·0	5·8	1·9
Bats	3·0	3·9	1·3

Data calculated from Hinegardner (1976), Sparrow *et al.* (1972), Sokurova (1973), Stanier & Cohen-Bazire (1977). Ranges for some groups e.g. Arthropods probably underestimated because strongly *r*-selected groups like copepods and mites have not been studied.

Cell volume determination and DNA content

A general correlation between the size of actively dividing cells and their DNA content is found in bacteria (Commoner, 1964) as well as in eukaryotes, which supports the idea that all cells have a common mechanism of size determination. There is evidence in both prokaryotes (Donachie, 1968) and eukaryotes (Fantes *et al.* 1975) that the initiation of DNA replication and subsequent cell division depends on the attain-

ment of a critical cell volume: models to explain this postulate the accumulation of an initiator or the dilution by cell growth of a repressor specific for replicon origins (Sompyrac & Maaloe, 1973; Donachie, 1974). Such a mechanism will automatically control the volume of the cell; moreover if the number of replicon origins is doubled or quadrupled by polyploidy the cell volume at initiation will also be doubled or quadrupled. Increasing the number of replicons by duplication (see below) or by aneuploidy would also increase the cell size.

If, as I argue below, cell size is highly adaptive, it is important to consider how it is genetically controlled.

If cell size in eukaryotes, as in prokaryotes, is primarily controlled by titrating replicon origins against cell volume or mass, then evolutionary adjustments to cell size could occur in 3 ways: (1) by varying the number of replicon origins, (2) by varying the amount of repressor or initiator synthesized per replication, and (3) by varying the affinity of repressor or initiator for the replicon origins. I suggest that the commonest and simplest way of varying cell size is to vary the number of replicons. Some variation in replicon number could occur without causing corresponding variations in C-value, by altering the lengths of replicons. However eukaryote replicon lengths are much more uniform than are cell volumes. Therefore major changes in replicon number must be accompanied by changes in overall DNA content and selection for particular cell sizes may be a major source of variation in C-value.

The adaptiveness of cell volumes and growth rates

Though there have been several suggestions concerning the adaptive significance of cell size (Szarski, 1976) and growth rates (Stebbins, 1966), their fundamental evolutionary significance has not been fully appreciated by cell and molecular biologists or geneticists. The reason why eukaryote cells differ so widely in size and growth rate is that they are subject to widely differing forms of selection. Of key importance is the relative significance for each species of the two opposing forms of natural selection which MacArthur & Wilson (1967) have called *r*-selection and *K*-selection. *r*-selection is most important in species that repeatedly need to colonize relatively empty or ephemeral environments (e.g. aphids or annual weeds), while *K*-selection is more characteristic of those (e.g. oak trees or elephants) that compete most effectively in stabler but crowded environments. Pianka (1970) has pointed out that *r*-selection favours rapid development, a high maximal intrinsic rate of population growth (r), early reproduction, a small body size and a short life time, whereas *K*-selection favours slower development, greater competitive ability, delayed reproduction, larger body size and a longer life time; for any species there is a compromise between *r*- and *K*-selection, but the compromise is different for different species – a continuous spectrum exists between the most highly *r*-selected organisms such as bacteria and highly *K*-selected ones such as elephants. I suggest that the most fundamental way organisms adapt to varying *r*- and *K*-selection is by evolving particular cell volumes and cell growth rates.

Since cell size and the length of the cell cycle are positively correlated, it is obvious that in unicellular organisms selection for large size will tend to select for slow growth

T. Cavalier-Smith

and a high C-value, and selection for rapid growth will select for small cell sizes and a low C-value. Thus the C-value of different unicellular organisms depends fundamentally on a compromise between 2 conflicting kinds of selection: (1) for a high growth rate, and (2) for large size. In unicellular saprophytes (e.g. fungi and bacteria) the balance of advantage is for small size and high growth rate, so C-values are small. In unicellular predators (e.g. amoebae – some may have the highest known C-values (Friz, 1968), ciliates, radiolarians, some dinoflagellates) selection is more for large size, so C-values are large. For unicellular photosynthesizers (e.g. *Chlamydomonas* or *Porphyridium*) cell size and C-values are often intermediate. Among algae (whose C-values vary by a factor of 5000) slow growing species with large cells have high C-values and rapid growers with small cells have low C-values. Clearly small-celled, rapidly growing low C-value organisms are extreme r-strategists and large-celled, slow growing, high C-value organisms are K-strategists.

The same is true of multicellular organisms, though the relationship is somewhat obscured by 3 extra factors: (1) the indirect relationship between cell growth rates and organismic growth rates, (2) the possibility of independent variation in cell size and body size, and (3) the possibility of differences in cell size and growth rates in different cells in the body. Despite these complications, it remains true that in most groups of animals and plants there is a good correlation between strong r-selection, small cells and low C-values on the one hand and between K-selection, large cells and high C-values on the other. In angiosperms annual plants have low C-values and perennial ones high C-values (Bennett, 1972); small-celled, rapidly reproducing insects like *Drosophila* have very small C-values, larger slower reproducers like grasshoppers have very high C-values (Bier & Müller, 1969); slow sluggish vertebrates with large cells (lungfish, *Amphiuma*) have the highest C-values and small-celled rapidly growing and metabolizing species, like birds and certain teleost fish, the lowest.

In view of the 3 extra factors mentioned above, it is at first surprising that there should be such a good correlation between the C-value of a multicellular organism and its position on the r-K-selection continuum. I suggest that the most important reason for this unexpectedly good correlation is that all multicellular organisms are unicellular at a critical stage of their life cycle: the fertilized egg. r-selection will favour rapid egg growth and development and therefore small cell size and low C-values: the length of meiosis will be an important rate limiting factor for r-selected species – a *Drosophila* could not tolerate the many months needed for egg production in many amphibia (even male meiosis takes longer in many high C-value organisms than the entire generation time of r-adapted species – Bennett, 1971). Since the selective advantage of a 10 % increase in developmental rate is equivalent to a 100 % increase in fertility (MacArthur & Wilson, 1977, p. 85) r-selection will strongly favour small cell size. K-selection which favours ability to compete in a crowded environment (Roughgarden, 1971) will on the other hand favour the production of relatively few large eggs with extensive food reserves, which can develop, without further nutrient, to a stage able to feed or photosynthesize in competition with others. In the absence of countervailing selection for small cell size this selection pressure would tend also to increase the volume of somatic cells. In addition there will be direct selection for increased size

in certain somatic cells, e.g. secretory cells. In large animals especially strong selection for large cell size would be expected for nerve cells and in large vascular plants for xylem and phloem cells.

The very existence of multicellular organisms is a response to extreme K-selection for increased bodily size, and food gathering and storage capacity in a crowded environment. Though the most strongly K-selected organisms are multicellular and the most strongly r-selected ones are unicellular, there is a considerable overlap in body size and reproductive rates between unicellular and multicellular organisms – extreme K-selection has produced macroscopic unicells like *Acetabularia* whereas extreme r-selection has produced microscopic multicellular animals like rotifers or copepods. It is therefore not in the least surprising that relatively r-selected multicellular organisms like *Drosophila* have less DNA per cell than many K-selected unicellular species, despite their greater structural and developmental complexity. Furthermore if one compares a multicellular organism like a rotifer with a unicellular organism like *Paramecium* having a similar overall body size and occupying a similar adaptive zone one would expect the multicellular species to have a lower C-value than the unicellular one because it has smaller cells. Multicellularity is a way of increasing body size without having to increase cell size and C-value in proportion: this immediately explains why strongly K-selected unicellular organisms, which lack this ability, have higher C-values than mammals (and some higher than any multicellular plants or animals). Coupled with the fact that the most highly r-selected species are unicells, this also explains why the range of C-values is greater in unicellular than in multicellular eukaryotes.

Though it could be argued that C-values vary for some mysterious unknown reason and that high C-value organisms simply happen to be pre-adapted to K-selected niches and low C-value ones to r-selected niches, it is more straightforward to postulate that the variation in C-values is simply the result of varying degrees of r- and K-selection: this solves the C-value paradox very simply.

Nuclear volume determination by nucleoskeletal DNA

It could be argued that cell size could be varied in evolution by increasing the number of replicons and varying the concentration and binding affinity of controlling proteins without changing the overall amount of DNA. Though there must be some scope for this, the fact that there is such a good correlation between cell size and DNA content suggests that such variation plays a minor evolutionary role. This implies that large cells actually require large amounts of DNA and small cells small amounts. I suggest that DNA acts as a nucleoskeleton determining the volume of the nucleus, and that larger cells require large nuclei and therefore correspondingly more DNA. There is, in fact, good evidence for a very close correlation between nuclear volumes, DNA content, and cell volumes of actively dividing cells (Price *et al.* 1973).

I argue that nuclear size is of key importance because it affects the rate of cell growth and thereby the length of the cell cycle. If one accepts the idea that messenger and ribosomal RNA pass from nucleus to cytoplasm via the nuclear pores (Franke, 1974), then the number of nuclear pores per cell will potentially limit the growth rate. Since in a wide variety of dividing cells the number of pores per unit area is fairly

constant (Maul & Deaven, 1977), the rate of RNA transport to the cytoplasm will depend on the area of the nuclear envelope. Since the nuclear envelope is normally attached to the outer surface of the chromatin mass, it is reasonable to suggest that the area of the envelope depends on and is determined by the volume (and shape) of the enclosed chromatin. Since the C-value determines the chromatin volume, it will also determine the nuclear volume and the nuclear surface area, and in consequence the rate of cell growth and the length of the cell cycle.

If selection increases cell size, it will tend to slow down the growth rate unless the surface area of the nucleus is increased in direct proportion. This would require a more than proportional increase in nuclear volume and therefore C-value, since for a spherical nucleus of volume V the surface area is proportional to $V^{\frac{2}{3}}$. If the cell volume increases by a factor n, the C-value would have to increase by $n^{1 \cdot 5}$ to allow the same growth rate. On this hypothesis the increased cell cycle time observed in angiosperms with increased C-value is the simple and inevitable consequence of the fact that the C-value increases only in direct proportion to the cell volume. In algae, however, C-values increase with the $1 \cdot 22$ power of the cell volume (calculated from Holm-Hansen, 1969) and in vertebrates also (Bachman, Goin & Goin, 1972) C-values increase more than proportionally with cell volume indicating that selection for increased cell size is accompanied by selection against excessive reductions in growth rate: in these cases higher C-value organisms will tend to have higher nucleocytoplasmic ratios and longer replicons. That greater cell volumes necessitate increases in nuclear volume to sustain rapid growth is strongly suggested by the immense growth of oocyte nuclei in animals and of the giant primary nucleus in *Acetabularia*. In both cases nuclear pores are maximally close-packed on the surface of the nuclear envelope, which also increases greatly in area to support the massive transfer of RNA to the cytoplasm: despite this huge increase, overall growth is slow compared with that of smaller cells.

The length of the cell cycle will be proportional to the amount of material that needs to pass into the cytoplasm (itself proportional to cellular volume) divided by the rate of transport (proportional to $V^{\frac{2}{3}}$). The cell cycle length is thus given by

$$\frac{bV}{aV^{\frac{2}{3}}}$$

which simplifies to $dV^{\frac{1}{3}}$, where a, b, d are arbitrary constants. If, as I have argued, V is directly proportional to the C-value, then the cell cycle length will be given by

$$kC^{\frac{1}{3}},$$

where C is the C-value and k is an arbitrary constant. Though other factors must also be involved, putting k = $6 \cdot 4$ gives a surprisingly good fit to the data for dicotyledons and putting k = $5 \cdot 4$ gives a rather less good fit to those for monocotyledons (Fig. 1). This supports the idea that the important quantitative or structural role for DNA that has been repeatedly postulated by numerous authors (Mather, 1943, 1949; Bennett, 1971; Szarski, 1976) is as a nucleoskeleton determining nuclear volume. The recent discovery of a proteinaceous nuclear matrix (Berezney & Coffey, 1977), which may also have a skeletal role, is in no way contrary to this suggestion, so long as one assigns

to DNA the primary role in the determination of nuclear volume; if, for example, the DNA doubles in a polyploid the amount of the proteinaceous matrix would double secondarily.
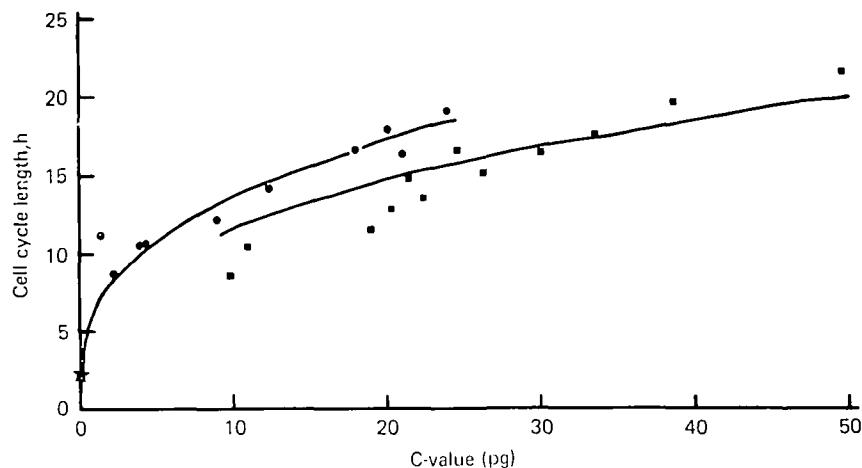

Fig. 1. Relationship between cell cycle length and C-value. ●, dicotyledons; ■, mono-cotyledons; ★, yeast. Angiosperm data from Rees (1972). The 2 curves are calculated from the formula $kC^{1/3}$, putting k equal to 5·4 and 6·4.

Replicon organization: G-DNA and S-DNA

I suggest that eukaryote nuclear DNA is of 2 distinct kinds, genic DNA and nucleo-skeletal DNA. Genic or G-DNA, in which nucleotide sequences are highly specific, codes for RNA and polypeptide sequences, or else acts as recognition sites for DNA-binding proteins (and RNAs if they exist) responsible for the control of transcription, replication and recombination. Nucleoskeletal, or S-DNA (an abbreviation N-DNA might be confused with nuclear or nucleolar DNA), by contrast does not code for protein but acts (together with G-DNA) as a nucleoskeleton to determine interphase nuclear volumes. The great variation in C-value is caused by variation in the amount of S-DNA in response to selection for particular nuclear sizes: its nucleotide sequence may be of little significance. If the amount of G-DNA is relatively constant in eukaryotes, and crossing over during meiosis is initiated by nucleases specific for sites present only in G-DNA, this would explain why meiotic recombination appears to be restricted to small regions of the genome in high C-value organisms (Thuriaux, 1977); mitotic recombination, which by contrast is roughly proportional to physical distance (Baker *et al.* 1976), would occur in both S- and G-DNA. There are probably minor DNA components apart from G- and S-DNA (e.g. K-DNA serving as attachment sites for kinetochores, and T-DNA forming telomeres (Cavalier-Smith, 1974)) which will not be considered here.

G-DNA and S-DNA as defined here do not correspond with euchromatin and heterochromatin; chromomeres are so much larger than genes that it is necessary to postulate that euchromatin contains both G- and S-DNA. Constitutive hetero-

chromatin (Yunis & Yasmineh, 1972) must, however, contain a relatively much greater concentration of S-DNA, and may often lack G-DNA altogether. I suggest that supernumary (including B) chromosomes, which are sometimes though not always heterochromatic, lack G-DNA altogether and consist entirely of S-DNA plus small amounts of K-DNA and T-DNA; this would explain their dispensibility and their effects on cell size and the length of the cell cycle (Jones, 1975, 1976). In ordinary (A) chromosomes G and S-DNA would normally be interspersed.
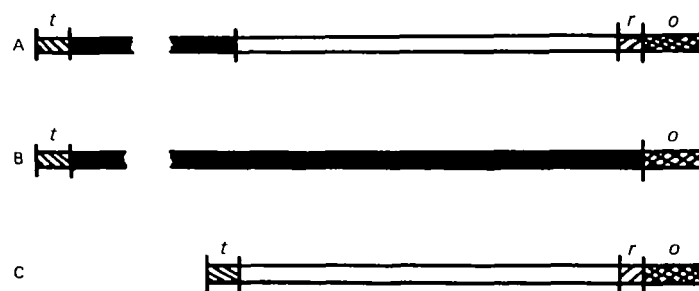

Fig. 2. Three possible types of replicon organization. In all three the replicon origin (o) and the terminus (t) consist of middle-repetitive DNA and replication proceeds to the left. In A both S-DNA (mainly unique, but sometimes including highly repetitive DNA) and G-DNA are present, whereas type B contains only S-DNA and type C only contains G-DNA. S-DNA is shown in black. G-DNA includes a unique sequence structural gene plus regulatory sequences (r) containing promoter, operator and recombinator sites.

Possible arrangements for S- and G-DNA in eukaryote replicons are shown in Fig. 2. I suggest that type A containing both S- and G-DNA is the usual kind of replicon in euchromatin, except in hypotrich ciliate macronuclei (Lauth, Spear, Heumann & Prescott, 1976) where they are type C; in those fungi with the lowest C-values also many, or even most, replicons may be of type C consisting only of G-DNA. In constitutive heterochromatin most (and in supernumary chromosomes all) replicons will be type B, consisting only of S-DNA. The recent discovery of gene inserts in eukaryote genes (Breathnach, Mandel & Chambon, 1977; Jeffreys & Flavell, 1977) raises the possibility that, in high C-value organisms at least, the non-coding intervening sequence contains some S-DNA as well as G-DNA (one would expect at least some specific sequence G-DNA near the ends to serve as recognition sites for excision). If this is so, then the G-DNA section of type A replicons would exist as 2 or more pieces separated by S-DNA.

Although it has been suggested that middle repetitive DNA is concerned with transcriptional control during differentiation (Britten & Davidson, 1969; Davidson, 1976), its distribution in eukaryotes is not what would be expected on that hypothesis, but supports the idea proposed here that it functions in initiation of replication of all 3 kinds of replicon. Its presence in fairly similar proportions of total DNA in simple unicellular organisms like *Dictyostelium* (Jacobson & Lodish, 1975) and dinoflagellates (Allen, Roberts, Loeblich & Klotz, 1975) as well as in highly differentiated multicellular organisms (regardless of immense variations in C-value) argues strongly against a

function in transcriptional control; on Britten & Davidson's model one would expect the overall number of middle repetitive sites to correlate with the number of structural genes and organismic complexity and not with C-value as on my model. It is dubious whether their more recent model (Davidson, Klein & Britten, 1977) is applicable to all eukaryotes, since Hn RNA is absent in most unicells (Jacobson & Lodish, 1975; Prescott, Stevens & Lauth, 1971; Hudspeth, Timberlake & Goldberg, 1977) – even ones with relatively high C-values and much middle repetitive DNA. I suggest
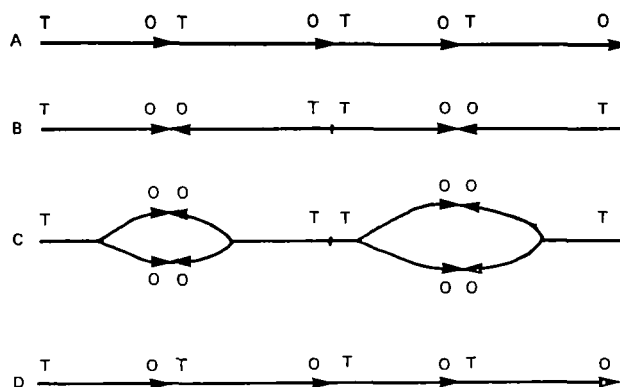

Fig. 3. Model for the control of replication by replicon inversion. A, in G_1 replicons are tandemly arranged with adjacent origins (O) and termini (T). B, replication is initiated by the inversion of alternate replicons as a result of crossing over between their origins and termini. After replication (C) is terminated at the end of S-phase reinversion occurs to produce the original sequence (D).

instead that evolutionarily conserved middle repetitive sequences serve as templates for the synthesis of RNA replicon initiator molecules like the *oop*-RNA of λ-phage (Hayes & Syzbalski, 1973). Various small nuclear RNAs (Goldstein, 1976; Tamm, 1977) in the size range 75–740 nucleotides may include replicon initiator RNAs; these would be quite distinct from the much shorter RNA molecules that initiate Okazaki fragments – the very small size of Okazaki fragments in eukaryotes (Edenberg & Huberman, 1975) suggests that one is initiated per nucleosome and that their initiation need not be sequence-specific. If replicon-initiator RNA is excised after use without degradation and is doubled in amount once per cell cycle, it could play an important role in the control of replication and the determination of cell size (Sompyrac & Maaloe, 1973; Donachie, 1974). Evidence that RNA covalently attached to pulse-labelled DNA hybridized preferentially with middle repetitive DNA (Taylor, 1974) is consistent with my model. So also is the unique interspersion pattern in *Drosophila* and *Chironomus* since these alone of the animals studied have polytene chromosomes, which are likely to require different replication controls from normal chromosomes.

To be consistent with bidirectional replication (Edenberg & Huberman, 1975) the replicons must be arranged in pairs with adjacent origins during replication (Fig. 3B, C). I suggest that in non-S-phase cells origins are adjacent to termini (Fig. 3A, D), and that S phase is initiated by the inversion of replicons by crossing over between origin

and terminus. Termination would cause inversion back to the G_1 state and also set in train the events leading to cell division. This inversion mechanism (which occurs during chromosomal replication of the G-segment of bacteriophage μ (Bukhari, 1976)) would greatly simplify the control of replication and division. It can explain the apparent mobility of foldback DNA (Perlman, Phillips & Bishop, 1976), since replication would create and destroy inverted repeats. In renatured foldback DNA OO and TT sequences would be seen as perfect hairpins and whole replicons as stem-loop structures (Cavalier-Smith, 1977). Moreover if the origin and terminus normally have to undergo recombination during replication, eukaryotes would be strongly predisposed to undergo replicon duplication (Keyl, 1965).

In many organisms the pattern of replicon initiation varies during the life cycle (Edenberg & Huberman, 1975; Callan, 1973): in rapidly growing embryonic cells and in yeast a maximum number of replicons are simultaneously active; in animal somatic cells only a subset is active at any given time, different subsets being activated in succession; in premeiotic *S*-phase even fewer replicons (perhaps only one per chromomere) are active. I suggest that this stage-specific modulation of replication is what requires middle repetitive RNA primers, and that the origins of replicons comprising a subset activated at a given time belong to a single family of middle repetitive DNA. Very low C-value organisms like yeast could economize on DNA by having several genes in a replicon, or by shortening their middle repetitive DNA.

Though the observed interspersion patterns of middle repetitive DNA generally fit in with such a model, *Drosophila* embryos (Blumenthal, Kriegstein & Hogness, 1974) and hypotrich ciliate macronuclei (Lauth *et al.* 1976) are at first sight apparent exceptions. In *Drosophila* the middle repetitive sequences are too infrequent to serve as embryonic replicon origins, though their separation agrees rather well with that of replicon origins (one per chromomere) in polytene chromosomes; however, in embryos there is no need for special sequences to ensure temporal control since all replicons initiate simultaneously (the short length of Okazaki pieces suggests that initiation of DNA chains is in principle possible without special initiation sequences). In hypotrich macronuclei also the usual temporal order of replication of particular sequences is absent – particular sequences are not replicated at corresponding times during successive *S*-phases, despite the sequential movement of the replication band across the nucleus (Gall, 1959): the degradation of middle repetitive sequences during macronuclear formation (Lauth *et al.* 1976) is therefore comprehensible.

S-DNA and polygenes

S-DNA can provide a simple explanation for polygenes, which Mather (1943, 1944, 1949) thought to be qualitatively distinct from major genes, and which have a major role in the quantitative inheritance of traits like height, weight or the number of insect chaetae (bristles). Some authors have questioned the idea that polygenes are qualitatively different from ordinary protein-coding genes; instead they consider that polygenes simply reflect the minor pleiotropic effects of major genes primarily affecting other traits (Mayr, 1970). Though some polygenic effects probably are produced in this way, the fact that polygenes are found in heterochromatin as well as in euchromatin

(Mather, 1943, 1949), shows that pleiotropy of major genes cannot be a universal explanation.

If S-DNA controls cell and nuclear size, and growth rates, as I have argued, then variations in its amount will alter these properties, just as occurs in polygenic inheritance (Barigozzi & Di Pasquale, 1953). In multicellular species they will have similar quantitative effects on the growth rate of the whole organism, and (except in species where size regulation is independent of the number of cells, e.g. most vertebrates) on overall body size. Even in organisms showing size regulation, one would predict numerous secondary consequences of changes in cell size and growth rate; any process of pattern formation (Summerbell, Lewis & Wolpert, 1973; Wilby & Ede, 1975) involving counting of cell numbers or precise timing will be altered. For example, if cell size increases in an animal of fixed size there must be fewer cells in each part; this could alter quantitative structural characters (e.g. chaetae number in insects – a classic case of polygenic inheritance (Mather, 1944, Thomson, 1975a) – since each chaeta is produced by one cell) and even behavioural properties dependent on the number or size of nerve cells (e.g. visual activity or learning – triploid salamanders do learn more slowly than diploid ones (Fankhauser, Vernon, Frank & Slack, 1955)). I therefore propose that polygenes in Mather's sense not only exist but consist simply of S-DNA, which since it acts without coding for proteins is, as he predicted, qualitatively different in its mode of action from major protein-coding genes. If polygenes consist of S-DNA, then large amounts of S-DNA could control quantitative characters without imposing a heavy genetic load, and controversies about the number of polygenes (Vetta, 1975; Thomson, 1975b) and their evolution are placed in a radically new light.

Though the quantity of DNA is clearly a major determinant of cell size, a few major genes must also be involved (e.g. to code for replication initiators or repressors); there is indeed genetic evidence that major genes as well as polygenes can affect cell size.

HnRNA, heterochromatin and nuclear volume modulation

If, as I suggested above, many developmental processes in multicellular organisms are highly sensitive to cell size and cell growth rates, then different sizes and growth rates will probably be optimal in different tissues; complex multicellular development will therefore require independent modulation of cell and nuclear volume in different cell types. I propose that the function of a major fraction of the heterogeneous nuclear RNA (HnRNA) sequences (Jelinek *et al.* 1974) is to increase nuclear volume and that only a minor part of it is messenger precursor. If a large part of the S-DNA is potentially transcribable into HnRNA the volume of the nucleus could be increased by increasing the amount of HnRNA in the nucleus. This nucleoskeletal RNA fraction of the HnRNA could affect the nuclear size in the same way as DNA itself by combining with 'swelling proteins', which cell fusion (Harris, 1970) and nuclear transplantation studies (Gurdon & Woodland, 1968) suggest cause the major increase in volume that occurs before shrunken 'switched off' avian erythrocyte nuclei become transcriptionally active. The presence of HnRNA in animals (Davidson, 1976; Jelinek *et al.* 1974) and its absence in the slime mould *Dictyostelium* (Jacobson & Lodish, 1975) and in other unicells (Prescott *et al.* 1971) fits in with this idea: its great length and kinetic

complexity does not favour a role for bulk HnRNA in gene regulation (Scherrer, 1974; Davidson *et al.* 1977) nor does the rapid and uniform evolutionary divergence (Davidson, 1976) of unique DNA in different species of the same genus. None of these pose problems for the functions postulated here, since, except for replicon and transcription origins and termini, the exact sequence in most HnRNA would be unimportant. Different batteries of HnRNA molecules could be transcribed in different tissues according to the nuclear volume requirements. The recent discovery (Chikaraishi, Deeb & Sueoka, 1978) that rat HnRNA sequences of kidney are a subset of those in liver, which in turn are a subset of those in brain, is hard to reconcile with the regulatory hypothesis but is in keeping with the idea of nuclear volume regulation. Liver nuclei tend to be larger than kidney nuclei, and brain nuclei are larger still (Altman & Katz, 1976). Some brain nuclei are larger than those of any other somatic cells. In high C-value organisms messenger precursor could form a very small fraction of total HnRNA molecules.

This proposal implies that S-DNA is of 2 kinds: non-transcribed S-DNA, which I suggest may correspond with constitutive heterochromatin (Brown, 1966; Yunis & Yasmineh, 1972), which will have a fairly constant effect on nuclear volume, and S-DNA that can be transcribed to produce nucleoskeletal RNA which could have a variable and potentially greater effect on nuclear volume according to how many different sequences and how many copies of each were transcribed. It means that not all regions of S-DNA will have identical phenotypic effects. Locatable polygenes (Thompson, 1975*a*) would be those from which especially large numbers of copies of nucleoskeletal RNA are transcribed in certain tissues. The need for controllable reductions in nuclear volume could be met by facultative heterochromatinization – normally of transcribable S-DNA, but in extreme cases (e.g. avian erythrocytes or the second X chromosome of female mammals) also of G-DNA.

Blood cells and the exceptionally low variance of amniote C-values

Most major groups of eukaryotes show wide interspecific variation in C-values (see Table 1) – usually over 1, 2, or even 3 orders of magnitude. The only exceptions are reptiles, mammals and, above all birds, which have remarkably uniform C-values: among mammals bats have particularly low and uniform C-values. I suggest that this uniformity results from exceptionally strong stabilizing selection for optimal erythrocyte size in these highly active air-breathing vertebrates. Vigorous activity requires a high concentration of haemoglobin in the blood, rapid circulation and rapid exchange of gases between erythrocytes and plasma. These requirements conflict: increased haemoglobin concentration requires larger and/or more numerous red cells. Larger cells will reduce gas exchange rates and make capillary flow more difficult: increased cell numbers will also increase the viscosity and therefore the load on the heart. The result of this will be strong selection for an optimal cell size and therefore a restricted range of DNA C-values.

Since mammals, and to a lesser extent birds, also experience a constant temperature throughout life, some of their developmental processes may have become much more dependent on constant cell cycle lengths as well as cell volumes than are those of other

organisms. This could explain why triploidy is invariably lethal in mammals (White, 1973) and normally also birds, but not in any other animals or in plants. As many as 4 % of spontaneous human abortions are triploid: this and the fact that the only triploid recorded as born alive showed severe respiratory distress and died 6 h later (Schindler & Mikamo, 1970) supports the idea that stabilizing selection on blood cell volumes is particularly strong in mammals. Significantly, 2 viable human diploid/triploid mosaics (Ellis, Marshall, Normand & Penrose, 1963) had diploid and not triploid blood cells: nuclear volumes were greater in the triploid than in the diploid cells (Mittwoch, 1968) as would be expected. It may be significant that birds and mammals are also the only major groups in which evolutionary polyploidy has never been demonstrated. It is also possible that mammals can tolerate an average C-value double that of reptiles and birds only because they alone are able to eliminate their erythrocyte nuclei to compensate for their increased volume.

The more frequent lethality of aneuploidy in animals, especially mammals, than in plants could also be because of their greater sensitivity to variations in cell size. In plants viable trisomics can often be obtained for all chromosomes, but in humans only trisomics involving the smaller chromosomes are viable and even these have greatly reduced fitness (White, 1973): this may indicate the critical importance even of small changes in cellular volume in mammals. It is possible that X-chromosome inactivation in female mammals (White, 1973) originated to keep cell volumes the same in the 2 sexes, rather than for gene dosage compensation. If facultative heterochromatinization prevents the binding to replicon origins of the initiator (but not repressor) molecules responsible for the control of cell proliferation, it will reduce cell size. I postulate that constitutive heterochromatin lacks such sites altogether, and that replication of both facultative and constitutive heterochromatin is triggered by the prior replication of euchromatin by a mechanism distinct from that which initiates S-phase: late replication of heterochromatin (Comings, 1972) whether constitutive or facultative, must inevitably result from this lack of involvement in the initiation of S-phase. Y-chromosomes, consisting mainly of constitutive heterochromatin, will therefore not affect cell size and, even when nearly as large as the X chromosomes, will have relatively little effect on nuclear volume compared with transcribed euchromatic X chromosomes. The absence of X-chromosome inactivation in birds is explicable by the relatively insignificant size of their sex chromosomes.

Many of the defects in viable human aneuploids might result from a nucleotypic effect on cell or nuclear volumes or growth rates. For example the abnormalities in XO humans (Turner's syndrome) and the mental deficiency of XXX humans, both of which like normal females have only one active X, might occur because the presence or absence even of inactivated X chromosomes significantly affects nuclear volumes and growth rates: it may not be necessary to postulate genic imbalance in the embryo prior to inactivation. Non-specific mental deficiency in human trisomics (Wahrman *et al.* 1976) might be a nucleotypic effect. The steady reduction in dermal ridge number on human fingers as the number of sex chromosomes increases (Polani, 1969) is just the sort of effect that the resulting increased cell size might cause if a cell-counting mechanism is involved in ridge formation. However in *Drosophila*, where triploids

are viable, gene imbalance (in *G*-DNA, *S*-DNA or both) presumably causes the inviability of most trisomics. Yet one would expect that transcriptional feedback control could compensate for different numbers of genes in a cell, since this will be necessary in every cell cycle when some genes are replicated and others not. The frequency of dominance and the fact that deletions of small regions of DNA are often not lethal in heterozygotes (Waddington, 1939) and the occurrence of dosage compensation without X-inactivation in *Drosophila* also shows that cells are usually well able to compensate for 2-fold variations in the relative numbers of different genes. The lethality of human triploids mentioned above, despite their balanced genome, also suggests that the effects of a changed amount of DNA may be more important in mammals than genic imbalance.

Evolution of diploidy and dikaryosis

No satisfactory evolutionary explanation has yet been given for the distribution of haploidy, diploidy and dikaryosis in eukaryote life cycles (Raper & Flexer, 1970; d'Amato, 1977). Existing explanations for diploidy are either genetic (buffering against recessive mutation or heterosis – Williams, 1966; Raper & Flexer, 1970; d'Amato, 1977), or very unspecific (e.g. to allow 'extensive tissue differentiation' or 'more subtle control of cellular activity' – Raper & Flexer, 1970). I believe that selection for optimum nuclear and cell size can explain all the observed variations in ploidy more simply and economically. Since diploid nuclei and cells are twice the volume of haploid ones, stages in the life cycle where large size is of selective advantage will be diploid, those where small size is preferable will be haploid.

The diploidy of animals, the more massive brown algae and the tracheophyte plants is immediately explicable in terms of selection for large cell size, whereas there is much evidence from the study of exceptional natural, as well as artificial, haploids in these groups that haploidy will allow complex differentiation and control of cell activity as well as does diploidy. In green (and red) algae, as well as brown, diploidy is characteristic of larger, slower growing plants with large cells such as the Siphonales and Oedogoniales, whereas small-celled rapidly dividing unicells like *Chlamydomonas* are haploid. In protozoa with obvious selection for large cell size (e.g. the predatory ciliates, amoebae, Foraminifera and Heliozoa) diploidy is the rule whereas where numerous small cells are more advantageous (e.g. the endoparasitic Sporozoa) haploidy prevails. The existence of dikarya in higher fungi, and their absence in other organisms, is readily explained by my theory. Because of their unique filamentous growth during which protoplasm, including nuclei, has to flow through minute pores in the transverse septa, they must be subject to intense selection for small nuclei, which has resulted in the lowest eukaryote C-values. Following hyphal fusion 2 separate haploid nuclei will be superior to one diploid one because their greater surface area will allow more rapid RNA transport and growth; they would also be less subject to damage when squeezing through pores. Selection for high spore productivity could be an additional reason for the low C-value, but the existence of complex adaptations for dikaryosis like the basidiomycete clamp connexions argues strongly for the view that diploids (which can be produced experimentally) are strongly selected against as my theory predicts.

The evolution of autopolyploidy in protozoa (Dogiel, 1965) as well as in plants and animals may commonly result from selection for large cell size.

Variations in ploidy (and in unicellularity versus multicellularity), like variations in C-value, are primarily the simple results of varying degrees of r- and K-selection, rather than subtle genetic or epigenetic adaptations: of course diploidy allows, as secondary effects, heterosis and dominance but its distribution among eukaryotes does not suggest that this is why it evolved; there would also be problems in explaining (without invoking group selection) how heterosis and dominance led to the origin of diploidy.

Nuclear volume variations and DNA inconstancy

In differentiated multicellular organisms one would expect different nuclear sizes to be optimal in different tissues. Tissues with large cells or ones highly active in protein synthesis would require a larger nuclear surface area (and therefore nuclear size) than in smaller or less active cells, possibly much larger than the HnRNA volume modulation mechanism would allow. If my theory is correct this diversifying selection should result in a wide range of DNA contents in different cells of an organism. Most multi-cellular animals and plants do contain certain tissues with multiples of the 2C amount of DNA (d'Amato, 1977; White, 1973) (though this fact understandably is ignored by those who think of DNA only in genetic terms); the mechanism is usually endomitosis which results in endopolyploidy. The obvious explanation is that it increases both the number of copies of each gene and the nuclear volume, thus allowing greater rates of transcription and of RNA transport to the cytoplasm (and also of proteins into the nucleus to associate with the RNA); one would expect the degree of polyploidy to be greatest in organisms with rapid life cycles and small C-values (e.g. certain insects) and lowest in those with slower life cycles and already high C-values (e.g. *Allium*), which is what is observed (Swanson, 1958). Polyteny, where the euchromatin only is redupli-cated, is found instead of endopolyploidy in most Diptera (significantly not in the large primitive Tipulidae) and in some Collembola, and is explicable as a response to selec-tion for extreme economy of material in very small C-value organisms. In certain Radiolaria (White, 1973) reversible polyploidy occurs in the primary nucleus at the stage in the life cycle when the cell is largest.

C-values and xylem evolution in plants

Many features of C-value variations in vascular plants become clear when one considers the selective forces acting on the size of xylem cells. An important clue is given by the hitherto paradoxical variation in C-values of coniferous trees: different species have remarkably similar and very high C-values (the ratio between the highest and lowest C-values recorded is only 3·5), yet within a species the C-value commonly varies by a factor of 2 (Grant, 1976). In many species there is a continuous cline in the Northern hemisphere from low C-values in the South to high ones in the North. This can simply be explained by increased selection for large size in xylem cells (tracheids) and for large nuclei in cambial cells in colder climates with shorter growing seasons.

Coniferous tree growth depends on the rapid ascent of huge quantities of sap in the

lumen of the dead tracheid cells and the production of new tracheids by division and differentiation of cambial fusiform initial cells: both become increasingly difficult in lower temperatures and shorter growth periods. The viscosity even of pure water increases by a factor of 2 as the temperature drops from 25 °C to 0 °C, which will slow the transpiration stream in direct proportion unless the diameter of the tracheids and their interconnecting pits increases: (by Poiseille's Law flow rates increase with the fourth power of the radius of a tube, so can be considerably increased by having larger cells and pits). It might be thought that this could be done without increasing the C-value by means of endopolyploidy during xylem differentiation, as occurs in the formation of primary xylem in angiosperm roots (List, 1963; Roberts, 1976) where future xylem cells expand greatly after their last division. However in conifer wood the actively dividing cambial cells are already giant (in *Pinus strobus* (Wilson, 1964, 1966) being 2500 times the volumes of root meristem cells of *Allium cepa*, itself a very high C-value species) and expand relatively little during differentiation: moreover I have found no reference to endopolyploidy in any conifer cells (does it even occur in angiosperm secondary xylem ?) The immense size of these cambial cells makes them grow very slowly: in May and June the cell cycle lasts 10 days even in mid latitudes (Wilson, 1964) and there will undoubtedly be strong selection to minimize this in latitudes where only a few successive cell divisions are possible per season: since this cannot be done by reducing cell size because of the need for high flow rates, there will be strong selection for increased nuclear volume and therefore higher C-values. The need for large tracheids must be especially great in the cooler spring growth season, which explains the greater size of the first tracheids produced in an annual growth ring: even in angiosperms, whose xylem contains vessels which conduct sap more rapidly than do tracheids, many advanced north temperate trees here evolved a ring porous wood with spring wood composed exclusively of large vessels (Esau, 1953); this shows the great selective advantage in improving the rate of sap flow in spring and supports the idea that selection for large xylem cells is the basic reason for the high C-values of vascular plants.

I suggest that the evolution of xylem vessels in angiosperms had a major impact on the subsequent evolution of genome sizes. Vessels consist of a linear array of dead cells with their ends perforated so as to form a long tube. Not only does this greatly speed up sap flow, but since each vessel is built from many separate cells its overall size can be much larger, or alternatively the size of the component cells (vessel elements) be considerably reduced while still maintaining a good flow of sap. This change in xylem construction therefore made possible the evolution in response to strong *r*-selection of herbaceous annual and ephemeral angiosperms with their small rapidly dividing cells and low C-values; the evolution of endopolyploidy during primary xylem development (absent in pteridophytes and gymnosperms ?) and the loss of cambium were additional adaptations needed to allow a further reduction in C-value and cell volume.

In gymnosperms, however, continued strong selection for large tracheids maintained a high C-value and slow rate of development and so prevented the evolution of annuals. Significantly, gymnosperm shrubs have lower C-values than trees, the lowest being *Ephedra* (Sparrow, Price & Underbrink, 1972) which unlike most gymnosperms

has vessels. Most pteridophytes are also slow-growing high C-value perennials, probably for the same reason: it would be well worth studying the C-values and xylem development of the few annual pteridophytes (the lowest C-value recorded among ferns is for *Salvinia*, which has a greatly reduced vascular system: the C-value of another water fern *Marsilea*, one of the few pteridophytes with vessels in its roots and which unlike angiosperms or gymnosperms has multinucleate primary xylem precursor cells (List, 1963), is unfortunately unknown).

The high frequency of evolutionary polyploidy in angiosperms and its rarity in gymnosperms (Grant, 1976), may be partly because angiosperms are developmentally preadapted to the successful evolution of polyploidy because endopolyploidy is a normal process in their development, and partly because most gymnosperm C-values are already so high that further doubling of cell volume would so slow growth as to be strongly selected against.

Chromatin diminution and chromosome elimination

If the main function of S-DNA is to increase the nuclear volume, one might expect some or all of it to be dispensed with in tissues having small cells. Chromosome diminution and chromosome elimination have long puzzled cytologists but can be easily explained as a response to 2 conflicting selection pressures. Both processes are made possible by the segregation of nuclei into soma and germ line, which occurs in animals and most ciliate protozoa, but not in other microorganisms or plants. In animals the selective forces acting on soma and the germ line are quite different: the most rapid development of a fertilized egg to produce an organism able to feed results if ova are as large as possible and somatic cells are as small as possible. In small rapidly reproducing organisms like nematodes and insects diversifying selection of this kind must be far more intense than in slow-growing vertebrates. Chromatin diminution in somatic cells of nematodes, e.g. *Parascaris* (Swanson, 1958; Hyman, 1951) is an obvious response to selection for large ova (and therefore a large oocyte nucleus) and small somatic cells. Diminution has been reported also in the copepod *Cyclops* (White, 1973), a microscopic strongly *r*-selected genus with relatively large eggs. More extreme diminution is found in the macronuclei of hypotrich ciliates (Lauth *et al.* 1976) where 97 % of the DNA (presumably S-DNA) is destroyed and the residue (presumably G-DNA) is reduplicated many thousandfold, and exists as numerous gene-sized pieces, which serve for transcription during vegetative growth. The micronucleus retains its S-DNA to allow rapid trans-envelope RNA and protein transport during the brief sexual phase when the macronuclei are destroyed and must be rapidly regenerated. Macronuclear diminution results from selection for economy of material; S-DNA is no longer necessary for large nuclear size which automatically results from the hyperreduplication of the G-DNA. Such extensive chromatin diminution by loss of interspersed S-DNA could not occur in the soma of animals since in the absence of hyperreduplication efficient segregation requires that each piece of G-DNA is attached to a centromere.

However if S-DNA was not interspersed with G-DNA but was instead concentrated on separate chromosomes these could be eliminated as a whole from somatic cells. This I suggest is the basis for chromosome elimination, which occurs in the Sciaridae,

Cecidomyidae and Orthocladiinae – tiny flies (Diptera) which lay small numbers of large rapidly developing eggs. Tiny *Miastor*, with the most bizarre chromosome behaviour, lays only 3 or 4 giant eggs each of which develops to produce many larvae. The sciarids have a few large (L-) chromosomes, and the cecidomyids and ortho-cladiins many small E-chromosomes, which are confined to the germ line; both are eliminated early on from somatic cells (White, 1973): I suggest that they consist exclusively of S-DNA which serves to enlarge oocyte nuclei so as to allow rapid oogenesis, and that their elimination from somatic cells allows these to be smaller and to produce viable offspring more quickly from limited materials. In plants elimination of B chromosomes from somatic cells could serve a similar function: tissue-specific variation in their numbers (Jones, 1975) would allow systematic variation in nuclear size.

Origin of haplodiploidy

The selective forces acting on male and female are also often very different. Where these cause males to be much smaller than the already small females, one might expect male-specific chromosome elimination to evolve, as has occurred in armoured scale insects (Diaspididae) where in males alone the entire paternal chromosome set is eliminated early in development (White, 1973); in certain cecidomyids male somatic cells alone are haploid. A more extreme result of the same selective pressures would be complete male haploidy and female diploidy; such haplodiploidy has evolved only in animal groups with a strong tendency towards extreme minuteness and often male–female dimorphism (i.e. Hymenoptera, rotifers, mites and some Homoptera, Thysan-optera and Coleoptera). Facultative heterochromatinization of the paternal chromo-some set, which occurs in some scale insects (Coccoidea), is a less extreme response, and one which could have been the evolutionary precursor (White, 1973) of male-specific elimination or haploidy.

Lampbrush chromosomes and nurse cells

Even where the need for economy of materials in somatic cells is less than in the minute Diptera there is still strong diversifying selection for much larger nuclear size in oocytes than in other cells. In oocytes, unlike somatic cells where polyploidy or polyteny can allow a size increase, an immense increase in volume has to occur with only the 4C DNA amount as in normal G_2 cells. A possible mechanism for this in most animals and in many plants would be the synthesis of a large amount of RNA by certain parts of the S-DNA which in combination with specific proteins could serve to swell the nucleus. I suggest that this is the function of the bulk of lampbrush hetero-geneous RNA (Davidson, 1976) (L-RNA), and of the lampbrush stage itself, which occurs in animals soon after the amplification of ribosomal DNA just at that time when the need for transport of ribosomal nucleoprotein to the cytoplasm is greatest. (A minority of loops is likely to be concerned with the synthesis of messengers for proteins needed in large amounts at this stage, e.g. nuclear pore complex protein). The fact that lampbrush chromosomes also occur in the giant primary nucleus of the unicellular alga *Acetabularia* at the time of ribosomal DNA amplification and maxi-

mum need for ribonucleoprotein transport into the cytoplasm (Spring, Scheer, Franke & Trendelenburg, 1975) strongly supports this interpretation; since meiosis does not occur at this stage, there is no justification for the idea that lampbrush chromosomes have a special role in meiosis; nor does their presence in *Acetabularia* support the idea that the great complexity of lampbrush RNA means that it is messenger needed to programme the development of many differentiated cell types in the growing embryo.

The probable existence of interspersed middle repetitive sequences in L-RNA, as in HnRNA, simply results from their great length – transcription will run through replicon origins. Even though only part of the S-DNA synthesizes lampbrush RNA, the total sequences involved can be 10 times that needed for oocyte messenger. Selection for small nuclei and low C-values should therefore favour alternative mechanisms. Insects again provide a good test case for the theory. Large primitive insects with high C-values use lampbrush chromosomes; but many more advanced smaller insects (e.g. Diptera, Coleoptera) with low C-values have evolved a different and much more rapid meroistic oogenesis (Davidson, 1976; Chapman, 1969) in which lampbrush chromosomes are absent and the oocyte's ribosomal RNA is synthesized not by the oocyte itself but by neighbouring nurse cells which transfer it to the oocyte. Predictably, nurse cells have highly polyploid or polytene giant nuclei. I suggest that meroistic oogenesis was a prerequisite for a lowering in C-value, which by allowing small somatic cells and rapid development was the foundation of their evolutionary success (insect species with meroistic oogenesis probably equal those of all other organisms combined – Coleoptera, Diptera and Hymenoptera alone have over 500 000 described species (Richards & Davies, 1977)); the existence of a limited lampbrush stage in spermatocytes suggests that S-DNA has not been totally eliminated in these insects. The apparently contradictory behaviour of the meroistic Cecidomyid *E*-chromosomes is an exception that proves the rule: in most meroists the oocyte nucleus does not make RNA, so does not need to be large; in the Cecidomyid *Wachtiella* the oocyte nucleus *E*-chromosomes do synthesize RNA, which I suggest is L-RNA, and the nucleus grows considerably after the early breakdown of the nurse cells, later synthesizing RNA on its somatic chromosomes also (presumably pre-mRNA). I predict that sciarids' and orthocladiins' *L*- and *E*-chromosomes also will synthesize RNA and show concomitant nuclear growth in oocytes and/or nurse cells. I should also expect a general correlation between low C-values, absence of oocyte lampbrush chromosomes, and the presence of nurse cells in invertebrates other than insects.

The absence of chiasmata in lampbrush loops (Callan, 1967) would be expected if they consist of S-DNA rather than G-DNA.

Double fertilization in angiosperms

In angiosperm plants the nutritive material for the growing embryo is the endosperm, which (in contrast to the haploid female gametophyte that nourishes gymnosperm embryos) develops only after fertilization. Selection for rapid seed development would therefore favour haploidy in endosperm cells. However endosperm is never haploid; its ploidy varies from 2–15 n as a result of fusion between one of the 2 haploid

sperm nuclei and 1, 2, 4, 8 or 14 haploid embryo sac nuclei, which has long been un-
explained (Foster & Gifford, 1974). Endoreduplication may further increase its ploidy.
The evolution of double fertilization and polyploidy perhaps resulted from selection
for larger cells because of their greater ratio of protoplasm to extracellular material;
endopolyploidy often increases succulence in this way in plants and in the nutritive
cells of plant galls (d'Amato, 1977). (Can endosperm type also be related to cell size?).

S-DNA and jumping genes

It is necessary to explain not only the selective forces which cause the genome size to
vary, but also the physical mechanism of change in DNA amount. In some cases poly-
ploidy or aneuploidy may be involved. There is however much evidence that major
changes in C-value are frequent in the absence of any change in chromosome number.
I suggest that 2 quite distinct mechanisms are involved: (1) whole replicon doubling,
for which Keyl (1965) has provided good evidence and Maclean (1973) a possible
mechanism; and (2) deletions, duplications and transpositions of parts of replicons,
occurring mainly in the S-DNA.

With the exception of replicon origins and termini, and HnRNA and L-RNA
initiation and termination sites, the base sequences of S-DNA will be relatively free of
the stringent stabilizing selection which keeps G-DNA sequences more or less con-
stant, and so will be free to diversify by genetic drift. Rapid divergence does occur
between the bulk of unique sequences (but not middle repetitive ones) in related
species (Davidson, 1976). Though base changes, duplications and deletions (e.g. from
unequal crossing over, Smith, 1976) and inversions could be sufficient basis for such
changes, I suggest that transpositions between different S-DNA segments also are
frequent in eukaryotes. There are 2 reasons for supposing such 'jumping genes' to be
especially important in S-DNA evolution. First is a general theoretical argument: the
existence of massive amounts of DNA, whose sequence is unconstrained by stabilizing
selection acting via the phenotype, provides a potent source for the origin of DNA-
virus-like self-replicating elements, as well as a superb environment for their increase
by direct genic selection (Williams, 1966). Mutations in a section of S-DNA enabling
it to replicate at a greater rate than the 'host' chromosome and to insert into other
regions of S-DNA will automatically be selected. Some pieces of S-DNA may also
acquire the capacity to move from organism to organism as full-blown DNA-viruses:
this seems to me the most likely origin for DNA viruses. The second reason lies in the
growing evidence for transposable genetic elements in eukaryotes.

I suggest that these jumping genes normally have no obvious effect on the cell
phenotype because of their restriction to the S-DNA, though in genetically well
studied organisms like maize (Fincham & Sastry, 1974) and *Drosophila* (Rasmusson,
Green & Carlson, 1974) they can be detected. My postulate that eukaryote trans-
posable controlling elements are primarily the result of phenotype-unrelated genic
selection on S-DNA is in line with Fincham & Sastry's (1974) scepticism concerning
McClintock's (1951, 1967) arguments for their involvement in normal gene regulation.
This does not mean however, that their existence has no effects on G-DNA and its
regulation: once evolved some of these elements could become involved in the regula-

tion of DNA (e.g. by sequence inversion). Long palindromes (Cavalier-Smith, 1976) which appear to move about the genome (Perlman *et al.* 1976) provide evidence for quite rapid variations in the sequence organization of DNA; it seems simplest to suppose that they are components of S-DNA unrelated to the phenotype-controlling function of DNA (Cavalier-Smith, 1977). They could be part of transposons or be generated during the transposition or inversion of S-sequences, or as suggested above by the inversion of replicons during replication. The fact that differences in fertilizer treatment can cause inherited changes in the DNA content and cell size of flax (*Linum usitatissimum*) (Evans *et al.* 1966), *Viola* (Pierce, 1937) and *Nicotiana* (Hill, 1976) – presumably S-DNA alone is involved – raises the interesting possibility that quasi-Lamarckian processes may sometimes be involved in the evolution of polygenic characters influenced by S-DNA. Heritable environmentally induced changes in S-DNA content could occur if the excision and reinsertion of certain jumping genes in S-DNA is sensitive (like λ prophage induction) to environmental stimuli.

Co-evolution of G- and S-DNA

It is instructive to think of G- and S-DNA as being in perpetual evolutionary conflict, since each represents a rather different mode of evolution. S-sequences will be subject to high rates of genetic drift, to selection for replicative independence from the rest of the genome as episomes or plasmids and for transposability (even into the G-DNA). However, unless the S-sequences become able to infect other individuals as viruses their tendency to increase will be held in check by selection acting (on both S- and G-DNA) via the fitness of the organism carrying them. Such selection will act to prevent the S-sequences affecting the phenotype and to control the overall amount of S-DNA, but it would not prevent certain sequences replicating greatly at the expense of others, which could be one way in which highly repetitive DNA (Flamm, 1972) originates (and disappears). The existence of non-translated insertions in several eukaryote genes (Breathnach *et al.* 1977; Jeffreys & Flavell, 1977) suggests that such selection may not always have been able to prevent the invasion of G-DNA by S-sequences, but that mechanisms have evolved to allow normal gene action despite such invasions: it remains to be seen whether such insertions, once evolved, are merely tolerated or whether they now have a function. The considerations put forward here suggest that such insertions will be longer and possibly more frequent in high than in low C-value organisms: the gene-sized DNA in hypotrich ciliate macronuclei (Lauth *et al.* 1976) does not have room for long insertions.

The ability of S-DNA segments to increase or decrease in size through the activities of jumping genes (as well as by chance deletions and duplications, e.g. through unequal crossing over) explains how large evolutionary changes in the DNA content of the chromosomes can occur without changing the relative positions of centromeres and nucleolar organizers (Lima-da-Faria, 1973) or the linkage relationships of genes (Ohno, 1973) (Fig. 4).

In prokaryotes the absence of a nuclear envelope allows DNA and ribosomes to intermingle: even large cells therefore need no S-DNA. This explains why many blue-green algae can have cells as large as those of eukaryotes which have much higher C-

values. Since the chromosome is a single replicon, increased DNA content will greatly slow reproductive rates as well as being metabolically wasteful: selection will therefore minimize the amount of non-coding DNA in prokaryotes. The origin of the high and variable C-values of eukaryotes, and the many associated features discussed in this paper, is thus directly attributable to the evolution of the nuclear envelope (Cavalier-Smith, 1975).
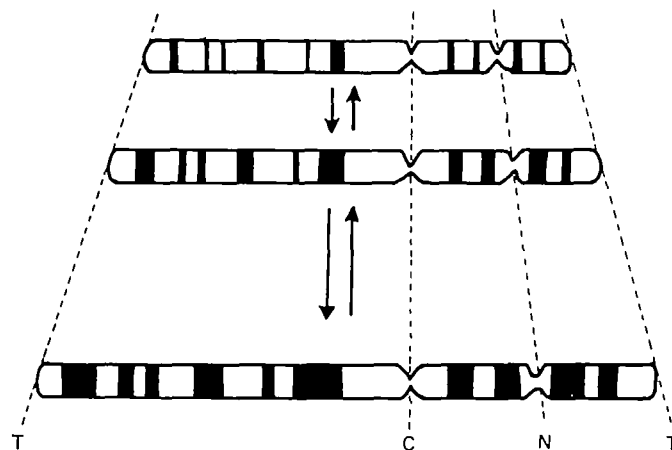

Fig. 4. Evolutionary changes in amount of S-DNA. Variation in amount of S-DNA (black) can change the DNA content and size of chromosomes without changing the amount of G-DNA (white) or the relative positions of nucleolar organizers (N) telomeres (T) and centromeres (C), or the order of the genes.

Heteroploidy and chromosome rearrangements

Many established animal cell lines show considerable increases in chromosome number despite having constant amounts of DNA (Kraemer, Deaven, Crissman & Van Dilla, 1972). DNA constancy can be explained by continuing stabilizing selection for constant nuclear volume in cell culture. Increases in chromosome number can be explained by a relaxation of selection pressure against mutations leading to fragmentation of chromosomes and rearrangement of replicons. In normal animals meiosis will regularly select against most such rearrangements: in culture such selection will be absent, except for rearrangements which separate replicons completely from K-DNA.

It seems likely that most spontaneous and induced chromosome rearrangements (the latter seem to be non-randomly located (Savage, Bigger & Watson, 1976)) occur through breaks in S-DNA, partly because S-DNA is usually more abundant than G-DNA, partly because breaks involving G-DNA would usually be lethal, but mainly because of the likelihood that S-DNA contains 'jumping genes' predisposed towards non-reciprocal recombination. Kinetochore fragmentation by means of jumping genes can explain the independent evolutionary origin of diffuse centromeres (Swanson, 1958).

Predictions

The best way of testing the ideas put forward here would be to study the C-values, nuclear sizes, DNA sequence organization, polysomal messenger complexity, and the

reproductive biology and ecology of a series of related invertebrates or algae of comparable structural complexity but of very different sizes and developmental rates.

I predict that C-values would vary, but that the complexity of the polysomal messenger would stay constant. Such studies are needed in a variety of different groups and might be especially valuable (though technically not easy) in groups which include some extreme *r*-strategists, e.g. algae, Crustacea (compare crabs and Copepods), Arachnida (compare *Limulus* and mites (Acari)), Hymenoptera (compare honeybees and hyperparasites), Diptera (compare large and minute flies).

Selection versus orthogenesis and random walk in the evolution of genome size

The evidence reviewed here argues strongly for the idea that C-values are highly adaptive and are controlled by a balance of selective forces. There is no reason to suppose that they are the mere expression of a temporary and haphazard 'balance' between an orthogenetic tendency for organisms to lose DNA and a random tendency to acquire it as suggested by Hinegardner (1976). Nor is there any reason to suppose that low amounts of DNA hinder evolution (Kimura, 1961; Kubitschek, 1974) or that high amounts hinder evolution (Bier & Müller, 1969) or cause evolutionary senescence (Fredga, 1977) or conversely evolutionary plasticity (Ohno, 1970, 1972; Hinegardner, 1976), or that major increases in DNA are needed for quantum evolution (Goin & Goin, 1968), or that DNA loop lengths exert a direct timing function (Watson, 1976): the 2-fold intraspecific variation of gymnosperm C-values suggests that the amount of DNA, far from directly controlling the rate of evolution, can easily be altered by selective forces.

These vague and often contradictory suggestions have little basis either in fact or in biological theory. A suggestion that does seem better substantiated by observations, though no theoretical explanation has been given, is that in certain groups of plants (Stebbins, 1966; Rees & Hazarika, 1969) and in fish (Hinegardner & Rosen, 1972) and other animals (Hinegardner, 1976) more specialized species tend to have lower DNA amounts. I suggest that the explanation of this is that in fish and angiosperms the species that taxonomists regard as advanced simply tend to be ones that are relatively *r*-selected: in both groups 'ancestral types' are large *K*-selected species, that would be expected to have high C-values on my hypothesis. In *Lathyrus* the data (Rees & Hazarika, 1969) clearly show that this is the case: the lower C-value species are mostly annuals and the highest C-values are perennials. I predict that a systematic study of the animal groups where specialized species are reported to have less DNA would show that minimum generation times correlate with C-values as in plants (Bennett, 1972). At present my suggestion that low C-values result from *r*-selection and high C-values from *K*-selection seems to be better supported by observations and also to be more firmly based on basic evolutionary theory and cell biological principles than other proposals. It explains the correlation between adult body size and C-value in molluscs (Hinegardner, 1974), since K-selected species have larger bodies and longer generation times (Pianka, 1970); it also explains the smaller C-values in speciose taxa – there are many more animal species in *r*-selected taxa, because small organisms experience a coarser-grained environment which therefore provides many more niches (Levins,

1968). Since it depends on individual selection and not group selection it is also preferable (Williams, 1966) to the hypothesis that postulates that apparent excess DNA is needed to speed up evolution (Edström & Lambert, 1975).

The frequent statements that DNA contents have increased over the whole evolutionary time scale are potentially highly misleading. There is no evidence for any increase in the 2000 million years or more between the first fully functional prokaryote cells and the origin of eukaryotes. Nor is there any evidence for an overall average increase in C-values in the 600 million years from the major diversification of eukaryote phyla in the late precambrian to the present day. The distribution of C-values in different organisms suggests that a major increase in DNA content occurred during or shortly after the origin of eukaryotes (Cavalier-Smith, 1975), and that the present wide range of C-values was established in response to differential r- and K-selection during the late precambrian adaptive radiation of eukaryote protists (Cavalier-Smith, 1978) even before the origin of multicellular eukaryotes. There must have been many increases and decreases in C-value during the origin and diversification of multicellular organisms but there is no evidence for an overall increase or decrease.

It is likely that there are evolutionarily significant variations in the number of protein-coding genes in different eukaryotes, but since G-DNA is probably normally only a tiny fraction of total DNA they will be hidden by the variations in C-value produced by differing levels of r- and K-selection. Moreover there is so far no good evidence that the number of protein-coding genes in different eukaryotes varies by more than a factor of 5 or 6, or that the simplest eukaryotes like yeast (Hereford & Rosbash, 1977) have very many more genes than *Escherichia coli* (Watson, 1976). I see no compelling reason why humans, for example, need even 5 times as many proteins as a unicellular photosynthetic dinoflagellate. If one accepts that most DNA is not 'genic' there would always be an excess – often a vast excess – of S-DNA for recruitment as protein-coding genes: this could, however, be quite irrelevant to the evolution of new proteins, since it would seem much easier for them to evolve by duplication of G-DNA because initiation and termination signals for both transcription and translation would already be present.

Though I have referred to S-DNA as 'non-genic' this is only shorthand for saying that it does not code for proteins – I do not wish to imply that it has no genetic function; this would be as misleading as calling heterochromatin inert: my basic argument is that S-DNA has a genetic function (the 'nucleotypic' function of controlling, in conjunction with G-DNA, nuclear and cell volumes and growth rates), but that it does not do this by coding for proteins and is therefore not divided into discrete genes: it mutates, recombines and is inherited like G-DNA, and its amount is determined by natural selection and not by genetic drift or orthogenesis. The genetic properties of DNA comprise both the protein-coding, genic, function of G-DNA and the joint nucleotypic function of both G- and S-DNA. I prefer to call the protein-coding function genic, not genotypic (Bennett, 1971, 1972), because 'genotype' in the usual meaning of the term includes both the genic and the nucleotypic functions of DNA.

Though the amount of S-DNA must be determined by selection, its sequences, on the other hand, are probably much freer than those of most G-DNA to evolve by genetic drift. However, it is unlikely that they are totally free to do so. Apart from the

general arguments of Fitch (1976) which show that the bulk of unique DNA is subject to mild stabilizing selection, certain probable mild constraints on S-RNA sequences are obvious. There would be a need for specificity to allow differential folding and unfolding (e.g. during facultative heterochromatinization, chromomere formation, and lampbrush transcription). L-RNA and HnRNA function may depend on regularities in secondary structure dependent on base pairing between inverted repeats, and must lack sequences that would cause confusion with rRNA, tRNA or mRNA (e.g. in intra-nuclear processing, transport to cytoplasm, ribosome assembly, or translation). Non-transcribed S-DNA must also lack sequences that would cause it to be transcribed or undergo recombination (or replication at the wrong place or time) or act as a kineto-chore. There could also be subsidiary functions, e.g. interaction with the synaptinemal complex, requiring a limited sequence specificity. These mild selective constraints, coupled with the fact that 21·5 % of bases in G-DNA could in principle change with-out altering amino acid sequences (Fitch, 1976), might make it difficult to distinguish S- and G-DNA by interspecific DNA hybridization. Melting temperature studies of DNA heteroduplexes have nonetheless shown that messenger-coding sequences have diverged much less between mouse and rat than have total DNA sequences (Rosbash, Campo & Gummerson, 1975). Though most S-DNA must be unique sequence DNA, much highly repetitive DNA and some middle repetitive DNA probably represents S-DNA in the process of evolution: it is noteworthy that in the pea *Lathyrus* (Narayan & Rees, 1976) and the salamander *Plethodon* (Macgregor, Mizuno & Vlad, 1976) – both high C-value genera – the species with highest C-values have the most repetitive DNA; in *Plethodon*, unlike lower C-value organisms (Galau *et al.* 1976), most repetitive sequences have diverged faster in different species than have unique sequences. The extra, disposable, DNA in flax genotrophs is also middle repetitive (Cullis, 1973).

Whether or not the distinction between S- and G-DNA will be easy or hard to make in practice, it is no longer permissible to equate genome size (whether measured by C-value (Rendel, 1966) or by the complexity of unique DNA (Searcy & MacInnis, 1970) with genetic complexity, if by this one means the number of protein-coding genes. Future discussions of genetic evolution will have to distinguish carefully and clarify the relative role of, four kinds of DNA sequences: (1) G-DNA coding for proteins, (2) G-DNA involved in control of transcription, replication and recombina-tion, (3) transcribed S-DNA, and (4) non-transcribed S-DNA. Though it is increas-ingly recognized that changes in transcriptional control may be especially important for macroevolutionary change, discussion of molecular evolution is still often confined to protein sequences. If S-DNA is as important in the determination of cell and nuclear volume and growth rates as I have suggested, then changes in S-DNA may have a large part to play in the numerous changes that accompany speciation as well as in macroevolution.

Note added in proof. Spring *et al.* (1978, *Expl Cell Res.* **114**, 203–215) suggest meiosis may after all occur in the primary nucleus of *Acetabularia*. However, their evidence that it is dip-loid not polyploid, and its high C-value compared with *Chlamydomonas*, fit my suggestion that lampbrush chromosomes are devices for increasing nuclear size in cells that cannot become polyploid.

REFERENCES

ALLEN, J. R., ROBERTS, T. M., LOEBLICH, A. R. III & KLOTZ, L. C. (1975). Characterisation of the DNA from the dinoflagellate, *Crypthecodinium cohnii* and implications for nuclear organisation. *Cell* 6, 161–169.

ALTMAN, P. L. & KATZ, D. D. (1976). *Biological Handbook I, Cell Biology*. Bethesda, Maryland: Fedn Am. Socs exp. Biol.

D'AMATO, F. (1977). *Nuclear Cytology in Relation to Development*. Cambridge, London & New York: Cambridge University Press.

BACHMANN, K., GOIN, O. B. & GOIN, C. J. Nuclear DNA amounts in vertebrates. In Smith (1972), pp. 419–447.

BAKER, B. S., CARPENTER, A. T. C., ESPOSITO, M. S., ESPOSITO, R. E. & SANDLER, L. (1976). The genetic control of meiosis. *A. Rev. Genet.* 10, 53–134.

BARIGOZZI, C. & DI PASQUALE, A. (1953). Heterochromatic and euchromatic genes acting on quantitative characters in *D. melanogaster*. *Heredity* 7, 389–399.

BENNETT, M. D. (1971). The duration of meiosis. *Proc. R. Soc.* B 178, 277–299.

BENNETT, M. D. (1972). Nuclear DNA content and minimum generation time in herbaceous plants. *Proc. R. Soc.* B 181, 109–135.

BEREZNEY, R. & COFFEY, D. S. (1977). Nuclear matrix: isolation and characterisation of a framework structure from rat liver nuclei. *J. Cell Biol.* 73, 616–637.

BIER, K. & MÜLLER, W. (1969). DNS-Messungen bei Insekten und eine Hypothese über retardierte Evolution und besonderen DNS-Reichtum in Tierreich. *Biol. Zbl.* 88, 425–449.

BISHOP, J. O. (1974). The gene numbers game. *Cell* 2, 81–86.

BLUMENTHAL, A. B., KRIEGSTEIN, H. J. & HOGNESS, D. S. (1974). The units of replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harb. Symp. quant. Biol.* 38, 205–223.

BREATHNACH, R., MANDEL, J. L. & CHAMBON, P. (1977). Ovalbumin gene is split in chicken DNA. *Nature, Lond.* 270, 314–319.

BRITTEN, R. J. & DAVIDSON, E. H. (1969). Gene regulation for higher cells: a theory. *Science, N.Y.* 165, 349–357.

BROWN, S. W. (1966). Heterochromatin. *Science, N.Y.* 151, 417–425.

BUKHARI, A. I. (1976). Bacteriophage mu as a transposition element. *A. Rev. Genet.* 10, 389–412.

CALLAN, H. G. (1967). The organization of genetic units in chromosomes. *J. Cell Sci.* 2, 1–7.

CALLAN, H. G. (1973). DNA replication in the chromosomes of Eukaryotes. In *Molecular Cytogenetics* (ed. B. A. Hamkalo & J. Papaconstantinou), pp. 31–47. New York: Plenum Press.

CAVALIER-SMITH, T. (1974). Palindromic base sequences and replication of eukaryote chromosome ends. *Nature, Lond.* 250, 467–470.

CAVALIER-SMITH, T. (1975). The origin of nuclei and of eukaryotic cells. *Nature, Lond.* 256, 463–468.

CAVALIER-SMITH, T. (1976). Long palindromes in eukaryotic DNA. *Nature, Lond.* 262, 255–256.

CAVALIER-SMITH, T. (1977). Visualising jumping genes. *Nature, Lond.* 270, 10–12.

CAVALIER-SMITH, T. (1978). The evolutionary origin and phylogeny of microtubules, mitotic spindles and eukaryote flagella. *Biosystems* 10, 93–114.

CHAPMAN, R. F. (1969). *The Insects: Structure and Function*. London: English Universities Press.

CHIKARAISHI, D. M., DEEB, S. S. & SUEOKA, N. (1978). Sequence complexity of nuclear RNAs in adult rat tissue. *Cell* 13, 111–120.

COMINGS, D. E. (1972). The structure and function of chromatin. *Adv. Hum. Genet.* 3, 237–436.

COMMONER, B. (1964). Roles of deoxyribonucleic acid in inheritance. *Nature, Lond.* 202, 960–968.

CULLIS, C. A. (1973). DNA differences between flax genotrophs. *Nature, Lond.* 243, 515–516.

DARLINGTON, C. D. (1937). *Recent Advances in Cytology*, 2nd edn. London: Churchill.

DAVIDSON, E. H. (1976). *Gene Activity in Early Development*, 2nd edn. New York: Academic Press.

DAVIDSON, E. H., KLEIN, W. H. & BRITTEN, R. J. (1977). Sequence organisation in animal DNA and a speculation on hnRNA as a coordinate regulatory transcript. *Devl Biol.* **55**, 69–84.

DOGIEL, V. A. (1965). *General Protozoology*, 2nd edn. Oxford: Clarendon Press.

DONACHIE, W. D. (1968). Relationship between cell size and time of initiation of DNA replication. *Nature, Lond.* **219**, 1077–1079.

DONACHIE, W. D. (1974). Cell division in bacteria. In *Mechanism and Regulation of DNA Replication* (ed. A. R. Kolber & M. Kohijama), pp. 431–445. New York: Plenum Press.

EDENBERG, H. J. & HUBERMAN, J. A. (1975). Eukaryotic chromosome replication. *A. Rev. Genet.* **9**, 245–284.

EDSTRÖM, J. E. & LAMBERT, B. (1975). Gene and information diversity in eukaryotes. *Prog. Biophys. molec. Biol.* **30**, 57–82.

ELLIS, J. R., MARSHALL, R., NORMAND, J. C. S. & PENROSE, L. S. (1963). A girl with triploid cells. *Nature, Lond.* **198**, 411.

ESAU, K. (1953). *Plant Anatomy.* New York: Wiley.

EVANS, G. M., DURRANT, A. & REES, H. (1966). Associated nuclear changes in the induction of flax genotrophs. *Nature, Lond.* **212**, 697–699.

EVANS, G. M., REES, H., SNELL, C. L. & SUN, S. (1972). The relationship between nuclear DNA amount and the duration of the mitotic cycle. *Chromosomes Today* **3**, 24–31.

FANKHAUSER, G., VERNON, J. A., FRANK, W. H. & SLACK, W. V. (1955). Effect of size and number of brain cells on learning in larvae of the salamander, *Triturus viridescens. Science, N.Y.* **122**, 692–693.

FANTES, P. A., GRANT, W. D., PRITCHARD, R. H., SUDBERY, P. E. & WHEALS, A. E. (1975). The regulation of cell size and the control of mitosis. *J. theor. Biol.* **50**, 213–244.

FINCHAM, J. R. S. & SASTRY, G. R. K. (1974). Controlling elements in maize. *A. Rev. Genet.* **8**, 15–50.

FITCH, W. M. (1976). Molecular evolutionary clocks. In *Molecular Evolution* (ed. F. J. Ayala), pp. 160–168. Sunderland: Sinauer Associates.

FLAMM, W. G. (1972). Highly repetitive sequences of DNA in chromosomes. *Int. Rev. Cytol.* **32**, 1–51.

FOSTER, A. S. & GIFFORD, E. M. (1974). *Morphology of Vascular Plants*, 2nd edn. San Francisco: Freeman.

FRANKE, W. W. (1974). Structure, biochemistry and functions of the nuclear envelope. *Int. Rev. Cytol.*, Suppl. **4**, 71–236.

FREDGA, K. (1977). Chromosomal changes in vertebrate evolution. *Proc. R. Soc.* B **199**, 377–397.

FRIZ, C. T. (1968). The biochemical composition of the free living amoebae *Chaos chaos, Amoeba dubia* and *Amoeba proteus. Comp. Biochem. Physiol.* **26**, 81–90.

GALAU, G. A., CHAMBERLAIN, M. E., HOUGH, B. R., BRITTEN, R. J. & DAVIDSON, E. H. (1976). Evolution of repetitive and non-repetitive DNA. In *Molecular Evolution* (ed. F. J. Ayala). Sunderland: Sinauer Associates.

GALL, J. G. (1959). Macronuclear duplication in the ciliated protozoan *Euplotes. J. biophys. biochem. Cytol.* **5**, 295–308.

GIERER, A. (1974). Molecular models and combinatorial principles in cell differentiation and morphogenesis. *Cold Spring Harb. Symp. quant. Biol.* **38**, 951–961.

GOIN, O. B. & GOIN, C. J. (1968). DNA and the evolution of vertebrates. *Am. Midland Nat.* **80**, 289–298.

GOIN, O. B., GOIN, C. J. & BACHMANN, K. (1968). DNA and amphibian life history. *Copeia* **1968**, 532–540.

GOLDSTEIN, L. (1976). Role for small nuclear RNAs in 'programming' chromosomal information? *Nature, Lond.* **261**, 519–521.

GRANT, W. F. (1976). The evolution of karyotype and polyploidy in arboreal plants. *Taxon* **25**, 75–84.

GURDON, J. B. & WOODLAND, H. R. (1968). The cytoplasmic control of nuclear activity in animal development. *Biol. Rev.* **43**, 233–267.

HARRIS, H. (1970). *Nucleus and Cytoplasm*, 2nd edn. Oxford: Clarendon Press.

HAYES, S. & SZYBALSKI, W. (1973). Synthesis of RNA primer for lambda DNA replication is controlled by phage and host. In *Molecular Genetics* (ed. B. A. Hamkalo & J. Papaconstantinou), pp. 277–283. New York: Academic Press.

HEREFORD, L. M. & ROSBASH, M. (1977). Number and distribution of polyadenylated RNA sequences in yeast. *Cell* **10**, 453-462.

HILL, J. (1976). The environmental induction of heritable changes in *Nicotiana rustica* parental and selection lines. *Genetics, Princeton* **55**, 735-754.

HINEGARDNER, R. (1974). Cellular DNA content of the mollusca. *Comp. Biochem. Physiol.* **47A**, 447-460.

HINEGARDNER, R. (1976). Evolution of genome size. In *Molecular Evolution* (ed. F. J. Ayala), pp. 179-199. Sunderland: Sinauer.

HINEGARDNER, R. & ROSEN, D. E. (1972). Cellular DNA content and the evolution of teleost fishes. *Am. Nat.* **106**, 621-644.

HOLLIDAY, R. (1970). The organization of DNA in eukaryotic chromosomes. *Symp. Soc. gen. Microbiol.* **20**, 359-380.

HOLM-HANSEN, O. (1969). Algae: amounts of DNA and organic carbon in single cells. *Science, N.Y.* **163**, 87-88.

HUDSPETH, M. E. S., TIMBERLAKE, W. E. & GOLDBERG, R. B. (1977). DNA sequence organization in the water mold *Achlya. Proc. natn. Acad. Sci. U.S.A.* **74**, 4332-4336.

HYMAN, J. H. (1951). *The Invertebrates*, vol. 3. New York & London: McGraw-Hill.

JACOBSON, A. & LODISH, H. F. (1975). Genetic control of development of the cellular slime mold, *Dictyostelium discoideum. A. Rev. Genet.* **9**, 145-185.

JEFFREYS, A. J. & FLAVELL, R. A. (1977). The rabbit β-globin gene contains a large insert in the coding sequence. *Cell* **12**, 1097-1108.

JELINEK, W., MOLLOY, G., SALDITT, M., WALL, R., SHEINESS, D. & DARNELL, J. E. (1974). Origin of mRNA in HeLa cells and the implications for chromosome structure. *Cold Spring Harb. Symp. quant. Biol.* **38**, 891-898.

JONES, R. N. (1975). B-chromosome systems in flowering plants and animal species. *Int. Rev. Cytol.* **40**, 1-100.

JONES, R. N. (1976). Genome organisation in higher plants. *Chromosomes Today* **5**, 117-130.

KAUFFMAN, S. (1971). Gene regulation networks: A theory for their global structures and behaviours. In *Current Topics in Developmental Biology* (ed. A. A. Moscona & A. Monroy), pp. 145-182. New York: Academic Press.

KEYL, H. G. (1965). A demonstrable local and geometric increase in the chromosomal DNA of *Chironomus. Experientia* **21**, 191-193.

KIMURA, M. (1961). Natural selection as the process of accumulating genetic information in adaptive evolution. *Genet. Res.* **2**, 127-140.

KRAEMER, P. M., DEAVEN, L. L., CRISSMAN, H. A. & VAN DILLA, M. A. (1972). DNA constancy despite variability in chromosome number. *Adv. Cell molec. Biol.* **2**, 47-108.

KUBITSCHEK, H. E. (1974). Operation of selection pressure on microbial populations. *Symp. Soc. gen. Microbiol.* **24**, 105-130.

LAUTH, M. R., SPEAR, B. B., HEUMANN, J. & PRESCOTT, D. M. (1976). DNA of ciliated protozoa: DNA sequence diminution during macronuclear development of *Oxytricha. Cell* **7**, 67-74.

LEVINS, R. (1968). *Evolution in Changing Environments*. Princeton: Princeton University Press.

LEWIN, B. (1974). *Gene Expression*, vol. 2. London: Wiley.

LIMA-DA-FARIA, A. (1973). Equations defining the position of ribosomal cistrons in the eukaryotic chromosome. *Nature, New Biol.* **241**, 136-139.

LIST, A. (1963). Some observations on DNA content and cell volume growth in the developing xylem cells of certain higher plants. *Am. J. Bot.* **50**, 320-329.

MACARTHUR, R. H. & WILSON, E. O. (1967). *The Theory of Island Biogeography*. Princeton: Princeton University Press.

MACGREGOR, H. C., MIZUNO, S. & VLAD, M. (1976). Chromosomes and DNA sequences in salamanders. *Chromosomes Today* **5**, 331-339.

MACLEAN, N. (1973). Suggested mechanism for increase in size of genome. *Nature, New Biol.* **246**, 205-206.

MARTIN, P. G. (1966). Variation in the amounts of nucleic acids in the cells of different species of higher plants. *Expl Cell Res.* **44**, 84-94.

MATHER, K. (1943). Polygenic inheritance and natural selection. *Biol. Rev.* **18**, 32-64.

MATHER, K. (1944). The genetical activity of heterochromatin. *Proc. R. Soc. B* **132**, 308-332.

MATHER, K. (1949). *Biometrical Genetics*. London: Methuen.

MAUL, G. G. & DEAVEN, L. (1977). Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. *J. Cell Biol.* **73**, 748–760.

MAYR, E. (1970). *Population, Species and Evolution.* Cambridge, Massachusetts: Belknap-Harvard.

McCLINTOCK, B. (1951). Chromosome organisation and genic expression. *Cold Spring Harb. Symp. quant. Biol.* **16**, 13–47.

McCLINTOCK, B. (1967). Genetic systems regulating gene expression during development. *Devl Biol.*, Suppl. 1, 84–112.

MITTWOCH, U. (1968). Nuclear sizes in a human diploid/triploid cell culture. *Nature, Lond.* **219**, 1074–1076.

NARAYAN, R. K. J. & REES, H. (1976). Nuclear DNA variation in *Lathyrus. Chromosoma* **54**, 141–154.

OHNO, S. (1970). *Evolution by Gene Duplication.* London: Allen & Unwin.

OHNO, S. (1972). An argument for the genetic simplicity of man and other mammals. *J. Hum. Evol.* **1**, 651–662.

OHNO, S. (1973). Ancient linkage groups and frozen accidents. *Nature, Lond.* **244**, 259–262.

OLMO, E. & MORESCALCHI, A. (1976). DNA and cell size in salamanders. *Chromosomes Today* **5**, 467.

PERLMAN, S., PHILLIPS, C. & BISHOP, J. O. (1976). A study of foldback DNA. *Cell* **8**, 33–42.

PIANKA, E. R. (1970). On r- and K-selection. *Am. Nat.* **104**, 592–597.

PIERCE, W. P. (1937). The effect of phosphorus on chromosome and nuclear volume in a violet species. *Bull. Torrey bot. Club.* **64**, 345–356.

POLANI, P. E. (1969). Abnormal sex chromosomes and mental disorder. *Nature, Lond.* **223**, 680–686.

PRESCOTT, D. M., STEVENS, A. R. & LAUTH, M. R. (1971). Characterisation of nuclear RNA synthesis in *Amoeba proteus. Expl Cell Res.* **64**, 145–155.

PRICE, H. J., SPARROW, A. H. & NAUMAN, A. F. (1973). Correlations between nuclear volume, cell volume and DNA content in meristematic cells of herbaceous angiosperms. *Experientia* **29**, 1028–1029.

RAPER, J. R. & FLEXER, A. S. (1970). The road to diploidy with emphasis on a detour. *Symp. Soc. gen. Microbiol.* **20**, 401–432.

RASMUSSON, B., GREEN, M. M. & CARLSON, B. M. (1974). Genetic instability in *Drosophila melanogaster*: evidence for insertion mutations. *Molec. gen. Genet.* **133**, 237–248.

REES, H. (1972). DNA in higher plants. In *Evolution of Genetic Systems, Brookhaven Symp. Biol.* **26**, 394–418.

REES, H. & HAZARIKA, M. H. (1969). Chromosome evolution in *Lathyrus. Chromosomes Today* **2**, 158–165.

REES, H. & JONES, R. N. (1972). The origin of the wide species variation in nuclear DNA content. *Int. Rev. Cytol.* **32**, 53–92.

RENDEL, J. M. (1966). The effects of genetic change at different levels. *Ideas in Modern Biology*, Proc. 6 XVI int. Congr. Zool. (ed. J. A. Moore), pp. 285–295. New York: National History Press.

RICHARDS, O. W. & DAVIES, R. G. (1977). *Imm's General Textbook of Entomology.* London: Chapman & Hall.

ROBERTS, L. W. (1976). *Cytodifferentiation in Plants: Xylogenesis as a Model System.* London & New York: Cambridge University Press.

ROSBASH, M., CAMPO, M. S. & GUMMERSON, K. S. (1975). Conservation of cytoplasmic poly (A) – containing RNA in mouse and rat. *Nature, Lond.* **258**, 682–686.

ROUGHGARDEN, J. (1971). Density-dependent natural selection. *Ecology* **55**, 453–468.

SAVAGE, J. R. K., BIGGER, T. R. L. & WATSON, G. E. (1976). Location of quinacrine mustard – induced chromatid exchange points in relation to ASG bands in human chromosomes. *Chromosomes Today* **5**, 281–291.

SCHERRER, K. (1974). Control of gene expression in animals: the cascade regulation hypothesis revisited. In *Control of Gene Expression* (ed. A. Kohn & Shatky), pp. 169–219. London: Plenum Press.

SCHINDLER, A. M. & MIKAMO, K. (1970). Triploidy in man. Report of a case and a discussion on etiology. *Cytogenetics* **9**, 116–130.

SEARCY, D. G. & MACINNIS, A. J. (1970). Measurements by DNA renaturation of the genetic basis of parasitic reduction. *Evolution* 24, 796–806.

SMITH, G. P. (1976). Evolution of repeated sequences by unequal crossover. *Science* 191, 528–535.

SMITH, H. H. (ed.) (1972). *Evolution of Genetic Systems*. New York: Gordon & Breach.

SMITH, J. B. & BENNETT, M. D. (1975). DNA variation in the genus *Ranunculus*. *Heredity* 35, 231–239.

SOKUROVA, E. N. (1973). The content of nucleic acids in the cells of yeasts belonging to various taxonomic groups (in Russian). *Mikrobiologiya* 42, 1020–1024.

SOMPYRAC, L. & MAALOE, O. (1973). Autorepressor model for control of DNA replication. *Nature, New Biol.* 241, 133–135.

SPARROW, A. H., PRICE, H. J. & UNDERBRINK, A. G. (1972). In Smith (1972), 451–493.

SPRING, H., SCHEER, U., FRANKE, W. W. & TRENDELENBURG, M. F. (1975). Lampbrush-type chromosomes in the primary nucleus of the green alga *Acetabularia mediterranea*. *Chromosoma* 50, 25–43.

STANIER, R. Y. & COHEN-BAZIRE, G. (1977). Phototrophic prokaryotes: the cyanobacteria. *A. Rev. Microbiol.* 31, 225–274.

STEBBINS, G. L. (1966). Chromosomal variation and evolution. *Science, N.Y.* 152, 1463–1469.

SUMMERBELL, D., LEWIS, J. H. & WOLPERT, L. (1973). Positional information in chick limb morphogenesis. *Nature, Lond.* 244, 492–496.

SWANSON, C. P. (1958). *Cytology and Cytogenetics*. London: Macmillan.

SZARSKI, H. (1974). Cell size and nuclear DNA content in vertebrates. *Int. Rev. Cytol.* 44, 93–111.

TAMM, I. (1977). Definition of subclasses of nucleoplasmic RNA. *Proc. natn. Acad. Sci. U.S.A.* 74, 5011–5045.

TAYLOR, J. H. (1974). Units of DNA replication in chromosomes of eukaryotes. *Int. Rev. Cytol.* 37, 1–20.

THOMAS, C. A. (1971). The genetic organisation of chromosomes. *A. Rev. Genet.* 5, 237–256.

THOMPSON, J. N. (1975a). Quantitative variation and gene numbers. *Nature, Lond.* 258, 665–668.

THOMPSON, J. N. (1975b). Evidence for polygenes. *Nature, Lond.* 261, 526.

THURIAUX, P. (1977). Is recombination confined to structural genes on the eukaryotic genome ? *Nature, Lond.* 268, 460–462.

VAN'T HOF, J. (1965). Relationships between mitotic cycle duration, S period duration and the average rate of DNA synthesis in the root meristem cells of several plants. *Expl Cell Res.* 39, 48–58.

VAN'T HOF, J. & SPARROW, A. H. (1963). A relationship between DNA content, nuclear volume, and minimum mitotic cycle time. *Proc. natn. Acad. Sci. U.S.A.* 49, 897–902.

VETTA, A. (1975). Evidence for polygenes. *Nature, Lond.* 261, 525–526.

WADDINGTON, C. H. (1939). *An Introduction to Modern Genetics*. London: Allen & Unwin.

WAHRMAN, J., GOITEN, R., RICHLER, C., GOLDMAN, B., AKSTEIN, E. & CHAKI, R. (1976). The mongoloid phenotype in man is due to trisomy of the distal pale G-band of chromosome 21. *Chromosomes Today* 5, 241–248.

WATSON, J. D. (1976). *Molecular Biology of the Gene*, 3rd edn. London: Benjamin.

WHITE, M. J. D. (1973). *Animal Cytology and Evolution*, 3rd edn. London & New York: Cambridge University Press.

WILBY, O. K. & EDE, D. A. (1975). A model generating the pattern of cartilage skeletal elements in the embryonic chick limb. *J. theor. Biol.* 52, 199–217.

WILLIAMS, G. C. (1966). *Adaptation and Natural Selection*. Princeton: Princeton University Press.

WILSON, B. F. (1964). A model for cell production by the cambium of conifers. In *The Formation of Wood in Forest Trees* (ed. M. H. Zimmerman), pp. 19–36. London: Academic Press.

WILSON, B. F. (1966). Mitotic activity in the cambial zone of *Pinus strobus*. *Am. J. Bot.* 53, 364–372.

YUNIS, J. J. & YASMINEH, W. G. (1972). Model for mammalian constitutive heterochromatin. *Adv. Cell molec. Biol.* 2, 1–46.