STADLER SYMP. Vol.7 (1975) University of Missouri, Columbia - 47

GENOME ORGANIZATION IN HIGHER PLANTS

(DNA, higher plants, repeated sequence, genome organization)

RICHARD B. FLAVELL and DEREK B. SMITH

Department of Cytogenetics
Plant Breeding Institute, Trumpington
Cambridge, CB2 2LQ England

SUMMARY

Higher plant chromosomes contain many families of thousands of similar nucleotide sequences as well as sequences present in only one or a few copies per haploid genome. We summarise here biochemical investigations into the linear chromosomal organisation of the different kinds of repeated and non-repeated sequences in the DNA of wheat. Some models of the fine structure organisation of common sequence arrangements are presented. The possible roles of these sequence arrangements in gene expression and chromosome behaviour are discussed.

INTRODUCTION

Biochemical and other kinds of studies on genetic variants in bacteria and particularly bacteriophage have enabled considerable information and understanding to be gained of the linear organization of the genes and their controlling sequences. In some instances much of the DNA can be accounted for in terms of gross function. Ti gain such extensive knowledge from studies of genetic variants in higher plants and animals is out of the question. There is far too much DNA in the chromosomes of higher organisms and much of it probably consists of nucleotide sequences in which mutations are not phenotypically expressed and therefore not available for genetic analysis. However, over the last few years, new biochemical and molecular approaches have been made in chromosomal DNA studies, which are beginning to provide detailed information on the organisation and function of the large amount of DNA in the chromosomes of higher organisms.

The chromosomes of *enopus* (DAVIDSON *et al.* 1973a, b), sea urchin (GRAHAM *et al.* 1974), *Drosophila* (LAIRD *et al.* 1973) and rat (BONNER *et al.* 1973) featured in the early studies and certain striking patterns of nucleotide sequence organisation

are common to these organisms. Our laboratory is particularly concerned with the nucleotide sequence organisation of higher plant chromosomes and it is the main conclusions from these studies that we wish to present in this contribution. Although we have made preliminary investigations on a wide range of plant species, our more detailed experiments have been made on members of the Gramineae, particularly common bread wheat, *Triticum aestivum* and rye, *Secale cereale*.

REPEATED AND NON-REPEATED SEQUENCES IN PLANT CHROMOSOMES

Biochemical analyses of the complex genomes of higher organisms were pioneered by BRITTEN and KOHNE (1965) who showed that the kinetic and genetic complexities of genomes could be estimated from the renaturation kinetics of denatured chromosomal DNA, sheared to small fragments. Such studies demand methods for quantitatively distinguishing between renatured duplex DNA and single stranded DNA. This can be achieved by column chromatography on hydroxyapatite (BERNARDI 1969). At a low concentration of phosphate (0.12 M) DNA fragments containing double stranded regions of approximately thirty or more paired bases, bind to hydroxyapatite while single stranded DNA does not. At high phosphate concentrations, (0.5 M), double stranded DNA can be recovered from the hydroxyapatite.

We have used hydroxyapatite fractionation extensively in our studies. The renaturation kinetics of denatured wheat DNA, sheared by sonication to an average fragment size of 470 nucleotides and incubated at 60°C in 0.12 M phosphate buffer is illustrated in Figure 1. The proportion of renatured DNA, determined by hydroxyapatite chromatography, is plotted against C_0t [concentration of DNA (moles nucleotide/litre) x incubation time (seconds)] on a log scale. The rate of DNA renaturation is proportional to sequence concentration (BRITTEN and KOHNE 1965). Thus sequences present in many copies find their complementary sequences and reanneal more rapidly than sequences present in only one copy per haploid genome. From knowledge of the genome size of *Escherichia coli* (4.5×10^6 nucleotide pairs) and of hexaploid wheat (34.9×10^9 nucleotide pairs), non-repeated sequences in wheat would not be expected to reanneal until a C_0t value in excess of 1000, given that *E. coli* DNA, consists entirely of non-repeated sequences (see Figure 1). The immediate conclusion from the DNA renaturation kinetics is that approximately 80% of wheat DNA renatures rapidly and must consist of highly repeated sequences (BENDICH and McCARTHY 1970, SMITH and FLAVELL 1974, 1975).

More detailed studies have indicated that the repeated sequence fraction of wheat DNA consists of sequences present in approximately 200 to 200,000 copies with a mean of approximately 10,000 per 1*C* genome (SMITH and FLAVELL 1975).

Wheat is not unusual in having such a high proportion of repeated sequence DNA. In a survey of twenty three higher

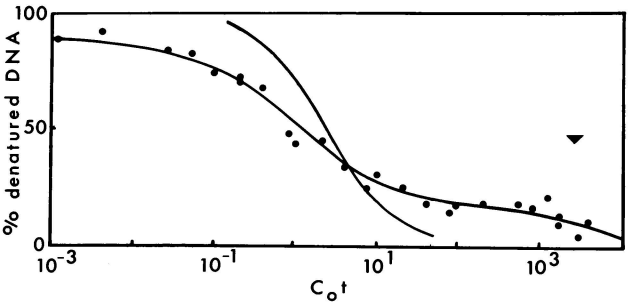

FIGURE 1. Renaturation kinetics of wheat and *Escherichia coli* DNAs. Wheat DNA, average single stranded fragment size 470 nucleotides containing a low concentration of ^3H labelled *Escherichia coli* DNA was heat denatured and incubated in 0.12 M phosphate buffer at 60°C to various 'C_0t' values before being fractionated into reannealed duplex and denatured fractions on hydroxyapatite. The arrow indicates the C_0t value when half of the DNA present in one copy per 1C wheat genome would be expected to have reannealed.

●▬● = wheat ▬▬ = *Escherichia coli*

plant species including monocotyledonous and dicotyledonous species with 2*C* DNA contents ranging from 1.5 to 98 pg (see Table 1) the mean percentage of repeated sequence DNA was found to be 74 ± 2.4% (FLAVELL *et al.* 1974). For many plant DNAs it is difficult to obtain accurate estimates of the renaturation kinetics of the sequences present in fewer than five copies. The small proportion of the total DNA represented by these sequences and the relatively long incubation times necessary for near complete renaturation, cause technical difficulties. However, by isolation of the non-repeated sequence fraction and studying its renaturation kinetics in a series of careful experiments, we have concluded that approximately 12 to 20% of the wheat genome consists of sequences present in six copies per hexaploid cell or one copy per constituent haploid genome (SMITH and FLAVELL 1975). This result indicates that the single copy sequences of the constituent diploids of hexaploid wheat are nearly identical, and gives support to the hypothesis that the wheat chromosome contains a single DNA helix, rather than many helices. The existence of sequences present in one copy per haploid constituent genome fulfills a prediction of the geneticist. Genes present in one copy per haploid chromosome set would, on simplest models, consist of a nucleotide sequence present in only one copy per haploid genome. It is therefore reasonable to conclude that the structural coding genes and probably many control genes are included in the non-repeated sequence fraction of the genome but how much of this fraction consists of structural genes or

coding sequences is not know. However, the proportion of the chromosomal DNA likely to be involved in coding for proteins is very small. Some proteins, for example the histones (KEDES and BIRNSTIEL 1971), and some RNAs, for example ribosomal

Table 1. Estimates of the proportion of repeated-sequence DNA[a] in species with a nuclear DNA mass between 1.5 and 98 pg.

Species	Ploidy	2C nuclear DNA content (pg)	Proportion of repeated sequences (%)
Linum usitatissimum	2x	1.5	59
Capsella bursa-pastoris	4x	1.7	46
Veronica persica	4x	1.9	63
Stellaria media	7x	2.5	69
Lamium purpureum	2x	2.7	60
Senecio vulgaris	8x	3.5	74
Daucus carota	2x	2.1	62
Beta vulgaris	2x	2.7	63
Poa trivialis	2x	6.9	82
Tropaeolum majus	2x	7.3	70
Pisum sativum	2x	9.9	75
Helianthus annuus	2x	10.7	69
Zea mays	2x	11.0	78
Hordeum vulgare	2x	13.4	76
Poa annua	4x	13.8	87
Triticum monococcum	2x	14.0	80
Secale cereale	2x	18.9	92
Vicia faba	2x	29.3	85
Allium cepa	2x	33.5	95
Triticum aestivum	6x	36.2	83
Avena sativa	6x	43.0	83
Tulipa kaufmanniana	2x	62.5	73
Hyacinth orientalis	4x -1	98.1	75

[a]Proportion of DNA sequences present in an excess of ten copies as indicated by hydroxyapatite chromatography.

(BIRNSTIEL, CHIPCHASE and SPIERS 1971), 5s (BROWN and SUGIMOTO 1973) and transfer (CLARKSON and BIRNSTIEL 1973) RNA are specified by repeated sequences but these are almost certainly exceptions.

Knowledge of the presence of many highly repetitious and a small proportion of non-repeated sequences immediately raises many questions. The most significant question is, of course, what are the function(s) of all the different kinds of sequences in the genome. It is a hope that studies on the molecular arrangement and evolution of the different

kinds of sequences will lead to an understanding of the functions of the different kinds of nucleotide sequences which constitute higher plant chromosomal DNAs.

THE SIZE AND ARRANGEMENT OF NON-REPEATED SEQUENCES

The arrangement of non-repeated sequences can be gained by estimating the proportion of DNA retained by hydroxyapatite in the renatured fraction after incubating different size fragments to a C_0t value at which all the repeated sequences have reannealed but none of the non-repeated sequences have started to reanneal (see Figure 1). If non-repeated sequences are interspersed between repeated sequences then the non-repeated sequence DNA exists as single stranded tails on duplexes formed between long fragments carrying short repeated sequences. This is illustrated in Figure 2. With increasing fragment size, the size of the single stranded tail increases, increasing the proportion of the genome retained by hydroxyapatite in the renatured DNA fraction. When the fragment length exceeds the length of the non-repeated sequence there is no further increase in the proportion of hydroxyapatite retained DNA, because the repeated sequence DNA at either end of the non-repeated sequence forms duplexes and is retained by hydroxyapatite with very short fragments (see DAVIDSON *et al.* 1973a; GRAHAM *et al.* 1974).

The relationship between average single stranded fragment size and the proportion of wheat DNA included in the renatured DNA fraction after incubation to C_0t 50 is shown in Figure 3. To simplify experimental design and interpretation, ^{3}H-labelled DNA of different average fragment sizes were renatured with a large excess of unlabelled DNA of a constant size (DAVIDSON *et al.* 1973a). The results imply that approximately 16% of the wheat genome consists of non-repeated sequences around 1000 nucleotides long interspersed between repeated sequences (see DAVIDSON *et al.* 1973a, b). It is particularly interesting to note that similar lengths of non-repeated sequences interspersed between repeated sequences have been found in other organisms including *Xenopus* (DAVIDSON *et al.* 1973a, b), sea urchin (GRAHAM *et al.* 1974), rat (BONNER *et al.* 1973) and *Dictiostelium* (LODISH, FIRTEL and JACOBSON 1973). The size of the sequence coding for an average protein would be expected to be about 800 to 1500 nucleotides long. It is possible therefore that this class of non-repeated sequence may include the coding sequences of the genome (see later).

The 9% non-repeated sequences not included in the repeated sequence fraction on hydroxyapatite even with very long starting DNA fragments (Figure 3) must consist of long non-repeated sequences, many thousands of nucleotides from a repeated sequence. A very similar arrangement with similar proportions of non-repeated sequences has been found in oats, rye and *Vicia faba* (FLAVELL, SMITH and RIMPAU, unpublished).

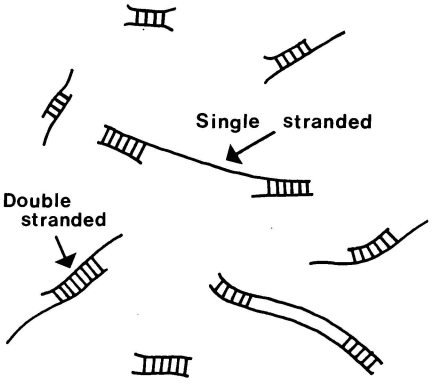

FIGURE 2. Duplex formation involving long and short single stranded DNA fragments where repeated sequences are interspersed with non-repeated sequences. Renaturation is to a C_0t value which allows only repeated sequences to reanneal. All the DNA illustrated, including the single stranded 'tails', would be retained by hydroxyapatite in the renatured DNA fraction.

Extrapolation of the curve in Figure 3 to very small fragments most of which would consist of either only repeated or only non-repeated sequences indicates that 75% of the wheat genome consists of repeated sequences.

EVIDENCE FOR PALINDROME SEQUENCES

For Hela, *Xenopus*, *Drosophila* and *Triturus* DNAs, WILSON and THOMAS (1974) have presented evidence for the existence of sequences with inverted complementary sequences as neighbours. These palindromic sequences, when not base paired with the other strand of the double helix, can pair with each other as shown in Figure 4. The intrastrand base pairing takes place extremely rapidly at very low DNA concentrations since it is not a bimolecular reaction, in contrast to interstrand hybridisation. Thus the formation of duplex DNA due to palindrome sequence association can be measured by hydroxyapatite chromatography immediately after denaturation of DNA at very low concentration.

The proportion of wheat DNA retained by hydroxyapatite immediately after melting very dilute solutions of ^{3}H-labelled DNA of different average fragment sizes is shown in Figure 5. The proportion bound to hydroxyapatite increases with fragment size almost certainly due to non-duplex sequences being retained as single stranded 'tails' on the hairpin duplexes

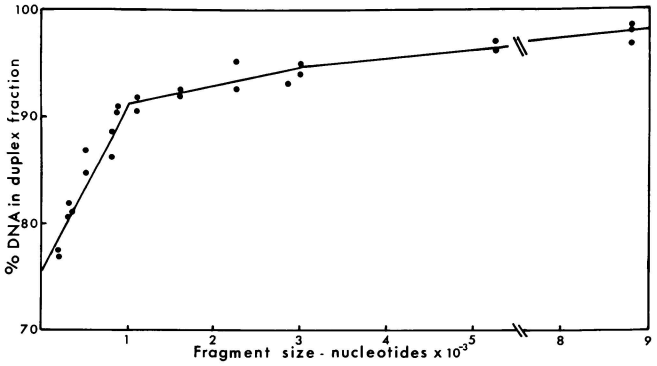

FIGURE 3. Interspersion of non-repeated sequences in repeated sequences. The proportion of ^{3}H-labelled wheat DNA included in the renatured DNA fraction after denaturation and renaturation of different size fragments to C_0t 50 at 60°C in 0.12 M phosphate buffer was determined by hydroxyapatite chromatography. The increase in the renatured DNA fraction up to a fragment size of 1000 nucleotides represents non-repeated sequences interspersed between repeated sequences.

(see Figure 2). This is supported by the fact that these additional sequences are degraded by the single strand specific nuclease S_1 from *Aspergillus* (ANDO 1966, SUTTON 1971, SMITH and FLAVELL 1975). Thus the sequences responsible for the zero-time binding (DAVIDSON 1973a) of denatured DNA to hydroxyapatite are spaced throughout at least 20% of the wheat genome. The linear relationship between the amount of zero-time binding DNA and fragment size indicates a regular spacing of groups of palindromes (HAMER and THOMAS 1974). Extrapolation of the curve to very small fragment sizes indicates that 3 to 5% of the wheat genome consists of zero-time binding sequences. Since the extrapolated line cuts the y axis at a positive value, the zero-time binding sequences are probably clustered in groups (see HAMER and THOMAS 1974, for the theory of palindrome distribution). Although the zero-time binding DNA in wheat could be duplex DNA formed because of interstrand cross linking, it is highly probable that most of it consists of hairpin structures from palindromes (see Figure 4) as in mammals (see BONNER 1974).

The existence of regularly spaced, clustered groups of palindrome sequences is an extremely provocative finding. They surely have a function, though what this is is speculation at present. Perhaps they act as recognition signals for initiation of replication. The inverted sequences appear to

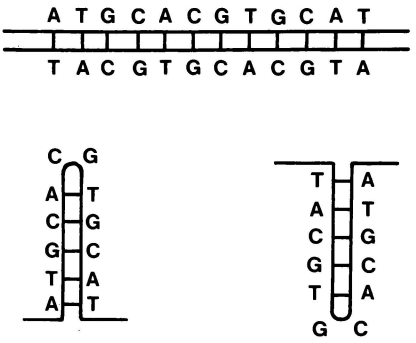

FIGURE 4. A palindrome sequence in a chromosome. After denaturation, intrastrand base pairing can occur essentially instantaneously to produce the hairpin duplex structures illustrated. These duplexes can be isolated by hydroxyapatite chromatography as a zero-time binding fraction. With long starting fragments the hairpin duplexes may have very long single stranded 'tails' (see Figure 2).

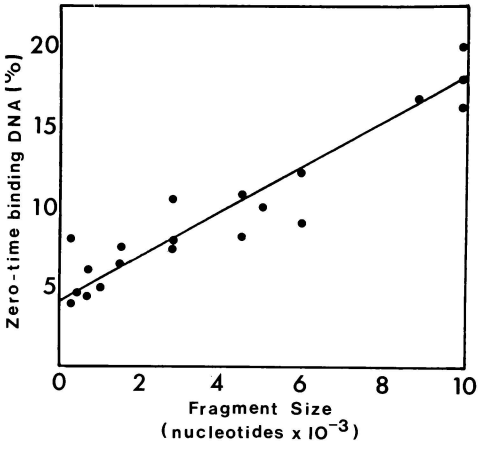

FIGURE 5. Relationship between the proportion of zero-time hydroxyapatite binding DNA and average single stranded fragment size. ^{3}H-labelled wheat DNA was denatured at 0.026 µg/ml and immediately chromatographed on hydroxyapatite at 60°C in 0.12 M phosphate buffer ($C_0t < 2x10^{-7}$). ^{3}H-labelled DNA average single stranded fragment sizes were determined by alkaline sucrose density gradient centrifugation.

have been highly conserved since the intrastrand duplexes are extremely stable to heat. Of great interest is whether intrastrand base pairing rather than interstrand base pairing exists in the chromosome *in vivo*.

THE SIZES AND ARRANGEMENTS OF THE REPEATED SEQUENCES

We have already indicated that approximately 75% of the wheat genome consists of repeated sequence DNA, as judged by the renaturation kinetics of fragments carrying no non-repeated DNA (see Figure 3). Of this, approximately 4% consists of zero-time binding (palindrome) sequences. Thus approximately 71% of the wheat genome consists of repeated sequences not organised into palindromes.

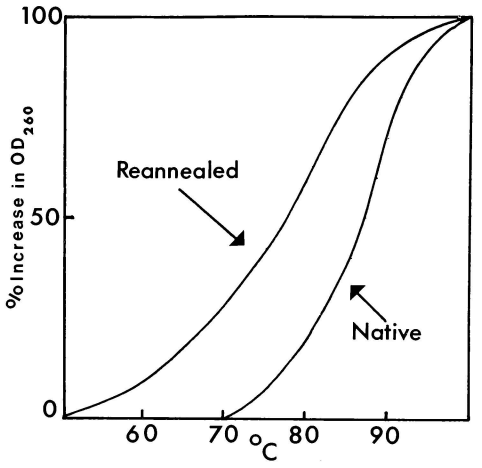

FIGURE 6. Thermal stabilities of wheat native DNA and repeated sequence DNA reannealed at 60°C in 0.12 M phosphate buffer. The melting of DNA was measured in a spectrophotometer by following the increase in optical density at 260 nm with increasing temperature.

When denatured, repeated sequence DNA is renatured at 60°C in 0.12 M phosphate buffer, the formed duplexes have a considerably reduced thermal stability compared with native DNA (Figure 6). This indicates that the repeated sequences are not identical, but consist of families of similar, related sequences. It is likely that some families consist of nearly identical sequences while others have diverged considerably and that each family has evolved from a single ancestral sequence (BRITTEN and KOHNE 1965).

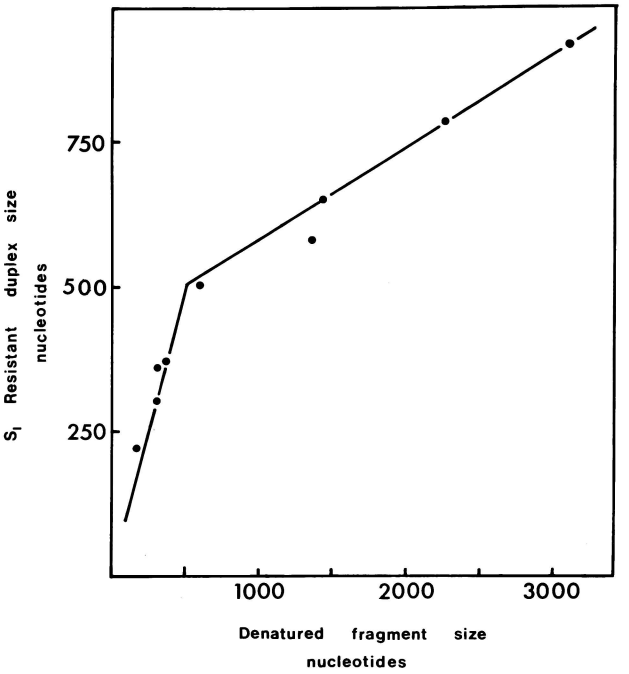

FIGURE 7. Dependence of the S_1 nuclease-resistant repeated sequence duplex size on the average denatured DNA fragment size after reannealing. Approximately 55% of the wheat genome is resistant to S_1 after reannealing, independent of the starting size. The S_1-resistant duplexes were recovered by hydroxyapatite chromatography after S_1 enzyme treatment as described by DAVIDSON *et al.* 1973b. Average fragment nucleotide sizes were estimated by analytical ultracentrifugation as described by STUDIER (1965).

THE SIZE OF S_1 NUCLEASE RESISTANT REPEATED SEQUENCE DUPLEXES

Highly mismatched duplexes, as well as long single stranded DNA regions are degraded by S_1 nuclease under the assay conditions we have used. Approximately 55% of wheat DNA is resistant to S_1 nuclease after renaturation to C_0t 50 in 0.12 M phosphate buffer at 60°C. Thus 27% of the repeated sequence duplexes (the least well base-paired) are degraded by the nuclease. Most of the S_1 resistant duplexes are between 500 and 600 nucleotides long. This information was gained from a study of the size of the S_1 resistant duplexes using different sizes of starting fragments for denaturation and incubation to C_0t 50 (Figure 7). With a starting average

fragment size below 500 nucleotides, the average S_1 resistant duplex size is similar. However, above 500 nucleotides, the S_1 resistant size is much smaller than the starting fragment size.

The two slopes in Figure 7 indicate that much of the wheat genome (85% of the S_1 resistant DNA) reanneals with S_1 sensitive sites spaced 500-600 nucleotides apart but some reanneal to form much longer S_1 resistant duplexes.

This latter DNA accounts for the small increase in the average S_1 resistant duplex size with starting fragment sizes above 500 nucleotides (Figure 7). The implications of these results for repeated sequence sizes and organisation are presented later.

HYPERPOLYMERS AND THE SIZE AND ARRANGEMENT OF REPEATED SEQUENCES

After renaturation of repeated sequence DNA, particularly with long average starting fragment sizes, very large networks of DNA form. These networks, or hyperpolymers (GRAHAM and BRITTEN 1973), are due to secondary hybridisations between single stranded tails of already formed duplexes. This is illustrated in Figure 8. Thus the kinetics of hyperpolymer formation provide an indicator of single stranded tail frequency. Furthermore, if the incubation times are too short for non-repeated sequences to reanneal, then the single

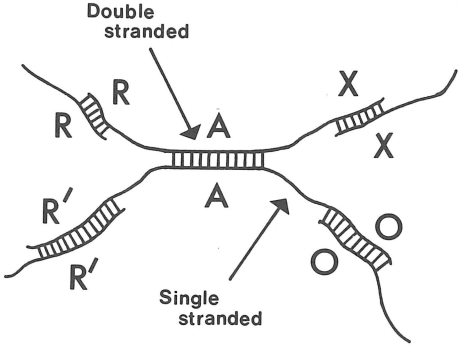

FIGURE 8. Hyperpolymer formation. Closely related or identical short repeated sequences, A, form the initial duplex. If these A sequences, coming from different places in the genome, have unrelated repeated sequences (O and X) as neighbours or evolutionarily related but diverged sequences (R and R') as neighbours then single stranded 'tails' will occur which will take part in secondary hybridisations to form large networks or hyperpolymers.

stranded tail hybridisations must also involve repeated sequences. Hyperpolymer formation indicates therefore the presence of more than one repeated sequence per fragment (GRAHAM and BRITTEN 1973).

If similar or identical repeated sequences have different repeated sequences as neighbours at different places in the chromosome, then four single stranded tails will result after the initial hybridisation event between them if the fragment lengths are long compared with the duplex repeated sequence length. This is also illustrated in Figure 8. It is not difficult to see how large networks result from secondary hybridisations on each of the four tails.

The kinetics of hyperpolymer formation with different average starting fragment sizes is shown in Figure 9. Hyperpolymer formation was assayed by the inability of hyperpolymers to be eluted from hydroxyapatite without being melted (FLAVELL and SMITH 1975). The rate of hyperpolymer formation is slow until the average starting fragment size exceeds 500-600 nucleotides. This indicates that there are few single

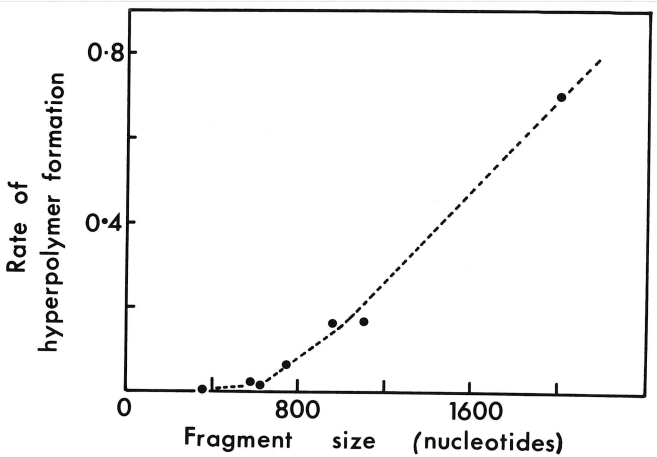

FIGURE 9. Rate of hyperpolymer formation with different average fragment sizes. The kinetics of hyperpolymer formation were determined at 60°C in 0.12 M phosphate buffer. The inability of hyperpolymers, in contrast to single duplexes, to be eluted from hydroxyapatite with 0.5 M phosphate buffer at 60°C without first melting at 95°C was used to assay hyperpolymer formation. Rate constants were determined from the reciprocal of the C_0t value at which 50% of the repeated sequence DNA was incorporated into a hyperpolymer. Average single stranded fragment sizes were determined by analytical centrifugation, as described by STUDIER (1965).

stranded tails after initial duplex formation with average starting fragment sizes below 600 nucleotides. This then is also consistent with a high proportion of the initial hybridisation events occurring between repeated sequences 500-600 nucleotides long.

The scheme illustrated in Figure 8 also accounts for the origin of S_1 sensitive sites spaced at 500-600 nucleotide intervals in much of the repeated sequence DNA since even when secondary hybridisation events occur, duplex formation is unlikely to occur without single stranded regions being left at each end of the initially formed duplex DNA. Hyperpolymer formation and S_1 sensitivity both support the model that related repeated sequences have different neighbouring repeated sequences at different places in the chromosome. The 8% of the wheat genome that forms long S_1 resistant duplexes upon renaturation (see Figure 7) must consist of repeated sequences much longer than 600 nucleotides or possibly short sequences arranged in tandem.

THE SIZE AND ORGANISATION OF THE HIGHLY DIVERGED REPEATED SEQUENCES

It was noted earlier that 73% of the duplex DNA between repeated sequences is resistant to S_1 nuclease. The degraded duplexes must be highly mismatched. These duplexes therefore involve families of repeated sequences within which there is considerable sequence divergence. To study the size and the organisation of these sequences a similar approach has been used to that already described in determining the size and organisation of non-repeated sequences (see Figure 3). Denatured, diverged repeated sequences can be prevented from forming stable duplexes upon incubation and thus made to behave as non-repeated DNA, by increasing the stringency of the incubation conditions (McCARTHY and FARQUHAR 1972). During incubation of denatured DNA in 62% formamide, 0.69 M NaCl at 42.5°C to C_0t 130 only well matched duplexes are stable. Consequently only 40 to 50% of the repeated sequence DNA forms stable duplexes with fragments below 500 nucleotides (BENDICH and McCARTHY 1970).

The proportion of wheat DNA retained by hydroxyapatite in the reannealed DNA fraction after denaturation and incubation under these conditions is shown in Figure 10 for average starting fragment sizes from 200 to over 10,000 nucleotides. Only approximately 40% of the genome is in the renatured fraction using very small fragment sizes but with starting fragments above 3,500 nucleotides long the proportion increases to 85% to include nearly all of the repeated sequences of the genome. The additional 53% of the genome retained in the renatured DNA fraction with 3,500 nucleotide fragment lengths is presumed to consist of single stranded tails attached to each duplex. This is supported by the fact that it is completely digested by S_1 nuclease, leaving only 32% of the genome in the form of stable, well base-paired duplexes.

This experiment provides several more important points about the organisation of the wheat genome. Approximately 28% (32% - 4% zero time binding DNA) consists of repeated sequences belonging to highly conserved families since it forms stable duplexes under the highly stringent incubation conditions. Separate measurements, as well as those in Figure 7, show that these sequences are short, with an average length of 500-600 nucleotides. They are often arranged contiguously in the chromosome but with adjacent sequences belonging to different families (see Figure 8).

Sixty three percent of the genome is retained in the duplex fraction with 1000 nucleotide long fragments. With longer fragments the rate of increase in renatured DNA with increasing fragment size decreases markedly (Figure 10). Therefore 31% (63 - 32%) must consist of single stranded tails of either the non-repeated sequence DNA of 1000 nucleotides average length or short diverged repeated sequences. If all the 16% non-repeated DNA of approximately 1000 nucleotides long is interspersed between conserved repeated sequences (see Figure 3) then the short, diverged repeated sequences interspersed between conserved repeated sequences would amount to 15% (31 - 16%) of the genome. Thus 43% (28 + 15%) of the wheat genome would consist of repeated sequences 500-600 nucleotides long. This is the same value as that found from using S_1 nuclease on DNA renatured at 60°C in 0.12 M phosphate buffer (see Figure 7).

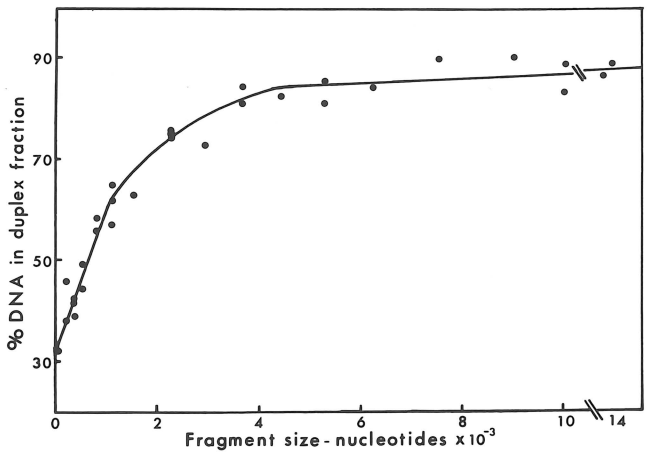

FIGURE 10. Interspersion of non-repeated and non-diverged and diverged repeated sequences. Wheat DNA was heat denatured and renatured in 62% formamide 0.69 M NaCl at 42.5°C to a C_0t of 100. The proportion in the renatured DNA fraction with different fragment sizes was determined by hydroxyapatite chromatography.

The remaining DNA (22%) included in the duplex fraction with DNA fragments between 1000 and 3500 nucleotides long (Figure 10) consists of repeated sequences 1000 to 3500 nucleotides long. These repeated sequences are heterogeneous and so form unstable duplexes under high stringency incubation conditions. They reside in the chromosome between repeated sequences, 500-600 nucleotides long, that are able to form stable duplexes under these incubation conditions. The long, mismatched repeated sequence duplexes are those repeated sequences (23% of the genome) previously found to be stable at 60°C in 0.12 M phosphate buffer but degraded by S_1 nuclease (see earlier).

Approximately 84% of the genome is included in the duplex fraction with fragment sizes of 4,000 nucleotides after incubation in 62% formamide, 0.69 M NaCl at 42.5°C (Figure 10) while approximately 95% is included in the duplex fraction with similar size fragments after incubation in 0.12 M phosphate buffer at 60°C (Figure 3). This difference is due to the absence of the 8% long repeated sequence duplexes (see Table 2) from the repeated sequence duplexes stable under the higher stringency incubation conditions.

ORGANISATION OF THE WHEAT GENOME - A SUMMARY

The proportion of the various kinds of nucleotide sequences we have uncovered in the wheat genome are summarised in Table 2, together with the corresponding results for *Xenopus* (DAVIDSON *et al.* 1973a, b). The wheat genome is also presented diagrammatically in Figure 11.

TABLE 2. Amounts of the different kinds of sequences in wheat and *Xenopus*.

Sequence	Wheat %	Wheat pg[a]	*Xenopus* %	*Xenopus* pg[a]
Non-repeated 800 to 1200 nucleotides	16	.96	38	1.18
Non-repeated > 4000 nucleotides	9	.54	35	1.10
Repeated, conserved 500-600 nucleotides	43	2.58	18[b]	0.56
Repeated diverged 2000-4000 nucleotides	20	1.20		
Repeated, long or tandemly repeated	8	.48	6	0.19
Palindrome (zero-time binding)	4	.24	2	0.06

[a]Calculated from haploid DNA contents. *Xenopus* 3.1 pg, wheat 6.0 pg.

[b]*Xenopus* short repeated sequences have a mean size of approximately 300 nucleotides (DAVIDSON *et al.* 1973a, b).

There are certain similarities in the kinds of sequences and their organisation in the two genomes. Both have a similar amount of non-repeated sequences about 800-1000 nucleotides long. Wheat has a considerably higher proportion and amount of repeated sequence DNA. While most of *Xenopus* DNA consists of alternating repeated and non-repeated sequences, most of wheat DNA consists of alternating kinds of repeated sequences (Table 2).

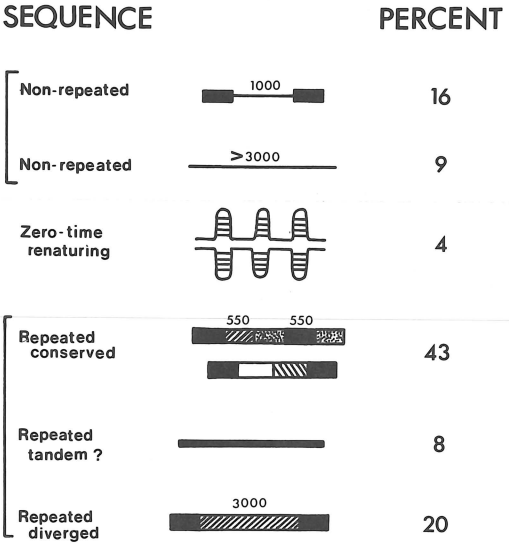

FIGURE 11. The sizes, organisation and proportions of the different kinds of nucleotide sequences in the wheat genome. The numbers indicate the approximate average lengths of the different sequences.

From the kinds of sequence organisation displayed in Figure 11, it is possible to make models of regions of wheat chromosomes. The model in Figure 12A shows a non-repeated sequence within an array of different 500 to 600 nucleotide repeated sequences. To fit our general findings, each repeated sequence would belong to a different family of conserved, nearly identical sequences. The number and order of repeated sequences adjacent to each coding sequence would be expected to be different for different coding sequences.

Other portions of the genome appear to consist of similar regions but with the non-repeated sequence replaced by longer 2000 to 4000 nucleotide repeated sequences (Figure 12B). There is considerable sequence heterogeneity between the longer repeated sequences.

These two units of organisation appear frequently and may account for up to 80% of the wheat genome if all the 500-600 nucleotide repeated sequences are associated with them.

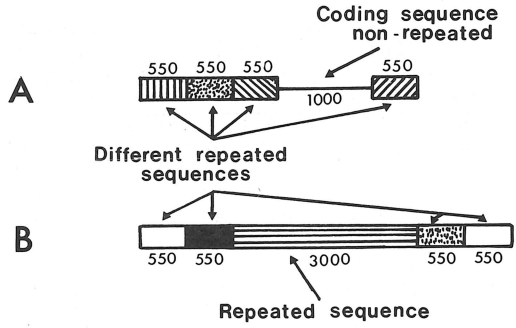

FIGURE 12. A schematic model for two common units of organisation within the wheat genome. A - a unit including a coding sequence. B - a unit including a repeated sequence approximately 3000 nucleotides long. The numbers indicate the approximate lengths in nucleotides of the various sequences.

FUNCTIONAL SIGNIFICANCE OF THE ORGANISATION OF REPEATED AND NON-REPEATED SEQUENCES.

Sixty four percent of the non-repeated DNA in wheat consists of sequences about the length of the average coding gene interspersed between short repeated sequences. Most of the *Xenopus*, sea urchin, rat and *Dictyostelium* genomes consist of single copy sequences alternating with short repeated sequences. A similar organisation also seems to be apparent in yeast mitochondrial DNA (PRUNNEL and BERNARDI 1974). This common arrangement is highly significant. At least some of the interspersed repeated sequences are transcribed. Furthermore, the nuclear very long primary transcription RNA products, containing messenger RNA (mRNA) sequences also contain repeated sequence transcripts on the same molecule (HOLMES and BONNER 1974; SMITH *et al.* 1974; JELINEK *et al.* 1974; VAN-DE-WALLE and DELTOUR 1974). The repeated sequences are degraded before the mRNA becomes bound to the polysome for translation, (KLEIN *et al.* 1974; GOLDBERG *et al.* 1973; BALSAMO, HIERRO and LARA 1973) but may preserve the mRNA from degradation (JELINEK *et al.* 1974). It is difficult to determine precisely the number of DNA sequences homologous to an mRNA species but it would appear, in the few cases investigated so far, that mRNA sequences are homologous to DNA sequences present in one copy per haploid genome (LODISH, FIRTEL and JACOBSON 1973, GREENBERG and PERRY 1971, BISHOP and ROSBASH 1973). The existence of repeated sequences as well as messenger on RNA nuclear transcripts strongly suppdrts the hypothesis that the single copy sequences interspersed

between repeated sequences in the genome are coding sequences. However, this in no way eliminates the possibility that some of the much longer non-repeated sequences (9% of the wheat genome) also code for proteins.

Attempts to explain from theoretical principles, the control of cell determination, pattern formation and differentiation in higher organisms have been based on complex regulatory mechanisms of a limited number of proteins (BRITTEN and DAVIDSON 1969, GEORGIEV 1973). The complex regulatory mechanisms suggested, include sites adjacent to the protein coding genes to which initiators, activators, inhibitors and tissue specific controlling elements can bind to influence transcription differentially in different developmental stages and tissues. Other sites may be essential for chromomere unfolding (PAUL 1972, CRICK 1971). The presence of multiple but different repeated sequences adjacent to possible coding sequences in the wheat genome (Figure 12A) is consistent with such models and invokes the speculation that the short 500-600 nucleotide repeated sequences are responsible for the control of gene expression.

There is little evidence available so far to enable us to speculate upon the function(s) of the other kinds of sequences in the wheat genome. However, repeating sequences or groups of sequences concerned with chromosome replication, chromosome folding and chromosome pairing might be expected to occur frequently.

Numerous studies in animal systems have shown constitutive heterochromatin to be associated with a high concentration of tandemly arranged highly repetitious sequences not regularly interspersed with other sequences (YUNIS and YASMINEH 1971). Wheat chromosomes contain only a small amount of constitutive heterochromatin as indicated by Giemsa staining (GILL and KIMBER 1974). This correlates with the wheat genome containing only a small amount of DNA (8%, Table 2) that could consist of similar highly repeated sequences tandemly arranged.

EVOLUTION OF SEQUENCE ORGANISATIONS

Families of repeated sequences could be formed from unequal crossing over or from aberrant DNA replication. Unequal crossing over is more plausible for relatively small changes in frequency of repeated sequences (SMITH 1973). To generate large numbers of nearly identical sequences however, DNA replication appears a more likely mechanism.

A likely feature of the families of repeated sequences, in the wheat genome, is that members of each family are distributed throughout the genome, on all chromosomes, with many different neighbouring sequences. There must clearly be mechanisms for transferring sequences to different places in the chromosomes (structural rearrangements within and be-

tween chromosomes) and selection for advantageous combinations of sequences. Unless the nucleotide sequences of the repeated sequences are maintained by natural selection there must also be mechanisms for keeping thousands of unlinked sequences essentially constant.

An alternative possibility for the evolution of the complex arrangements of the 500-600 nucleotide sequences (Figure 12) is that at some point in a primitive plant genome a cluster of short sequences was amplified and the newly formed clusters distributed throughout the genome. Subsequent modification and rearrangements of the sequences of individual clusters could have occurred to produce the variable patterns present today.

The evolution of palindrome sequences may have been achieved by a process of sequence duplication and inversion. However, since the presumed palindromes occur in clusters more complicated processes must also have been at work. A tremendous amount needs to be uncovered if we are to be able to understand the evolution of the complex higher plant genome.

Speciation is likely to have its genetic origin in control sequences rather than enzyme protein sequences. If repeated sequences are responsible for the control of gene expression and differentiation, then evolutionary changes in the repeated sequences of a genome may play an important part in species divergence.

AMPLIFICATION IN RYE EVOLUTION

We have recently compared the nucleotide sequence organisation in wheat with that in rye. Since diploid rye has a nuclear DNA content approximately 50% larger than the diploid progenitors of hexaploid wheat (BENNETT 1972) there is every reason to believe that with divergence of *Secale cereale* from the common wheat-rye ancester a DNA amplification event(s) occurred. Wheat and rye chromosomes are genetically very similar and most of the repeated nucleotide sequences in wheat are also present in similar proportions in rye (SMITH and FLAVELL 1974, RIMPAU and FLAVELL unpublished). Thus most of the additional DNA in rye consists of sequences also present in wheat. When very long rye DNA fragments are denatured, the repeated sequences reannealed and treated with S_1 nuclease, approximately 50% of the S_1 resistant DNA (32.5% of the genome) is in the form of very long duplexes. The remaining 50% consist of 500-600 nucleotide long duplexes similar to those found in wheat DNA (see Figure 7 for equivalent experiment with wheat DNA). Since only a small proportion of the wheat genome consists of long S_1 resistant duplexes after denaturation and renaturation, a major difference between these two closely related species is the amount of the repeated sequence DNA which when reannealed *in vitro* forms long S_1 nuclease resistant duplexes. Resistance to S_1 nuclease implies that the repeated sequences involved are at least of similar length to the fragments used in the experiments or that smaller identical repeating units

are tandemly arranged in relatively long stretches of DNA. An amplified complex chromosomal unit would be expected to be S_1 resistant after reannealing *in vitro*, if insufficient time has been available for extensive divergence of the identifical sequences or the repeats have been conserved for functional reasons. We therefore conclude that most of the S_1 resistant DNA in rye, after reannealing *in vitro*, is the DNA amplified during or after divergence of *Secale cereale* from the common ancestor of wheat and rye. This DNA is approximately 33% of the rye genome and therefore could numerically account for nearly all the additional DNA in diploid rye not in the diploid wheats.

Secale cereale, variety Petkus, with which we have worked, has large areas of constitutive heterochromatin at the telomeres of most, if not all chromosomes in contrast to wheat (BENNETT 1974 and personal communication). It is particularly interesting to investigate whether the additional 33% of the rye genome not in wheat is localised in the telomeric constitutive heterochromatin.

Constitutive heterochromatin is very common in higher plants (STEBBINS 1971). The underlying DNA sequence organisation of heterochromatin in plants is known. Simple sequence satellites, such as form the basis of heterochromatic regions of animal chromosomes, may not be common in plants (INGLE, PEARSON and SINCLAIR 1973). The recent DNA amplification in rye is a useful system for investigating some of the consequences of amplification of multiple complex sequences and their possible role in heterochromatic chromatin configurations.

Although we have stressed the role of amplification in the organisation of plant genomes, it must not be forgotten that deletion, duplication, inversion, translocation and polyploidy have all occurred frequently in plant evolution to create quantitative chromosomal variation.

CONCLUDING REMARKS

The kinds of sequences and their organisation described in this paper have been confined to species in the Gramineae since virtually no other species have been investigated in such detail. However, we believe that the results will be applicable to higher plant species in general.

Cytological observations of chromosomes provide descriptions of chromosome organisation on a gross scale, recognising centromeres, nucleolar organiser constrictions, inversions and translocations. More recent applications of fluorescence microscopy and specific stains, e.g. Giemsa, have enabled individual chromosomes to be characterised and differentiated into numerous discrete regions. However, while of use in chromosome recognition, such staining procedures allow little to be concluded about the underlying DNA sequences. The very specific banding patterns of the polytene chromosome in Diptera provide a much

more detailed picture of chromosome organisation but their counterparts in plants cannot be interpreted with such precision. Genetic studies in higher plants can do little more for chromosome organisation studies than approximately locate loci on chromosomes, fine structure analysis being extremely time and energy consuming, even where sufficient allelic variation is present in the species. Without doubt, for a detailed understanding of the molecular organisation and function of the DNA in the chromosomes of higher plants, we must also use biochemical methods. The work described in this paper constitutes a beginning for the application of such methods to plant genomes.

LITERATURE CITED

ANDO, T. 1966 A nuclease specific for heat-denatured DNA isolated from a product of *Aspergillus oryzae*. Biochim. Biophys. Acta 114:158-68.

BALSAMO, J., J. M. HIERRO and F. J. S. LARA 1973 Transcription of repetitive DNA sequences in *Rhynchosciara* salivary glands. Cell Differentiation 2:119-30.

BENDICH, A. J. and B. J. McCARTHY 1970 DNA comparisons among barley, oats, rye and wheat. Genetics 65:545-66.

BENNETT, M. D. 1972 Nuclear DNA content and minimum generation time in herbaceous plants. Proc. Roy. Soc. Lond. B. 181:109-35.

BENNETT, M. D. 1974 Meiotic, gametophytic and early endosperm development in Triticale. In 'Triticale', Ed. R. MacIntyre and M. Campbell Pp. 137-148. International Development Research Centre, Ottawa, Canada.

BERNARDI, G. 1969 Chromatography of nucleic acids on hydroxyapatite 1. Chromatography of Native DNA. Biochim. Biophys. Acta 174:423-34.

BIRNSTIEL, M. L., M. CHIPCHASE and J. SPIERS 1971 The ribosomal RNA cistrons. Prog. Nucleic Acid Res. Mol. Biol. 11:351-89.

BISHOP, J. O. and M. ROSBASH 1973 Reiteration frequency of duck haemoglobin genes. Nature New Biol. 241:204-7.

BONNER, J., W. T. GARRARD, J. GOTTESFELD, D. S. HOLMES, J. S. SEVALL and M. WILKES 1973 Functional organisation of the mammalian DNA. Carnegie Inst. Year Book 73:1079-88.

BRITTEN, R. J. and E. H. DAVIDSON 1969 Gene regulation for higher cells: a theory. Science 165:349.

BRITTEN, R. J. and D. E. KOHNE 1965 Nucleotide sequence repetition in DNA. Carnegie Inst. Year Book 65:78-106.

BROWN, D. D. and K. SUGIMOTO 1973 The structure and evolution of ribosomal and 5s DNAs in *Xenopus laevis* and *Xenopus mulleri*. Cold Spring Harbor Symp. Quant. Biol. 38:501-5.

CLARKSON, S. G. and M. L. BIRNSTIEL 1973 Clustered arrangement of tRNA genes of *Xenopus laevis*. Cold Spring Harbor Symp. Quant. Biol. 38:451-60.

CRICK, F. H. C. 1971 General model for the chromosome of higher organisms. Nature 234:25-7.

DAVIDSON, E. H., B. R. HOUGH, C. S. AMENSON and R. J. BRITTEN 1973a General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. J. Mol. Biol. 77:1-23.

DAVIDSON, E. H., D. E. GRAHAM, B. R. NEUFELD, M. E. CHAMBERLIN, C. S. AMENSON, B. R. HOUGH and R. J. BRITTEN 1973b Arrangement and characterisation of repetitive sequence elements in animal DNAs. Cold Spring Harbor Symp. Quant. Biol. 38:295-302.

FLAVELL, R. B., M. D. BENNETT, J. B. SMITH and D. B. SMITH, 1974 Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem. Genet. 12:257-69.

FLAVELL, R. B. and D. B. SMITH 1975 In preparation.

GEORGIEV, G. P., A. J. VARSHAVSKY, A. P. RYSKOV and R. B. CHURCH 1973 On the structural organisation of the transcriptional unit in animal chromosomes. Cold Spring Harbor Symp. Quant. Biol. 38:869-84.

GILL, B. S. and G. KIMBER 1974 Giemsa C-banding and the evolution of wheat. Proc. Nat. Acad. Sci. US 71:4086-90.

GOLDBERG, R. B., G. A. GALAU, R. J. BRITTEN and E. H. DAVIDSON 1973 Non-repetitive DNA sequence representation in sea urchin embryo messenger RNA. Proc. Nat. Acad. Sci. US 70:3516-20.

GRAHAM, D. E. and R. J. BRITTEN 1973 Hyperpolymers and another approach to repetitive sequence organisation. Carnegie Inst. Year Book 72:223.

GRAHAM, D. E., B. R. NEUFELD, E. H. DAVIDSON and R. J. BRITTEN 1974 Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell 1:127-37.

GREENBERG, J. R. and R. P. PERRY 1971 Hybridisation properties of DNA sequences directing the synthesis of messenger RNA and heterogeneous nuclear RNA. J. Cell Biol. 50:774-86.

HAMER, D. H. and C. A. THOMAS Jr. 1974 Palindrome theory. J. Mol. Biol. 84:139-44.

HOLMES, D. S. and J. BONNER 1974 Interspersion of repetitive and single copy sequences in nuclear RNA of high molecular weight. Proc. Nat. Acad. Sci. US 71:1108-12.

INGLE, J., G. C. PEARSON and J. SINCLAIR 1973 Species distribution and properties of nuclear satellite DNA in higher plants. Nature New Biol. 242:193-97.

JELINEK, W., G. MOLLOY, R. FERNANDEZ-MUNOZ, M. SALDITT and J. E. DARNELL 1974 Secondary structure in heterogeneous nuclear RNA. Involvement of regions from repeated DNA sites. J. Mol. Biol. 82:361-70.

KEDES, L. and M. L. BIRNSTIEL 1971 Reiteration and clustering of DNA sequences complementary to histone messenger RNA. Nature New Biol. 230:165-69.

KLEIN, W. H., W. MURPHY, G. ATTARDI, R. J. BRITTEN and E. H. DAVIDSON 1974 Distribution of repetitive and non-repetitive sequence transcripts in Hela mRNA. Proc. Nat. Acad. Sci. US 71:1785-89.

LAIRD, C. D., W. Y. CHOOI, E. H. COHEN, E. DICKSON, N. HUTCHINSON and S. H. TURNER 1973 Organisation and transcription of DNA in chromosomes and mitochondria of *Drosophila*. Cold Spring Harbor Symp. Quant. Biol. 38:311-27.

LODISH, H. F., R. A. FIRTEL and A. JACOBSON 1973 Transcription and structure of the genome of the cellular slime mold *Dictyostelium discoideum*. Cold Spring Harbor Symp. Quant. Biol. 38:899-14.

McCARTHY, B. J. and M. N. FARQUHAR 1972 The rate of change of DNA in evolution. Brookhaven Symp. Biol. 23:1-43.
PAUL, J. 1972 General theory of chromosome structure and gene activation in eucaryotes. Nature 238:444-46.
PRUNELL, A. and G. BERNARDI 1974 The mitochondrial genome of wild type yeast cells. J. Mol. Biol. 86:825-41.
SMITH, D. B. and R. B. FLAVELL 1974 The relatedness and evolution of repeated nucleotide sequences in the genomes of some Gramineae species. Biochem. Genet. 12:243-56.
SMITH, D. B. and R. B. FLAVELL 1975 Characterisation of the wheat genome by renaturation kinetics. Chromosoma (Berl.) 50:223-242.
SMITH, G. P. 1973 Unequal crossover and the evolution of multigene families- Cold Spring Harbor Symp. Quant. Biol. 38:507-14.
SMITH, M. J., B. R. HOUGH, M. E. CHAMBERLIN and E. H. DAVIDSON 1974 Repetitive and non-repetitive sequences in sea urchin heterogeneous nuclear RNA. J. Mol. Biol. 85:103-26.
STEBBINS, G. L. 1971 Chromosomal evolution in higher plants. Arnold, London.
STUDIER, F. W. 1965 Sedimentation studies of the size and shape of DNA. J. Mol. Biol. 11:373-90.
SUTTON, W. D. 1971 A crude nuclease preparation suitable for use in DNA reassociation experiments. Biochim. Biophys. Acta 240:522-31.
VAN-DE WALLE, C. and R. DELTOUR 1974 Presence of heterodisperse nuclear RNA in a plant : *Zea mays*. FEBS Letters 49:87-91.
WILSON, D. A. and C. A. THOMAS Jr. 1974 Palindromes in chromosomes. J. Mol. Biol. 84:115-44.
YUNIS, J. J. and W. G. YASMINEH 1971 Heterochromatin, satellite DNA and cell function. Science 174:1200-09.

Richard B. Flavell

The lecture about to begin

Informal discussions

Gary Kikudome

The Eisenstarks