Genet. Res., Camb. (1982), **39**, *pp.* 1–30
With 4 text-figures
Printed in Great Britain

Nucleotide sequences of highly repeated DNAs; compilation and comments

BY GEORGE L. GABOR MIKLOS AND AMANDA CLARE GILL

Department of Population Biology, Research School of Biological Sciences, The Australian National University, Canberra, A.C.T., Australia

(*Received* 4 *February* 1981 *and in revised form* 23 *June* 1981)

SUMMARY

The nucleotide sequence data from highly repeated DNAs of invertebrates and mammals are summarized and briefly discussed. Very similar conclusions can be drawn from the two data bases. Sequence complexities can vary from 2 bp to at least 359 bp in invertebrates and from 3 bp to at least 2350 bp in mammals. The larger sequences may or may not exhibit a substructure. Significant sequence variation occurs for any given repeated array within a species, but the sources of this heterogeneity have not been systematically partitioned. The types of alterations in a basic repeating unit can involve base changes as well as deletions or additions which can vary from 1 bp to at least 98 bp in length. These changes indicate that sequence *per se* is unlikely to be under significant biological constraints and may sensibly be examined by analogy to Kimura's neutral theory for allelic variation. It is not possible with the present evidence to discriminate between the roles of *neutral* and *selective* mechanisms in the evolution of highly repeated DNA.

Tandemly repeated arrays are constantly subjected to cycles of amplification and deletion by mechanisms for which the available data stem largely from ribosomal genes. It is a matter of conjecture whether the solutions to the mechanistic puzzles involved in amplification or rapid redeployment of satellite sequences throughout a genome will necessarily give any insight into biological functions.

The lack of significant somatic effects when the satellite DNA content of a genome is significantly perturbed indicates that the hunt for specific functions at the *cellular* level is unlikely to prove profitable.

The presence or in some cases the *amount* of satellite DNA on a chromosome, however, can have significant effects in the germ line. There the data show that localized condensed chromatin, rich in satellite DNA, can have the effect of rendering adjacent euchromatic regions *rec*$^-$, or of altering levels of recombination on different chromosomes. No data stemming from natural populations however are yet available to tell us if these effects are of adaptive or evolutionary significance.

INTRODUCTION

Many eukaryote genomes contain a large proportion of highly repeated DNA sequences. In this summary we discuss only those which exist in long tandem arrays. In *Drosophila nasutoides*, 60% of the nuclear DNA consists of four such sequences, all of which are located on one pair of giant chromosomes (Cordeiro-Stone & Lee, 1976; Wheeler *et al.* 1978). In the desert rodent, *Dipodomys ordii*, more than half the genome is made up of variants of AAG, GGGTTA and ACACAGCGGG (Salser *et al.* 1976). In other eukaryotes such as the fungus, *Aspergillus nidulans*, there are virtually no highly repeated sequences (Timberlake, 1978), nor is there much more than 1% in some species of wild rats (Miklos, Willcocks & Baverstock 1980).

Moreover, whether a eukaroyte is unicellular or multicellular, invertebrate or vertebrate, the amount of highly repeated DNA in its genome seems independent of its phylogenetic 'position'. What then are the functions (if any) of this class of DNA? What kinds of structural change contribute to variation at the DNA sequence level? What types of sequence variability occur during evolution and what implications might they have for possible functions?

PART 1. SEQUENCE DATA ON HIGHLY REPEATED DNAs

Invertebrates

(i) *Crustacea*

A very simple repeated DNA is that of the crab, *Cancer borealis*, which is greater than 90% AT (Table 1) (reviewed by Skinner, 1977). The two highly repeated DNAs of the hermit crab *Pagurus pollicaris* are more complex, being respectively 4 bp and a minimum of 16 bp in length. Variation in the basic repeating unit of the latter satellite is due to the variable number of CAG triplets so that the most frequent fragment is a 31 bp unit of the form $(CAG)_8$ CTGCACT (Chambers, Schell & Skinner, 1978). The red crab *Geryon quinquedens* on the other hand has approximately 2% of its genome as a repeating unit based on a series of 81 bp (Christie & Skinner, 1980*a*, *b*). Thus between these three species of *Crustacea*, the minimum complexities vary over at least a 40-fold range.

(ii) *Drosophila*

In this genus the various satellites can be closely related to each other at the sequence level. *Drosophila virilis* has four satellites (termed I, II, III and Ic) which constitute 23, 9, 8 and 0·1% of the diploid genome and each is a 7-base pair repeat (Table 2). The first three satellites differ from each other by a single base change (Gall and Atherton, 1974) whereas satellite I(c) has a sequence which shows characteristics of both *D. virilis* and *D. melanogaster* sequences (Mullins & Blumenfeld, 1979). Some members of the *D. virilis* group however, such as *D.*

ezoana and *D. littoralis*, are almost totally devoid of some satellite DNAs (Cohen & Bowman, 1979).

In *D. melanogaster*, each of the five major satellites comprises about 4 % of the genome, so that approximately 20 % of the nuclear DNA consists of highly repeated sequences (reviewed by Brutlag, 1980). Besides the 5 bp and 7 bp variants of the 1·672 and 1·705 g/cm³ satellites there are 10, 254 and 359 bp variants (Table 2).

Table 1. *Major average nucleotide sequences of Crustacean highly repeated DNAs. All sequences in this and accompanying tables and figures run 5′–3′ and are shown in stretches of 10 nucleotides*

Cancer borealis		A T			
Pagurus policaris		T A G G $(C A G)_n$ C T G C A C T			
Geryon quinquedens	1	A G C T T A T C A C	C A C C T G T A A C	A A C T T T T T T T	30
	31	G T A T A A G T C C	C C A A G C G C T T	C A C C G T G C C A	60
	61	C A G C C C T G C T	T T T G G C C G T C	A	81

Analysis of the 359 bp sequence reveals that there is no obvious relationship between it and the simpler sequences of *D. melanogaster* (Hsieh & Brutlag, 1979*a*). The 359 and 254 bp variants are however closely related to each other. They differ by one large 98 bp deletion (or addition), six smaller deletions, numerous base changes, and share only about 60 % homology (Carlson & Brutlag, 1979).

In three species of Hawaiian Drosophila (Table 3) a major satellite DNA at 1·690 g/cm³ comprises 40 % of the diploid genome and a large part of this DNA consists of sequence variants which are derivatives of a 189 bp sequence (Miklos & Gill, 1981). *Drosophila grimshawi* has a major repeat of 189 bp whereas *D. gymnobasis* and *D. silvarentis* have repeat lengths of 177, 179 and 185 base pairs. All the variants are derivable from the 189 base pair sequence by 'deletions' and single base changes. The consensus sequences have hints of a substructure. The tetranucleotide 5′ TTAA 3′ can be spaced at distances of 35, 37, 35 and 71 base pairs. The Hawaiian *Drosophila* data reinforce the points that the repeating unit can be large and that very closely related species can share similar sequences.

The results from the *Drosophila virilis* complex, *D. melanogaster*, Hawaiian *Drosophila*, as well as *D. nasutoides* reveal the following:

(1) Sequence complexities can vary from 5 to 359 bp.

(2) The proportion of highly repeated DNA in different *Drosophila* genomes can range from almost zero to in excess of 60 %.

(3) Some satellite sequences are similar to other satellites, whereas others show no obvious homologies.

(4) Complex sequences sometimes reveal weak indications of a substructure.

(5) Some sequences can be confined to a single chromosome but more usually they occupy heterochromatic regions on most chromosomes of the set.

Mammals

(i) *The kangaroo rats*

One species of kangaroo rat (*Dipodomys ordii*) has three satellite sequences with basic repeating units of AAG, GGGTTA and ACACAGCGGG respectively (Table 4). These sequences have been determined from uncloned material by ribosubstitution

Table 2. *Major average nucleotide sequences of various* Drosophila *highly repeated DNAs. In the 359 bp and 254 bp variants of the 1·688 g/cm³ satellite, the deletions are denoted —, and undetermined bases as X.*

D. virilis	Satellite I	A C A A A C T			
	Satellite II	A T A A A C T			
	Satellite III	A C A A A T T			
	Satellite Ic	A A T A T A G			
D. melanogaster:	Satellite 1·672 g/cm³	A A T A T			
		A A T A T A T			
	Satellite 1·705 g/cm³	A A G A G			
		A A G A G A G			
	Satellite 1·686 g/cm³	A A T A A C A T A G			
359 bp		C C A C A T T T T G	C A A A T T T T G A	T G A C C C C C C T	30
254 bp		– C A X A T T T – G	C A A A T T T – A A	T G A A C C C C – –	
359 bp		C C T T A C A A A A	A A T G C G A A A A	T T G A T C C A A A	60
254 bp		C C T T – C A A A A	A A T G C G A A A A	T T A A C G C A A A	
359 bp		A A T T A A T T T C	C C T A A A T C C T	T C A A A A A G T A	90
254 bp		A A T T G A T T T C	C C T A A A T C C T	T C A A A A A G T A	
359 bp		A T A G G G A T C G	T T A G C A C T G G	T A A T T A G C T G	120
254 bp		A – – – – – – – – –	– – – – – – – – – –	– – – – – – – – – –	
359 bp		C T C A A A A C A G	A T A T T C G T A C	A T C T A T G T G A	150
254 bp		– – – – – – – – – –	– – – – – – – – – –	– – – – – – – – – –	
359 bp		C C A T T T T T A G	C C A A G T T A T A	A C G A A A A T T T	180
254 bp		– – – – – – – – – –	– – – – – – – – – –	– – – – – – – – – –	
359 bp		C G T T T G T A A A	T A T C C A C T T T	T T T G C A G A G T	210
254 bp		– – – – – – – – – A	T A A C A A C T T T	T T G G C A A A A T	
359 bp		C T G T T T T T C C	A A A T T T C G G T	C A T C A A A T A A	240
254 bp		C T G A T T – C C C	T A A T T T C G G T	C A T T A A A T A A	
359 bp		T C A T T T A T T T	T G C C A C A A C A	T A A A A A A T A A	270
254 bp		T C A G T T T T T T	T G C C A C A A C T	T T A A A A A T A A	
359 bp		T T G T C T G A A T	A T G G A A T G T C	A T A T C T C A C T	300
254 bp		T T G T C T G A A T	A T G G A A T G T C	A T A C C T C G C X	
359 bp		G A G C T C G T A A	T A A A A T T T C C	A A T C A A A C T G	330
254 bp		X A G C T X G T A A	T T A A A T T T C C	A A T G A A A C T G	
359 bp		T G T T C A A A A A	T G G A A A T T A A	A T T T T T T G G	359
254 bp		T G T T C A A C A A	T G A A A A T T A C	A T T T T T C G G	

sequencing techniques (Salser *et al* 1976; Fry & Salser, 1977). The MS satellite which constitutes 22% of this genome is better expressed as $\begin{matrix} A & & A\ G \\ /\backslash & & | \\ C\ G & & A \end{matrix}$. The HS-$\alpha$ satellite which makes up to 19% of the genome ought to be written as

https://doi.org/10.1017/S0016672300020711 Published online by Cambridge University Press

Table 3. *Major average nucleotide sequence of the 1·690 g/cm³ satellite from three species of Hawaiian* Drosophila, D. gymnobasis, D. silvarentis *and* D. grimshawi.

D. gymnobasis	G A A T T C G C A (T/C) G A A T G T T G A (A/C) A C A A A A C T G C	30
D. silvarentis	G A A T T C G C A C G (A A C / G C T) G T T G A (A/C) A (C/T) A A A A C T G C	
D. grimshawi	G A A T T C A C T C G A T C T T T G G A A C A A A A A C T G C	
D. gymnobasis	A T T T A A A A A A A T (T/G) G A C T T (T/C) A G A C T C A T C T A	60
D. silvarentis	A T (T T A / C A C) A A A A A A T G G A C T T T A G A C T C A T C T A	
D. grimshawi	A T T A A A A T A A T T T G A C T T T C G A C T C A T C T A	
D. gymnobasis	A G A A A G A T A T T A A (C/G) G G A T A T C C (G/–) T C A A A A A	90
D. silvarentis	A G A A A G A T A T T A A (C/G) G G A T A T C C (G/–) T C A A A (A/T) A	
D. grimshawi	A G A A A G A T A T T A A G G G A T A T T C T T C G A T A A	
D. gymnobasis	C T C G A (– – – T A / A A A – –) T A A G T T (G̅/–) A (C/–) A T G A C (A T A G G / – – – – –)	120
D. silvarentis	(C/A) T C G A – – – T A T (A A G T T A A – C / – – – – – – – –) A T G C C (A T (A/T) G (G/C) / – – – – –)	
D. grimshawi	A T A G A A A A T (A/T) T A (T/A) G G A A A (G C / A T) A T T A A A T A G G	
D. gymnobasis	(C/A) T C A T C C T (A/T) T A A A C A A A T G G C C A T A T C T T C	150
D. silvarentis	(C/A) T C A T C C T A T A A A (G/C) A A A T (G/A) G C C A T A T C T T C	
D. grimshawi	G T C A A C C T T G A A A T A A A T G G C C A T A T C T T C	
D. gymnobasis	G T C A A A A A T T G T C C A A T T T C (A/C) A A T A T T A A G	180
D. silvarentis	G T C A A A A A T T G T C C A A (T/A) T T (C/T) A A A T A T T A A G	
D. grimshawi	G T C A A A A A T T G C C C C A T T T C A A T T A T A A A G	
D. gymnobasis	G T (A G / G T) T T T T (G/A)	189
D. silvarentis	G X G T T T T T A	
D. grimshawi	G T G T T T T T A	

https://doi.org/10.1017/S0016672300020711 Published online by Cambridge University Press

G G G T T A
  AATG
A C A C A G C G_{3-5}
  A G

. The HS-β satellite (11% of the genome) is

Since the DNA used in some sequencing studies has been obtained from different individuals, the observed heterogeneity can have a number of origins. There may be variation between (*a*) individual repeating units in a tandem array on a chromosome, (*b*) chromosomes in a genome, (*c*) individuals in a population and (*d*) tissues in the same individual. The heterogeneity observed within the kangaroo

Table 4. *Major average nucleotide sequences of the three kangaroo rat and guinea-pig satellites*

MS satellite	A A G
HS-α satellite	G G G T T A
HS-β satellite	A C A C A G C G G G
Guinea pig α satellite	T T A G G G

rat data may stem from any or all of these four sources. It should also be stressed that the available data refer solely to *D. ordii*. In fact the amounts of the satellite DNAs (and presumably similar sequences) vary dramatically amongst closely related species of kangaroo rats with some species having hardly any.

(*ii*) *The guinea pig*

The major sequence of the guinea pig α-satellite (TTAGGG) was deduced from fingerprints of pyrimidine tracts (Southern, 1970). The studies of Salser *et al.* (1976) demonstrated that the major repeating unit of the HS-α DNA in the kangaroo rat was identical to the major repeat in the guinea pig α-satellite. Futhermore, many of the sequence variants of the kangaroo rat HS-α were identical to the sequence variants of the guinea pig α-sequence. Since guinea pigs and kangaroo rats are thought to have diverged approximately 50 million years ago, Fry & Salser (1977) argued that satellite DNAs could be 'conserved' over long evolutionary times and that the reasons for this 'conservation' could be explained by assumed protein/DNA binding capabilities and a postulated role in speciation through effects on chromosome pairing.

(iii) *The mouse* (Mus musculus)

The sequence of uncloned mouse satellite DNA is a 234 bp unit with extensive internal substructure (Table 5) (Horz & Altenburger, 1981). The major internal repetition is one of essentially 28 bp units alternating with 30 bp units. As originally postulated by Southern (1975), the sequence could have come about in at least four multiplication steps with sequence divergence between steps.

(iv) *The 'laboratory rat'*

Rattus norvegicus has only a small proportion (< 3 % of the genome) of very highly repeated DNA with a copy number in excess of 100000. The sequence shown

Table 5. *Major average nucleotide sequence of mouse satellite DNA*

1	G G A C C T G G A A	T A T G G C G A G A	A A A C T G A A A A	30
31	T C A C G G A A A A	T G A G A A A T A C	A C A C T T T A G G	60
61	A C G T G A A A T A	T G G C G A G G A A	A A C T G A A A A A	90
91	G G T G G A A A A T	T T A G A A A T G T	C C A C T G T A G G	120
121	A C G T G G A A T A	T G G C A A G A A A	A C T G A A A A T C	150
151	A T G G A A A A T G	A G A A A C A T C C	A C T T G A C G A C	180
181	T T G A A A A A T G	A C G A A A T C A C	T A A A A A A C G T	210
211	G A A A A A T G A G	A A A T G C A C A C	T G A A	234

Table 6. *Major average nucleotide sequence of rat highly repeated DNA*

1	G A A T T C A C A G	A G A A A C A G T G	T T T C A G T T C G	30
31	T T A A A A C G T T	G C T C T A T C T T	G A A T A A C A A G	60
61	C T T A T T A C A T	G C G A A T C C T A	T T G G G A A C C T	90
91	A C T G A A T T C A	C C A T G A T A C T	T A G A T T C C G T	120
121	T C C T C A A A A T	G T T G C T C C A T	A T T G A A A A A G C	150
151	A A A C T C A T - A	C A A G C A T G T C	C C A T T G G A A A	180
181	C T C A C T G A A T	T C G C C T A G A A	A T T T T G A T T C	210
211	C A T T C G T G A A	A A T T T T T C T A	T A T C C C G A A C	240
241	A G T C C A C T T A	T T A C T A C T G C	G G C C C A C T G G	270
271	G A A C T A A C C G	A A T T C A C C A T	G T T A C T C A G A	300
301	T T C G G C T C A C	C - A A A T T T T G	A T A A A T C T T T	330
331	A A A A G T A C A C	A T A T T A C A A G	A G C A G G C T A C	360
361	T G G G A A C T A A	C T		372

in Table 6 represents an average of the most frequent nucleotides at each single position in the 370 bp monomer (Pech, Igo-Kemenes & Zachau, 1979*a*). This sequence has a clear substructure based on alternating 92 and 93 bp repeats. The 92 and 93 bp fragments are also found in two wild species of rats (*R. sordidus* and *R. villosissimus*) but in only one-thousandth of the amount found in *R. norvegicus* (Miklos *et al.* 1980). The rat data parallel the situation seen in the *Drosophila virilis* group where some species have almost zero quantities of sequences whereas in others satellites constitute nearly a quarter of the genome.

The rodent data can be summarized in an almost identical manner to those for *Drosophila:*

(1) Complexities vary from 3 bp to at least 370 bp.

(2) The proportion of highly repeated sequences can vary from a fraction of 1 % to more than half the genome.

(3) Some sequence variants can be identical or very close even between species, whereas others show no such relationships.

(4) Some sequences yield clear evidence of substructure.

(5) The sequences are heterogeneous within a species but the exact source of the heterogeneity has not been determined.

(v) *The calf* (Bos taurus)

Two DNAs of the calf, the 1·706 and the 1·720 g/cm³ satellites, have been analysed by cloning and sequencing (Pech, Streeck & Zachau, 1979*b*; Poschl & Streeck, 1980). The minimum repeating unit is 23 bp (Table 7). There is a longer periodicity however of around 2350 bp within which four segments (termed A, B, C and D) can be recognized. The A and C segments, which contain many sites for

Table 7. *Major average nucleotide sequences from four regions of the calf 23 bp 1·706 g/cm³ satellite, together with the 46 bp sequence of the calf 1·720 g/cm³ satellite*

Cloned *Sau* segments		G A T C A C G T G A	C T G A T C A T G C	A C T	23
Uncloned *Sau* segment C		G A T C A C G T G G	C T G A T C A T G C	A C T	
PvuD segment		A A T C A A G C A G	C T C A G C A G G C	A A T	
PvuB segment		A A T C A T G C A G	C T C A G C A G G C	A A/G T	
1·720 g/cm³ satellite	1	T A T C A G G C A G	A T G A G C G G G C	A G G	23
		T G T C G C G C G G	C T C A G C T G G C	G A G	46

the enzyme *Sau 3A1*, are designated *Sau* segments, and vary in length by multiples of 22–23 bp. The B and D segments of 247 and 251 bp also consist of an underlying 23 bp repeating unit (Tables 8 and 9) in which there is extensive variation. This variation is clearly seen in the data of Tables 8 and 9, but is 'lost' when only the core sequences are presented (as in Table 7). The 1·720 g/cm³ satellite has a 46 bp repeating unit (Table 7) which is in effect a 23 bp dimer. Thus it exhibits a high degree of homology to the 23 bp prototype sequence (Poschl & Streeck, 1980).

Primates

Satellite DNAs

About 4% of the human genome is readily recoverable as satellite DNA and its distribution within the human genome, as well as its homologies to related primates, have been described (Gosden *et al.* 1975, 1977; Gosden, Lawrie & Cooke, 1981; Miklos & John, 1979). Within each of the satellites there is extensive heterogeneity as revealed by restriction endonuclease analyses (Mitchell, Beauchamp & Bostock, 1979) but again the precise source of this heterogeneity is open to question. In the case of satellite III, different subpopulations of the satellite have been located on different chromosomes (Bostock, Gosden & Mitchell, 1978; Beauchamp *et al.* 1979).

The heterogeneity revealed by these techniques is also obvious at the sequence level. Table 10 illustrates the three incomplete sequences from a cloned fragment

Table 8. *Major average nucleotide sequence of the* PvuB *segment of calf 1·706 g/cm³ satellite*

1	G A T C A T G T G C	C T C T G G A G G C	A A T	23
24	G A A C A T A G A G	C T G A G T A G G C	A G G	46
47	A A T C A G – – – –	– T C A G C A G G C	T A T	69
70	A A T C A T G G A A	C T C A C C T C T G	A G A	92
93	A A T C A T G C T G	C T C A G C T G G C	A A T	115
116	A A T C A A G T A G	A T G A G C A G G C	A G G	138
139	A A T T A C G C A G	C A C A G G T G G C	A A T	161
162	T G T C A A G G A G	A T G A – C A G G C	A G G	184
185	A A T C G T G C A G	C T C A G C T G G A	A A T	207
208	T G T C A A G C A G	A T G A G C A G A C	A G T	230
231	A A T C A C G C A G	C T C A G C A G G C	C C T	253
	A A T C A T G C A G	C T C A G C A G G C	A A/G T	

Table 9. *Major average nucleotide sequence of an uncloned as well as two cloned fragments from the* PvuD *segment of calf 1·706 g/cm³ satellite*

1	G A T C A T G T G C	C T C T G G A G G C	A A T	23
24	A A A C A A A G A G	C T G A G T A G G C	A C G	46
47	A A T C A A A C A G	– T C A G C A G G C	A A T	69
70	A A T C A T G G A A	C T C A G C T G T G	A G G	92
93	A A T C A T G C T G	C T C A G C T G G C	A A T	115
116	A A T C A A G C A G	C T G A G C A G G C	A G G	138
139	A A T T A C A C A G	C A C A G C T G G C	A A T	161
162	T G T C A A G C A G	A T G A – C A G G C	A G G	184
185	A A T C G T G C A G	C T C A G C T G G C	A A T	207
208	T G T C A A G C A G	A T G A G C A G A C	A C T	230
231	A A T C A C G C A G	C T C A G C A G G C	C C G	253
	A A T C A A G C A G	C T C A G C A G G C	A A T	

of human satellite III. They are neither simple nor complex, with a common repeat being ATTC. This satellite thus has a substructure and may have arisen from a simple repeat (Cooke & Hindley, 1979).

Sequence analyses of restriction endonuclease fragments

The bulk of the primate sequence data are from uncloned restriction fragments and are summarized in Table II. In human beings, approximately 1% of the genome can be isolated as a 340 bp restriction fragment which consists of two unequal halves of 171 and 169 bp. These share 72% homology with each other and

https://doi.org/10.1017/S0016672300020711 Published online by Cambridge University Press

are denoted human-1 and human-2 (Manuelidis & Wu, 1978; Wu & Manuelidis, 1980). The west African baboon (*Papio papio*) has a similar sized fragment of 343 bp which consists of 172 and 171 bp halves (denoted baboon-1 and baboon-2 in Table 11). These exhibit 70 % homology (Donehower *et al.* 1980). The closely related bonnet monkey (*Macaca radiata*) has about 6 % of its genome as a 343 bp restriction fragment again with two unequal halves of 172 and 171 bp. These halves (bonnet-1 and bonnet-2) share 67 % homology (Rubin *et al.* 1980). The African

Table. 10. *Nucleotide sequences from three regions of human satellite III*

1	A A T T C A T T T G	A A G A C A A T T C	C A T T C A A T A C	30
31	C A A T T G A T G A	T G G T T A T T T T	T G A T T C C A T T	60
61	T G A T G A T G A T	T A C A T T C C A T	T T C A T C A T A A	90
91	T T C C A T T C G A	T T C C A C T C G A	G A T T C C A T T C	120
121	G A T T C C A T T C	A A		132
1	C G A A T G A A T G	A G T C C A T C C A	T T T C A A T T T C	30
31	A T G A T A A T T C	C A T T C G T T T C	A A T T C G A T G G	60
61	T G T T T C C A T T	C G A T T		75
1	T T C A T T C G A T	T C A T T T G A T G	A T G A T T C A T G	30
31	C G C G A T T C A T	T A G A T G A T G A	C C C C T T T C A T	60
61	T T C C A T T C A A	T G G A G G A T T C	C A T T C G G T T C	90
91	C A T			93

green monkey (*Cercopithecus aethiops*) contains several million copies of a 172 bp sequence which constitute a variable proportion of the genome. This sequence is a major component of the α-satellite which can make up 20–25 % of the nuclear DNA. In a BSC-1 cell line however the 172 bp fragment makes up only 7 % of the genome (Rosenberg, Singer & Rosenberg, 1978) and there are the expected variants (Thayer, Singer & McCutchan, 1981).

Since it is difficult to readily appreciate the sequence alterations from the mass of data of Table 11, we have summarized them as deviations from a consensus sequence (Fig. 1). It can be seen that some parts of the sequence have more base changes than others (see also Donehower *et al.* 1980). No long run of bases is absolutely conserved, the longest uninterrupted stretch being CAGAGTT. In total, only 62 of the possible 172 positions are unaltered by base changes or deletions so that sequence conservation is far from complete. In keeping with the nonrandom character of changes, it can be seen that the first 50 bases of the non-human sequences are hardly perturbed.

One hypothesis is that an ancestral sequence of 172 bp or its dimer occurred in the progenitors of these four primates, and has been inherited and undergone non-random sequence divergence. It has probably been altered by amplification, base changes, deletion/addition events and possibly rearrangements.

Table 11. *Major average nucleotide sequences of restriction fragments from human, African green monkey (AGM), west African baboon and bonnet monkey.*

human 1	AGAATTCTCA	GTAACTTCCT	TGTGTTGTGT		30
human 2	ATGATTCTCA	GAAACTCCTT	TGTGATGTGT		
AGM	AGCTTTCTGA	GAAACTGCTC	TGTGTTCTGT		
baboon 1	AGCTTTCTGA	GAAACTTCTT	TGTGTTCTGT		
baboon 2	AGCTTTCTGA	GAAACTGCTT	AGTGTTCTGT		
bonnet 1	AGCTTTCTGA	GAAACTTCTT	TGTGTTCTGT		
bonnet 2	AGCTTTCTGA	GAAACTGCTT	AGTGTTCTGT		
human 1	GTATTCAACT	CACAGAGTTG	AACGATCCTT		60
human 2	GCGTTCAACT	CACAGAGTTT	AACCTTTCTT		
AGM	TAATTCATCT	CACAGAGTTA	CATCTTTCCC		
baboon 1	GAAATCATCT	CACAGAGTTA	CAGCTTTCCC		
baboon 2	TAATTCATCT	CACAGAGTTA	CATCTGTATT		
bonnet 1	GAAATCATCT	CACAGAGTTA	CAGCTTCCCC		
bonnet 2	TAATTCCTCT	CGCAGAGTTA	CATCTGTATT		
human 1	TACACAGAGC	AGACTTGAAA	CACTCTTTTT		90
human 2	TTCATAGAGC	AGTTAGGAAA	CACTCTGTTT		
AGM	TTCAAGAAGC	CTTTCGCTAA	GGCTGTTCTT		
baboon 1	CTCAAGAAGC	CTTTCGCTAA	GACAGTTCTT		
baboon 2	TCGTGGATCT	CTTTGCTAGC	CTTATTTCT-		
bonnet 1	CTCAAGAAGC	CCTTCGCTAA	GACAGTTCTT		
bonnet 2	TCGTGGATCT	CTTTGCTAGC	CTTATTTC-T		
human 1	GTGGAATTTG	CAAGTGGAGA	TTTCAG-CCG		120
human 2	GTAAAGTCTG	CAAGTGGATA	TTCAGA-CCT		
AGM	GTGGAATTGG	CAAAGGGATA	TTTGGAAGCC		
baboon 1	GTGGAATTGG	CAAAGTGATA	TTTGGAAGCC		
baboon 2	GTGGAATCTG	AGAACAGATA	TTTCGGATCC		
bonnet 1	GTGGAATTGG	CAAAGTGATA	TTTGGAAGCC		
bonnet 2	GTGGAATCTG	AGAACAGATA	TTTCGGATCC		
human 1	CTTTGAGGTC	AATGGTAGAA	TAGGAAATAT		150
human 2	CTTTGAGGCC	TTCGTTGGAA	ACGGGATT-T		
AGM	CATAGAGGGC	TATGGTGAAA	AAGGAAATAT		
baboon 1	CATAGAGGGC	TATGGTGAAA	AAGGAAATAT		
baboon 2	CTTTGAAGAC	TATAGGGCCA	AAGGAAATAT		
bonnet 1	CATAGAGGGC	TATGGTGACC	AAGGAAATAT		
bonnet 2	CTTTGAAGAC	TATAGGGXAA	AAGGAAATAT		
human 1	CTTCCTATAG	AAACTAGACA	GA		172
human 2	CTTCATATTA	TG-CTAGACA	GA		
AGM	CTTCCGTTCA	AAACTGGAAA	GA		
baboon 1	CCTCAGATGA	AATCTGGAAA	GA		
baboon 2	CCTCCGATAA	CAAAGAGAAA	GA		
bonnet 1	CCTCCGATAA	CAAAGAGAAA	GA		
bonnet 2	CCTCAGATGA	AATCTGGAAA	GA		

Conclusions from the primate data

Remembering that the primate data have been obtained overwhelmingly from large restriction fragments, rather than clones derived from buoyant density satellites, it is found that:

(1) Sequence complexities appear longer on the average than most highly repeated DNAs, but this is likely to be due to a sampling component since at least human satellite III has a very simple substructure (ATTC).

(2) The proportion of a given restriction fragment can vary from less than 1% to near 20% of the diploid genome.

(3) The four species examined share variants of a 172 bp consensus sequence.

(4) The major repeat lacks a definitive substructure.

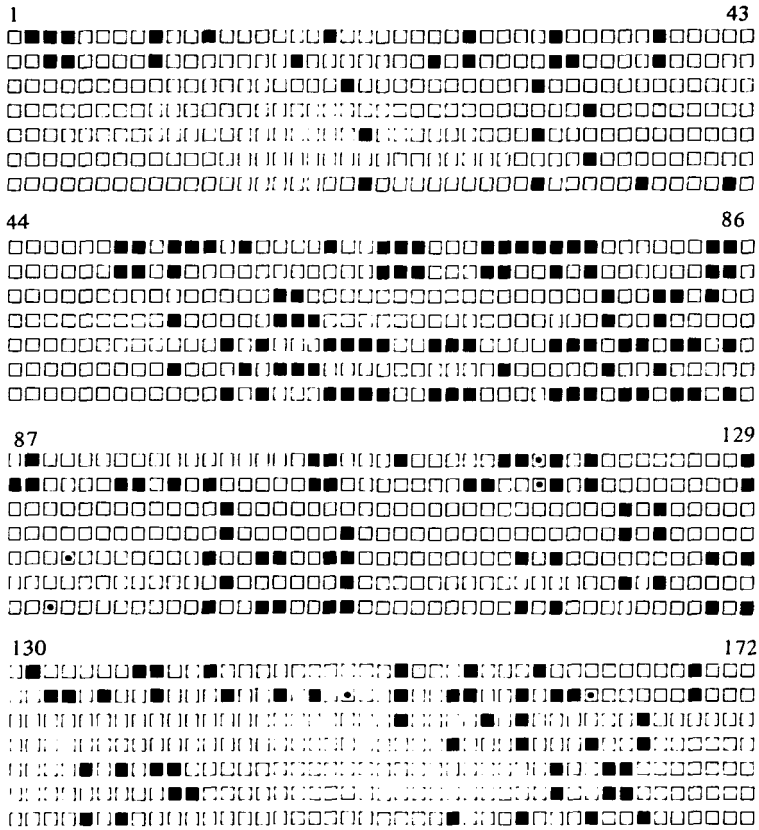

Fig. 1. Visual display of Table II illustrating the *between* species sequence alterations in the 172 bp consensus sequence. ■ Base changes; ⊡, deletions.

(5) The *within* species heterogeneity is unknown, but more than 60% of the sequence is not absolutely conserved *between* species.

(6) The sequence can be perturbed either by deletions or base changes and its evolution may have been constrained by non-random unequal crossovers (Donehower *et al.* 1980). Whether the nonrandom nature of the changes has biological significance has yet to be decided.

PART 2. DATA ON SEQUENCE VARIATION AND SECONDARY STRUCTURE

A widely held belief is that there is little sequence variation in highly repeated DNAs within or between species. The level of variation within a species is generally discussed in terms of *homogeneity* whereas that between species is referred to as *conservation*. There is confusion over these terms for a number of reasons. Firstly, *conservation* has not been strictly defined. Secondly, although it is believed that satellite sequences are highly conserved for important functional reasons, these reasons are speculative. Thirdly, it is assumed that if a low level of sequence variation is found within or between species, this automatically indicates functional significance. As we will point out below, unless we have a measure of heterogeneity within a species we are unable to make accurate statements about the degree of conservation between related species.

A point which is frequently overlooked is that the low level of variation within a stretch of repeated DNA must depend to a considerable extent on the time since the most recent amplification, as well as on the efficiency of the mechanisms which monitor or alter an array. Homogeneity need have no functional significance whatsoever. Ideally, measurements of variation depend on being able to obtain reasonable samples of most of the sequence variants within a species but in tandemly repeated arrays this task is complicated by processes of amplification and deletion. Amplification of a variant to a high copy number predisposes it to analysis owing to its ease of isolation. Such amplification also obscures those variants which remain in low copy numbers. Thus a meaningful prerequisite to obtaining estimates of conservation is to determine the levels and origins of variation within a species. As we shall see, this has not been systematically attempted for any system.

Within species: How much sequence variation?

(i) *The Calf*

The level of variation within a mammalian species can be gauged from cloned material derived from the 1·706 g/cm^3 satellite of the calf. Once again the DNA is from an unknown number of individuals so that the total variation cannot be accurately partitioned. However since the DNA has been cloned, the variation between adjacent repeating units on a chromosome can be estimated from any particular clone (Table 12, Fig. 2, Pech *et al.* 1979*b*). It is clear that adjacent 23 bp repeating units differ markedly from each other. Thus significant sequence variation is already present on a single chromosome.

In terms of the 23 bp sequence itself, the 8 clones reveal that 20 of the possible 23 positions are altered by base changes or deletions. In conventionally accepted molecular terms these 8 samples yield an estimate of 12 % variation (total number of base changes and deletions divided by the total number of nucleotides). Furthermore, the distribution of perturbations within the 23 bp unit is not random with the middle of the repeat suffering more heavily than either extreme (Fig. 2).

(ii) *The kangaroo rat*

Table 13 shows the variants from the HS-α satellite of *Dipodomys ordii* obtained from uncloned material from a number of individuals (Fry & Salser, 1977). Many variants other than TTAGGG do occur, but their relative yields are lower than the minimum cut-off point in this table. The level of variation is 15%. Clearly a population of sequence variants exists in which the variation is very similar to that of the 23 bp unit of the calf.

Table 12. *Nucleotide sequences of* 8 *clones from the* Sau *segments of the calf 1·706 g/cm³ satellite. Deletions are denoted —*

Clone 22		G A T C A T G C	A C T	23
	G T T C A C G T G A	C T		
Clone 23		G A T C A T A C	A C T	
	G C T C A C G T A A	C T G A T G A T G C	A C T	
	G A T C A C G T G G	C T – A T C A T G C	A C T	
Clone 42		G A T C A T T C	A C T	
	G A T C A C G T G A	C T G A C C A T G C	A C G	
	G A T C A C G T G A	C T		
Clone 46		G A T C C T G C	A C T	
	G A T C A C A T G G	C T C A T C A C G C	A C T	
	G A T C A C G T G A	G T A A T C A T G C	A C T	
	C T T C A C G T G C	A T G A T C A T G C	A C T	
	G A T C A C G T G G	C T – A T C A T G C	A C T	
	G C T C A C G T G G	C T		
Clone 3691	G A T C A C G T G A	C T – T A A T G G C	G C T	
	G A T C A C G T G G	C T – A T C A T G C	A C T	
	G A T C A C G T G G	A T – A T C A T G C	A C T	
Clone 3688		G A T C A T G C	A C T	
	G A T A C C G T G A	C T G A G C A T G G	A C G	
	G A T C A C G T G G	C T – A T C C T G C	A C T	
Clone 176	G A T C A C G T C T	A T – A T C A T G C	A C T	
	G A T C A C A T G A	C T G C C C A T G C	A C T	
	G A T C A C G T G A	T		
Clone 101		G A T C A T G C	A C T	
	G A T C A C G T G G	C – T A T C A T G C	A C T	
	G C T C A C G T G G	C T G A T C A T G C	A A T	
	G A T C A C G T A G	C T – A C C A T G C	A C T	
	G G T C A C G T G A	C T G A T C G T G C	A C T	
	G C T C A G G T A A	T T G A T C A T G C	A C T	
	G A T C A C G T G A	C T		

(iii) *Hawaiian* Drosophila

The variation in a complex invertebrate repeat is provided by cloned material from pooled *D. gymnobasis* DNA from a laboratory strain (Miklos & Gill, 1981). Here 183 positions are available for scrutiny and the level of variation is 11%, a figure similar to the 12 and 15% in the mammalian cases where the repeat unit

length is very much smaller (Table 14 and Fig. 3). Since the cloning protocol in this case yields only single repeating units, we are unable to estimate the variation between adjacent repeats on a chromosome.

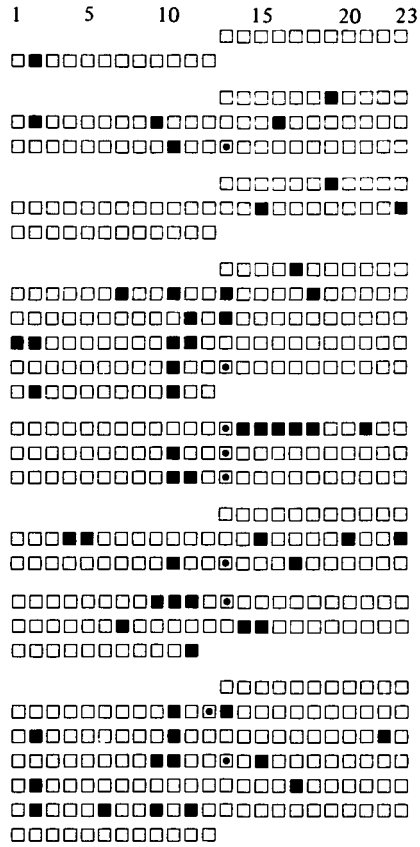

Fig. 2. Visual display of Table 12 illustrating the *within* species sequence alterations in the 23 bp repeating unit of the calf. Base changes; ⊡, and deletions.

(iv) D. melanogaster

In the 359 bp variant of *D. melanogaster* the cloned sequences reveal that the variation between adjacent repeats on a chromosome can be as low as 4 % (Hsieh & Brutlag, 1979*a*). However it should be borne in mind that if one examines independent clones of the same sequence such as the 254 and 359 bp variants, there can be as little as 60 % homology. Thus whilst regions of a chromosome can be fairly homogenous in sequence, an enormous variation can still exist within a single laboratory strain.

https://doi.org/10.1017/S0016672300020711 Published online by Cambridge University Press

Résumé

The variation at the sequence level is very similar for long or short repeat units from mammals or invertebrates. All four organisms referred to above reveal that portions of a satellite can exist as a population of variants and that base changes and 'deletions' are nonrandomly distributed along a given sequence.

The existence of sequence variants has implications for the putative importance of protein/satellite DNA binding. If there are proteins which bind to the variant 23 bp sequences of the calf (in which all but 3 positions suffer some form of alteration) then two possibilities exist: either (1) the base sequence recognizing

Table 13. *Variant sequences and their relative yields from the heavy strand of the HS-α satellite of the kangaroo rat*

Sequence	Relative yield
T T A G G G	1·0
T T A G A G	0·5
T T A G G T	0·5
T G A G G G	0·5
T T A G G A	0·3
T T A G T G	0·3
T A A G G G	0·2
T T A G T T	0·2
G A A G G G	0·2
T T A G G C	0·2
T G A G G A	0·1
T T A A G G	0·1

abilities of these theoretical proteins may be fairly nonspecific or (2) if the interactions are specific, then the number of protein variants will approach the number of DNA sequence variants. Taking into consideration the known variants for just the 1·706 g/cm^3 satellite (Tables 7–9; Fig. 2) this would mean that the number of binding proteins for just one satellite of the calf would exceed the normal coding capacity of the genome. A plausible interpretation is that repeated arrays diverge without significant functional constraints.

Between species: How much sequence variation?

Without an accurate estimate of sequence variation *within* a species, it is not possible to meaningfully evaluate 'conservation' between species. For example, heterogeneity within an ancestral species might now be represented as within or between species variation in new species. Alternatively ancestral sequences may have been totally homogenous at one time and have diverged, irrespective of speciation, to yield the observed spectrum of variation. We shall summarize the existing sequence data with these constraints in mind.

Table 14. *Nucleotide sequences of 10 clones of* D. gymnobasis. *Deletions are denoted – and undetermined bases by* X

6	1	GAATTCACAC	GGCTGTTGAC	ATAAAACTGC	30
15		GAATTCGCAC	GGCTGTTGAC	ATAAAACTGC	
9		GAATTCGCAC	GGCTGTTGAC	ATAAAACTGC	
7		GAATTCGCAC	GGCTGTTGAC	AAAAAACTGC	
10		GAATTCGCAT	GAATGTTGAA	ACAAAACTGC	
12		GAATTCGCAT	GAATGTTGAA	ACAAAACTGC	
14		GAATTCXXAT	GAATGTTGAA	ACAACACTGC	
11		GAATTCGCAT	GAATGXTGAA	ACAAAACTGC	
8		GAATTCGCAT	GAATGTTGAA	ACAACACTGC	
5		GAATTCGCAT	GAATGTTGAA	ACAACACTGC	
6		ATTTAAAAGA	ATGCACTTTA	GACTCATCTA	60
15		ATTTAAAAAA	ATGGACTTTA	GATTCATTTA	
9		ATTTAAAAAA	ATGGACTTTA	GACTCATCTA	
7		ATTTAAAAAA	ATGCACTTTA	GACTCATCTA	
10		ATTTAAAAAA	ACTGACTTCA	GACTCATCTT	
12		ATTTAAAAAA	ATTGACTTTA	GACTCATCTA	
14		ATTTAAAAAA	ATTGACTTCA	GACTCATCTT	
11		ATTTAAAAAA	ACTGACTTCA	GACTCATCTT	
8		ATTTAAAAAA	ATTGACTTCA	GACTCATCTA	
5		ATTTAAAAAA	ATTGACTTCA	GACTCATTAA	
6		AGAAAGATAT	TAAGGGATAT	CC–TCAAAAA	90
15		AGAAAGATAT	TAAGGGATAT	CC–TCAAAAA	
9		AGAAAGATAT	TAAGGGATAT	CC–TCAAAAA	
7		AGAAAGATAT	TAAGGGATAT	CC–TCAAAAA	
10		AGAAAGATAT	TAACGGATAT	CCGTCAAAAA	
12		AGAAAGATAT	TAACGGATAT	CCGTCAAAAA	
14		AGAAAGATAT	TAACGGATAT	CCGTCAAAAA	
11		AGAAAGATAT	TAACGGTTTT	CCGTCAAAAA	
8		AGAAAGATAT	TAACGGATAT	CCGTCAAAAA	
5		AGAAAGATAT	TAACGGATAT	CCGTCAAAAA	
6		ATCGAAAA––	––––––––––C	ATGACATAGG	120
15		ATCGAAAA––	––––––––––C	ACGACATAGG	
9		ATCGAAAA––	––––––––––C	ATGACATAGG	
7		ATCGAAAA––	––––––––––C	ACGACATAGG	
10		CTCGA–––TA	TAAGTTGA––	–––––ATAGC	
12		CTCGA–––TA	TAAGTTGA––	–––––ATAGC	
14		CTCGA–––TA	TAAGTTGA––	–––––ATAGC	
11		CTCGA–––TA	TAAGTTAA–C	ATGACATAGG	
8		CTCGA–––TA	TAAGTTAA–C	ATGACATAGG	
5		CTCGA–––TA	TAAGTTAA–C	ATGACATAGG	
6		ATCATCCTTT	AAACAAATAG	CCATATCTTC	150
15		ATTATCCTTT	AAACAAATAG	CCATATCTTC	
9		ATXATCCTTT	AAACAAATAG	CCATATCTTC	
7		ATCATCCTTT	AAACAAATGG	CCATATCTTC	
10		CTCATCCTAT	AAACAAATGG	CCATATCTTC	
12		CTCATCCTAT	AAACAAATGG	CCATATCTTC	
14		CTCATCCTAT	AAACAAATGG	CCATATCTTC	
11		CTCATCCTAT	AAACAAA–––	–––TATCTTC	
8		CTCATCCTAT	AAACACCTGG	CCATATCTTC	
5		CTCATCCTAT	AAACACCTGG	CCATATTTTC	

https://doi.org/10.1017/S0016672300020711 Published online by Cambridge University Press

Table 14—(*cont.*)

6	G T C A A A A A T T	G T C C A A A T T X	A A A T A T T A A G	180
15	G T C A A A C A T T	G T C C A A A T T C	A A A T A T T A A G	
9	X X X A A A A A T T	G X X X X X X X X X	A A A T A T T A A G	
7	G T C A A A A A T T	G T C C A A A T T X	A A A T A T T A A G	
10	G T C A A A A A A X T	G T C C A T T T T C	C A A T A T T A A G	
12	G T C A A A A A T T	G T C C A A T T T C	C A A T A T T A A G	
14	G T C A A A A A T T	G T C C A A T T T C	C A A T A T T A A G	
11	G T C A A A A A T T	G T C C A A T T T C	A A A T A T T A A G	
8	G T C A C A A A T T	G T C C A A T T T C	C A A T A T T A A G	
5	G T C A A A A A T T	G T C C A A T T T C	T A A T A T T A C G	
6	G T G T T T T T T A			189
15	G T G T T T T T T A			
9	G X G T T T T T T A			
7	G T G T T T T T T A			
10	G T A G T T T T T G			
12	G T A G T T T T T G			
14	G T A G T T T T T G			
11	G T A G T T T T T G			
8	G T A G T T T T T G			
5	G T A G T T T T T G			

(i) *Primates*

The conventional yardstick by which the functional significance of any DNA sequence is evaluated is how well it has fared in a comparative evolutionary sense. The primate data of Fig. 1 and Table 2 provide minimum estimates of change for a DNA of specific fragment length. The African green-monkey sequence differs by 9% from a consensus sequence, the baboon and bonnet monkey by 29 and 31% respectively, and the human by 39%. Furthermore, only 36% of the 172 positions are free of mutational or deletional alterations. In spite of the non-randomness of the changes this sequence would be regarded as having significantly diverged as a result of evolutionary events spanning at least 20 million years.

(ii) *Hawaiian* Drosophila

D. gymnobasis and *D. silvarentis* are very close relatives, their estimated divergence times lying between 100000 and 1 million years. *D. grimshawi* is considered a more distant member of the same group. Since the oldest Hawaiian island is less than 6 million years of age, and since the radiation of most of the Hawaiian *Drosophila* occurred on these younger islands, 6 million years is an absolute upper limit to any divergence times.

Comparison of the consensus sequence of *D. gymnobasis* with *D. silvarentis* (Table 3) reveals that the two are virtually indistinguishable. With two such extremely close homosequential species, it is perhaps not so surprising that speciation occurred in the apparent absence of significant sequence alteration in this particular restriction fragment.

Comparison of the *D. gymnobasis* consensus sequence with *D. grimshawi* on the

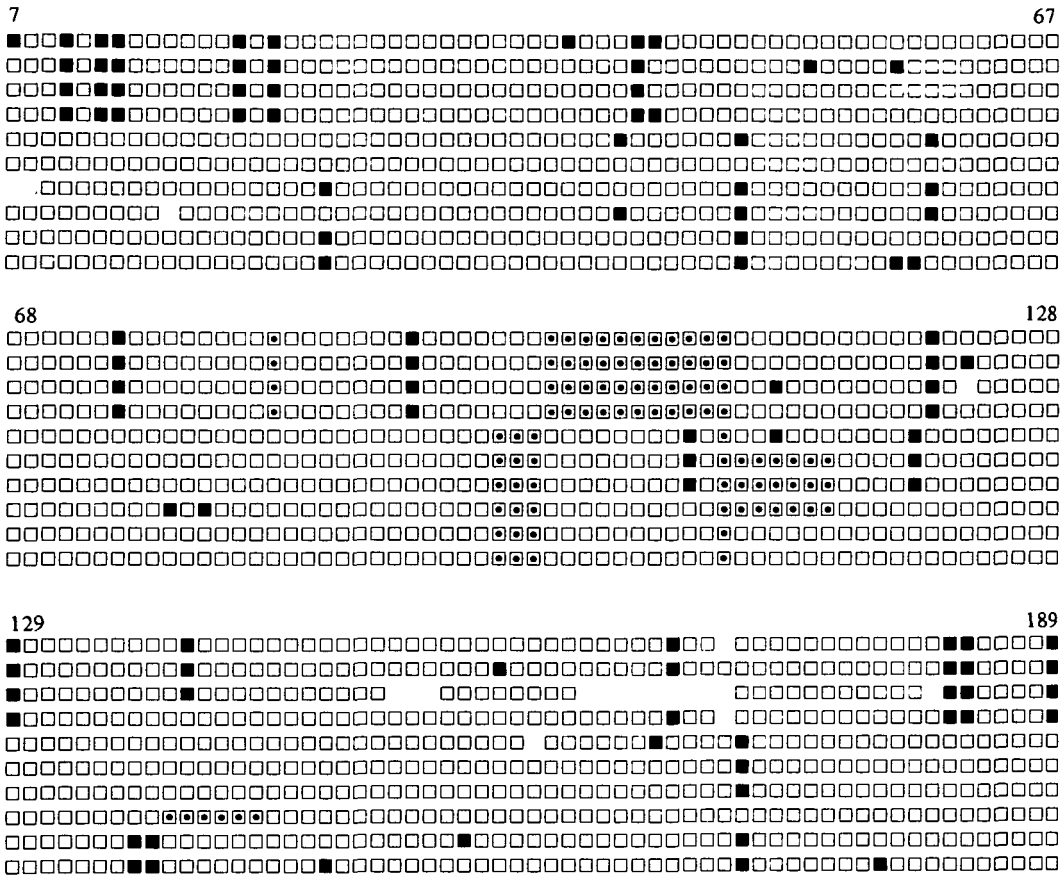

Fig. 3. Visual display of Table 14 illustrating the *within* species sequence alterations in the 189 bp unit of *D. gymnobasis*. ■, Base changes; ⊡, deletions.

other hand, shows that they remain absolutely unaltered at only 130 of the 183 positions. If we do not include the substantial differences due to deletions, but consider only those positions altered by base changes, then *D. gymnobasis* and *D. grimshawi* differ by approximately 15%.

Conclusions on levels and rates of variation

DNA sequences which code for protein products face stringent constraints and show characteristic levels of evolutionary change. As an example, the exonic regions of the β-globins of the mouse show only 5% divergence after 50 million years. Compare this with changes in repeated arrays. In conventional terms the human/African green monkey divergence would be 35% after 20–50 million years, and the Hawaiian *Drosophila* comparison would be a minimum of 15% after only 6 million years.

However, such conventional interpretations are in fact invalidated because of the substantial within species sequence variation. It is not possible with the existing data to estimate rates of divergence in repeated arrays. A further complication is that whilst some genic sequences occur in just one copy per genome, repeated arrays are subjected to continuing cycles of amplification and deletion.

The kinds of variation that are seen at the sequence level in repeated arrays can however be meaningfully compared to those in genic sequences. Repeated sequences seem to have accumulated deletions of a variety of lengths as well as base changes in almost any position. In this respect they may conform to an extension of the neutralist viewpoint about allelic variation. If we draw an analogy to Kimura's (1979) theory, then its general applicability to repeated sequences *per se* would be as follows:

(1) Sequences are selectively neutral if they are neither more nor less advantageous than the ones they replace.

(2) Neutral theory does not assume that sequences are functionless, only that different sequences have about the same effect on fitness.

(3) As the functional constraints on a molecule diminish, so the evolutionary rate of mutant substitutions increases.

As the mechanisms involved in generating variation differ for tandem arrays and alleles, the analogy with Kimura's theory cannot be taken to predict levels of variation in tandem arrays within a species as it does for allelic variation.

The future of satellite sequence analyses

It must be quite obvious that the structural approach has proved disappointing in a functional sense. It is also clear that sequence analysis is at too fine a level to discriminate between various alternatives. For example, repeated arrays that appear as if they arose from unequal exchange events can theoretically come about by a number of disparate mechanisms such as unequal sister chromatid exchange (Petes, 1980; Szostak & Wu, 1980), intrachromosomal conversion (Klein & Petes, 1981) or even extrachromosomal plasmid replication combined with subsequent plasmid recombination. To distinguish between such mechanisms for satellite sequences one needs to ask questions such as (*a*) do these plasmids actually exist, and (*b*) how much spontaneous USCE actually occurs? More detailed sequence analysis will not help solve the problem. The only presently convincing data for USCE in tandem arrays comes from ribosomal genes (Petes, 1980; Szostak & Wu, 1980).

Dyadic symmetries

One structural feature of a DNA sequence which is often held to be of potential functional significance for protein/DNA interactions is a region of dyadic symmetry. The primate data of Table 11, allow us to evaluate the conservation of such regions. Consider for example the nucleotide sequence from positions 163 to 026. In the bonnet monkey there are 15 matched base pairs out of 18 positions whereas

only 9 matched pairs occur in the human sequence and these 9 are distributed in such a way as to make the resultant hairpin highly improbable (Fig. 4).

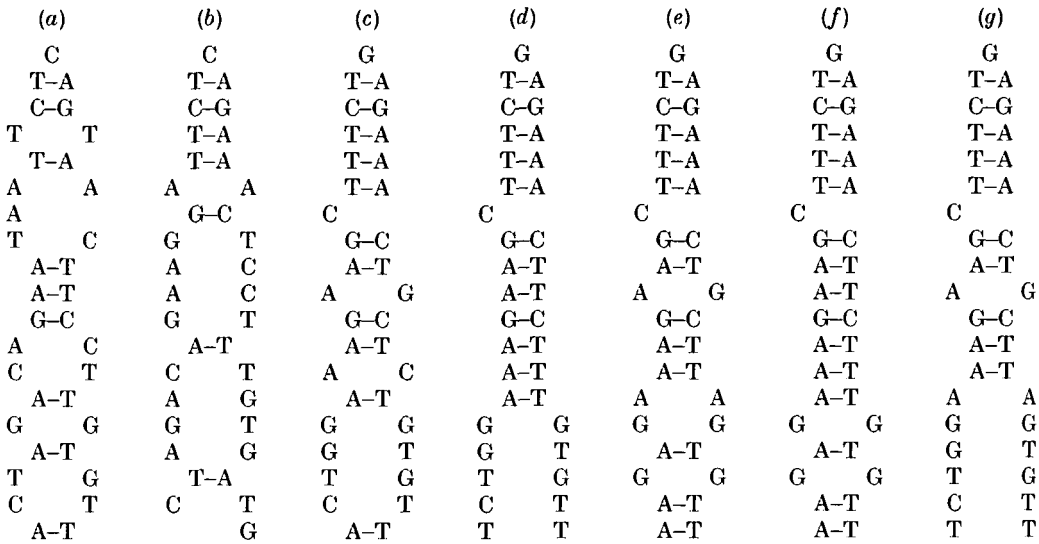

Fig. 4. Potential dyadic symmetries in the various primate sequences: (*a*) human-1; (*b*) human-2; (*c*) African green monkey; (*d*) baboon-1; (*e*) baboon-2; (*f*) bonnet-1 and (*g*) bonnet-2.

Other dyadic symmetries also occur but suffer the same fate. In the human-1 sequence, positions 034–041 and 046–052 form a perfect match

$$\underline{\underline{\text{TTCAACTC}}}$$
$$\underline{\underline{\text{AAGTTGAG}}}.$$

The equivalent position in the bonnet monkey (bonnet-2) shows a badly mismatched sequence

$$\begin{array}{ccc} & \text{TCCT} & \\ \underline{\text{T}} & & \underline{\underline{\text{CTC}}} \\ \underline{\text{A}} & & \underline{\underline{\text{GAG}}} \\ & \text{CATT} & \end{array}$$

If symmetry elements such as these were to be of significance, their time course of usage and obsolescence must be very short in evolutionary terms. Wu & Manuelidis (1980) have extensively analysed the primate data and pointed out that dyadic symmetries are not of convincing importance. They stress that 'overall length and tandem repetition are the critical features of satellite DNA rather than individual repeat length or secondary structure potential'. A similar point relating to sequence *per se* has been emphasized by Miklos & Gill (1981).

https://doi.org/10.1017/S0016672300020711 Published online by Cambridge University Press

Structural features

Can a given repeated DNA sequence influence the local molecular geometry of a helix and so endow it with biologically interesting properties such as protein binding (Klug *et al.* 1979) or chromatin condensation? It is known that poly (dA-dT) . poly (dA-dT) can exist in a special D form of DNA rather than the more usual A and B forms. Can satellite DNA of unusual base composition do likewise? Selsing & Arnott (1976) studied three satellite DNAs of varying complexities precisely from this point of view, these being an almost pure d(AT) from crab, the ACAAACT sequence from *D. virilis* and the classical mouse satellite. All three DNAs displayed only the orthodox conformations, leading the authors to conclude that hypotheses which held highly repeated DNA as being important to chromosome folding were unsupported. For *D. virilis* at least, the compaction of satellite DNAs into heterochromatin may well be achieved by phosphorylation of histone H1 (Blumenfeld *et al.* 1978; Billings *et al.* 1979).

DISCUSSION

The molecular data presented in this article represent the backbone of the current sequence knowledge. We now turn to the remaining molecular and biological data and provide an array of recent references which yield a cross section of the various issues, the relative importance of many of which is in dispute. We have also included a small number of references which do not deal with satellite DNA directly, but which we think are relevant to the hypotheses and data under consideration. Finally we present our own particular viewpoint which in many respects departs radically from the more conventional paths in this field.

The reference base

(1) *Structural characteristics of sequences*

Brutlag (1980), Chambers *et al.* (1978), Christie & Skinner (1980*a*, *b*), Cooke & Hindley (1979), Donehower *et al.* (1980), Horz & Altenburger (1981), Miklos & Gill (1981), Mullins & Blumenfeld (1979), Pech *et al.* (1979*a*, *b*), Poschl & Streeck (1980), Rubin *et al.* (1980), Salser *et al.* (1976), Schmookler-Reis & Biro (1978), Southern (1970), Rubin *et al.* (1980), Thayer *et al.* (1981), and Wu & Manuelidis (1980).

(2) *Sequence arrangements, rearrangements, expansion and contraction of arrays, polymorphisms*

Beauchamp *et al.* (1979), Bedbrook *et al.* (1980*a*, *b*), Bostock *et al.* (1978), Brown & Dover (1979, 1980*a*, *b*), Brutlag (1980), Cordeiro-Stone & Lee (1976), Flavell, O'Dell & Hutchinson (1981), Gosden *et al.* (1981), Holmquist & Dancis (1979), John (1981), Klein & Petes (1981), Kurnit (1979), McKay, Bobrow & Cooke (1978), Miklos *et al.* (1980), Petes (1980), Roizes, Pages & Lecou (1980), Rosenberg *et al.* (1978), Scherer & Davis (1980), Szostak & Wu (1980), and Wheeler *et al.* (1978).

(3) *Sequence configurations and DNA binding proteins*

Comings *et al.* (1977), Hseih & Brutlag (1979*a*, *b*), Kao-Huang *et al.*(1977), Klug *et al.* (1979), Lin & Riggs (1975), Selsing & Arnott (1976), Will & Bautz (1980).

(4) *Periodicities in repeated arrays*

Maio, Brown & Musich (1977), Brown, Musich & Maio (1979), Igo-Kemenes, Omori & Zachau (1980), Musich, Brown & Maio (1977, 1980).

(5) *Compaction of DNAs*

Billings *et al.* (1979), Blumenfeld *et al.* (1978), Chahal, Matthews & Bradbury (1980), Comings & Okada (1976), Matsumoto *et al.* (1980).

(6) *Transcription of repeated sequences*

Varley, Macgregor & Erba (1980).

(7) *Chromosome mechanics and germ line effects*

Bostock (1980), John & Miklos (1979), John (1981), Maynard-Smith (1977), Miklos & John (1979), Rees & Dale (1974), Yamamoto & Miklos (1977, 1978), Yamamoto (1979*a*, *b*).

(8) *Somatic effects, nucleotype, C-value paradox*

Barlow (1973), Bennet (1971), Cavalier-Smith (1978, 1980*a*, *b*), Hutchinson *et al.* (1979), Macgregor (1980), Miklos (1981), Olmo & Morescalchi (1978).

(9) *Evolution*

Amos & Dover (1981), Barnes, Webb & Dover (1978), Cohen & Bowman (1979), Cseko *et al.* (1979), Dover (1978, 1980), Flavell *et al.* (1981), Holmquist & Dancis (1979), Miklos & Gill (1981), Musich, Brown & Maio (1980), Stanley (1975).

(10) *General reviews, selfish DNA, neutrality, selection*

Bostock (1980), Brutlag (1980), Cavalier-Smith (1978), Doolittle & Sapienza (1980), Dover (1980), John & Miklos (1979), Kimura (1979), Kurnit (1979), Macgregor (1980), Nagl (1978), *Nature*, News and Views (1980*a*, *b*), Orgel & Crick (1980), Walker (1979).

Viewpoints

The bulk of the recent data understandably deals with sequence structure and associated spinoffs. Conventionally it is thought that a continuing analysis of sequence structure will provide a satisfying array of solutions at the *cellular* level to the present imbroglio concerning highly repeated sequences. For the solution of some problems this is the only sensible direction in which to proceed, since arrays of tandem sequences can be used to analyse fundamental problems such as chromatin compaction, DNA binding proteins, gene regulation as regards position effect variegation and nucleosomal phasing.

However solutions to such fundamental problems do not neccessarily enlighten us on the functions (if any) of satellite DNAs *per se*. As we saw from the analyses of DNA sequence data, the main characteristic of satellites is their extensive variation. The magnitude of the biological problem comes home with some force when it is realized that many sequences breeze in and breeze out of genomes in rather short evolutionary time periods, such as the 'spring cleaning' that occurs in ribosomal cistrons (Dover & Coen, 1981). It ought to be somewhat disturbing to begin molecular characterization of a sequence in a given species when it is virtually guaranteed that nature will have supplanted it with another in a different species. The critical factor will of course be a lack of generality in extrapolating the results.

We believe on the other hand that the available data are telling us something quite different. The central issue is not really *how* the replicative and recombinative machinery churns out such a bewildering pastiche of sequences, but whether the final product in a particular chromosomal location has biological significance. Does it really matter if the increase in DNA on a chromosome is brought about by the insertion of an extra chromosomally replicating plasmid, unequal sister chromatid exchange or some as yet undefined conversion mechanism? The end result is what is important, namely a tandem pack of sequences. The issue is the ability of the organism or the population to cope with such added arrays arising in the germ line. The important question of elucidating mechanisms for their own sake is separate and not necessarily germane to our biological enquiries.

Satellite sequences have little somatic effects on an organism (Kurnit, 1979; Hutchinson, Rees & Seal, 1979; Miklos, 1981) as is also evidenced by the ease with which polymorphisms are tolerated. In the germ line however satellite sequences can exert significant influences by rendering rec^- long regions of neighbouring DNA (John & Miklos, 1979). The consequences of such effects are *not* on the phenotype of the organism, but on the potential array of DNA sequences in succeeding generations (see Rees & Dale, 1974). This general category of germ line effects has been missed in formulations on selfish DNA for example, where the overlying emphasis has been on *cellular* events and *mechanisms*.

Conclusions

Although the precise mechanisms themselves are unclear, we have reached the following views about the generation, maintenance and significance of highly repeated sequences organized in tandem arrays.

(1) The demonstrated biological effects of such sequences are predominantly in the germ line.

(2) One consequence of the experimentally induced changes in the amounts of localized sequences is that recombinational probabilities along a chromosome can be modified. Whether these effects have adaptive or evolutionary significance has yet to be experimentally determined.

(3) The incidence of deletion/addition events of virtually any length as well as the almost unconstrained occurrence of base changes indicates that the sequence *per se* of a basic repeating unit may well be profitably examined by analogy to Kimura's neutralist theory.

(4) Repeated arrays are continually generated, expanded, reduced and dispersed as a natural consequence of the inbuilt quirks of the replicative and recombinative machinery of eukaryotic cells. The somatic effects of such changes are generally small within the boundary conditions that normally occur in natural polymorphisms.

(5) Repeated arrays continuously fluctuate in size, but are constrained by the low density of genes within them. An array cannot be expanded or deleted with impunity since serious genetic effects may arise as a consequence of gene deletion or amplification. Note that this damage is not done by repeated sequences *per se* but by the genes sequestered within them.

(6) The wholesale rearrangement and movement of arrays can initially be seriously hampered by position effect variegation for genes contained within them (Baker, 1971). Furthermore the rearrangment of such arrays is constrained by the needs of centromeres and telomeres to function properly (Holmquist & Dancis, 1979).

We are indebted to our colleagues, Des Clark-Walker, Adrian Gibbs, John Gibson, Bernard John and Hiroto Naora for their stimulating discussions and for gently clarifying our misconceptions.

REFERENCES

Amos, A. & Dover, G. A. (1981). The distribution of repetitive DNAs between regular and supernumerary chromosomes in species of *Glossina*: a two step process in the origin of supernumeraries. *Chromosoma* **81**, 673–690.

Baker, W. K. (1971). Evidence for position effect suppression of the ribosomal RNA cistrons in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences (USA)* **68**, 2472–2476.

Barlow, P. (1973). The influence of inactive chromosomes on human development. *Humangenetik* **17**, 105–136.

Barnes, S. R., Webb, D. A. & Dover, G. (1978). The distribution of satellite and main-band

DNA components in the *melanogaster* species subgroup of *Drosophila*. *Chromosoma* (*Berlin*) **67**, 341–363.

Beauchamp, R. S., Mitchell, A. R., Buckland, R. A. & Bostock, C. J. (1979). Specific arrangements of human satellite III DNA sequences in human chromosomes. *Chromosoma* (*Berlin*) **71**, 153–166.

Bedbrook, J. R., Jones, J., O'Dell, M., Thompson, R. D. & Flavell, R. B. (1980*a*). Molecular characterization of telomeric heterochromatin in *Secale* species. *Cell* **19**, 545–560.

Bedbrook, J. R., O'Dell, M. & Flavell, R. B. (1980*b*). Amplification of rearranged sequences in cereal plants. *Nature* **288**, 133–137.

Bennett, M. D. (1971). The duration of meiosis. *Proceedings of the Royal Society* B. **178**, 259–275.

Billings, P. C., Orf, J. W., Palmer, D. K., Talmage, D. A., Pan, C. G. & Blumenfeld, M. (1979). Anomalous electrophoretic mobility of Drosophila phosphorylated H1 histone: is it related to the compaction of satellite DNA into heterochromatin? *Nucleic Acids Research* **6**, 2151–2164.

Blumenfeld, M., Orf, J. W., Sina, B. J., Kreber, R. A., Callahan, M. A. & Snyder, L. A. (1978). Satellite DNA H1 histone, and heterochromatin in *Drosophila virilis*. *Cold Spring Harbour Symposium on Quantitative Biology* **42**, 273–276.

Bostock, C. J. (1980). A function for satellite DNA? *Trends in Biochemical Sciences* **5**, 117–119.

Bostock, C. J., Gosden, J. R. & Mitchell, A. R. (1978). Localization of a male-specific DNA fragment to a sub-region of the human *Y* chromosome. *Nature* **272**, 324–328.

Brown, F. L., Musich, P. R. & Maio, J. J. (1979). The repetitive sequence structure of component α DNA and its relationship to the nucleosomes of the African green monkey. *Journal of Molecular Biology* **131**, 777–799.

Brown, S. D. M. & Dover, G. A. (1979). Conservation of sequences in related genomes of *Apodemus:* constraints on the maintenance of satellite DNA sequences. *Nucleic Acids Research* **6**, 2423–2434.

Brown, S. D. M. & Dover, G. A. (1980*a*). The specific organisation of satellite DNA sequences on the *X* chromosome of *Mus musculus*: partial independence of chromosome evolution. *Nucleic Acids Research* **8**, 781–792.

Brown, S. D. M. & Dover, G. A. (1980*b*). Conservation of segmental variants of satellite DNA of *Mus musculus* in a related species: *Mus spretus*. *Nature* **285**, 47–49.

Brutlag, D. L. (1980). Molecular arrangement and evolution of heterochromatic DNA. In *Annual Reviews of Genetics*, vol. 14 (ed. H. L. Roman, A. Campbell and L. M. Sandler), pp. 121–144.

Carlson, M. & Brutlag, D. (1979). Different regions of a complex satellite DNA vary in size and sequence of the repeating unit. *Journal of Molecular Biology* **135**, 483–500.

Cavalier-Smith, T. (1978). Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *Journal of Cell Science* **34**, 247–278.

Cavalier-Smith, T. (1980*a*). r- and k-tactics in the evolution of protist developmental systems: cell and genome size, phenotype diversifying selection, and cell cycle patterns. *BioSystems* **12**, 43–59.

Cavalier-Smith, T. (1980*b*). How selfish is DNA? *Nature* **285**, 617–618.

Chahal, S. S., Matthews, H. R. & Bradbury, E. M. (1980). Acetylation of histone H4 and its role in chromatin structure and function. *Nature* **287**, 76–79.

Chambers, C. A., Schell, M. P. & Skinner, D. M. (1978). The primary sequence of a crustacean satellite DNA containing a family of repeats. *Cell* **13**, 97–110.

Christie, N. T. & Skinner, D. M. (1980*a*). Evidence for nonrandom alterations in a fraction of the highly repetitive DNA of a eukaryote. *Nucleic Acids Research* **8**, 279–298.

Christie, N. T. & Skinner, D. M. (1980*b*). Selective amplification of variants of a complex repeating unit in DNA of a crustacean. *Proceedings of the National Academy of Sciences* (*USA*) **77**, 2786–2790.

Cohen, E. H. & Bowman, S. C. (1979). Detection and location of three simple sequence DNAs in polytene chromosomes from *virilis* group species of *Drosophila*. *Chromosoma* (*Berlin*) **73**, 327–355.

COMINGS, D. E. & OKADA, T. A. (1976). Fine structure of the heterochromatin of the kangaroo rat *Dipodomys ordii*, and examination of the possible role of actin and myosin in heterochromatin condensation. *Journal of Cell Science* **21**, 465–477.

COMINGS, D. E., HARRIS, D. C., OKADA, T. A. & HOLMQUIST, G. (1977). Nuclear proteins. *Experimental Cell Research* **105**, 349–365.

COOKE, H. J. & HINDLEY, J. (1979). Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acids Research* **6**, 3177–3197.

CORDEIRO-STONE, M. & LEE, C. S. (1976). Studies on the satellite DNAs of *Drosophila nasutoides*: their buoyant densities, melting temperatures, reassociation rates and localizations in polytene chromosomes. *Journal of Molecular Biology* **104**, 1–24.

CSEKO, Y. M. T., DOWER, N. A., MINOO, P., LOWENSTEIN, L., SMITH, G. R., STONE, J. & SEDEROFF, R. (1979). Evolution of polypyrimidines in *Drosophila*. *Genetics* **92**, 459–484.

DONEHOWER, L., FURLONG, C., GILLESPIE, D. & KURNIT, D. (1980). DNA sequence of baboon highly repeated DNA: evidence for evolution by nonrandom unequal crossovers. *Proceedings of the National Academy of Sciences (USA)* **77**, 2129–2133.

DOOLITTLE, W. F. & SAPIENZA, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603.

DOVER, G. (1978). DNA conservation and speciation: adaptive or accidental? *Nature* **272**, 123–124.

DOVER, G. A. (1980). Problems in the use of DNA for the study of species relationships and the evolutionary significance of genomic differences. In *Chemosystematics: principles and practice* (ed. F. A. Bisby, J. G. Vaughn & C. A. Wright) (In the Press). London, New York, San Francisco: Academic Press.

DOVER, G. A. & COEN, E. (1981). Springcleaning ribosomal DNA: a model for multigene evolution? *Nature* **290**, 731–732.

DOVER, G. & DOOLITTLE, W. F. (1980). Modes of Genome evolution. *Nature* **288**, 646–647.

FLAVELL, R. B., O'DELL, M. & HUTCHINSON, J. (1981). Nucleotide sequence organization and evidence for sequence transloçation during evolution. *Cold Spring Harbor Symposium* **45** (In the Press).

FRY, K. & SALSER, W. (1977). Nucleotide sequences of HS-α satellite DNA from Kangaroo Rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* **12**, 1069–1084.

GALL, J. G. & ATHERTON, D. D. (1974). Satellite DNA sequences in *Drosophila virilis*. *Journal of Molecular Biology* **85**, 633–664.

GOSDEN, J. R., MITCHELL, A. R., BUCKLAND, R. A., CLAYTON, R. P. & EVANS, H. J. (1975). The location of four human satellite DNAs on human chromosomes. *Experimental Cell Research* **92**, 148–158.

GOSDEN, J. R., MITCHELL, A. R., SEUANEZ, H. N. & GOSDEN, C. M. (1977). The distribution of sequences complementary to human satellite DNAs I, II and IV in the chromosomes of chimpanzee (*Pan troglodytes*), Gorilla (*gorilla gorilla*) and Orang Utan (*Pongo pygnaeus*). *Chromosoma (Berlin)* **63**, 253–271.

GOSDEN, J. R., LAWRIE, S. S. & COOKE, H. J. (1981). A cloned repeated DNA sequence in human chromosome heteromorphisms. *Cytogenetics and Cellular Genetics* (In the Press).

HOLMQUIST, G. P. & DANCIS, B. (1979). Telomere replication, kinetochore organizers and satellite DNA evolution. *Proceedings of the National Academy of Sciences (USA)* **76**, 4566–4570.

HORZ, W. & ALTENBURGER, W. (1981). Nucleotide sequence of mouse satellite DNA. *Nucleic Acids Research* **9**, 683–696.

HSIEH, T-S. & BRUTLAG, D. (1979*a*). Sequence and sequence variation within the 1·688 g/cm^3 satellite DNA of *Drosophila melanogaster*. *Journal of Molecular Biology* **135**, 465–481.

HSIEH, T-S. & BRUTLAG, D. (1979*b*). A protein that preferentially binds *Drosophila* satellite DNA. *Proceedings of the National Academy of Sciences (USA)* **76**, 726–730.

HUTCHINSON, J., REES, H. & SEAL, A. G. (1979). An assay of the activity of supplementary DNA in *Lolium*. *Heredity* **43**, 411–421.

IGO-KEMENES, T., OMORI, A. & ZACHAU, H. G. (1980). Non-random arrangement of nucleosomes in satellite I containing chromatin of rat liver. *Nucleic Acids Research* **8**, 5377–5390.

John, B. (1981). Heterochromatin variation in natural populations. *Proceedings of the 7th International Chromosome Conference, Oxford (In the Press)*.

John, B. & Miklos, G. L. G. (1979). Functional aspects of satellite DNA and heterochromatin. In *International Review of Cytology* (ed. G. H. Bourne and J. F. Danielli), pp. 1–114. New York, San Francisco, London: Academic Press.

Kao-Huang, Y., Revzin, A., Butler, A. P., O'Conner, P., Noble, D. W. & Von Hippel, P. H. (1977). Nonspecific DNA binding of genome-regulating proteins as a biological control mechanism: Measurement of DNA-bound *Escherichia coli lac* repressor *in vivo*. *Proceedings of the National Academy of Sciences (USA)* **74**, 4228–4232.

Kimura, M. (1979). The neutral theory of molecular evolution. *Scientific American* **241**, 94–104.

Klein, H. L. & Petes, T. D. (1981). Intrachromosomal gene conversion in yeast. *Nature* **289**, 144–148.

Klug, A., Jack, A., Viswamitra, M. A., Kennard, O., Shakked, Z. & Steitz, T. A. (1979). A hypothesis on a specific sequence – dependent conformation of DNA and its relation to the binding of the *lac*-repressor protein. *Journal of Molecular Biology* **131**, 669–680.

Kurnit, D. M. (1979). Satellite DNA and heterochromatin variants: the case for unequal mitotic crossing over. *Human Genetics* **47**, 169–186.

Lin, S. & Riggs, A. D. (1975). The general affinity of *lac* repressor for *E. coli* DNA: Implications for gene regulation in procaryotes and eukaryotes. *Cell* **4**, 107–111.

Macgregor, H. C. (1980). Recent developments in the study of lampbrush chromosomes. *Heredity* **44**, 3–35.

McKay, R. D. G., Bobrow, M. & Cooke, H. J. (1978). The identification of a repeated DNA sequence involved in the karyotype polymorphism of the human Y chromosome. *Cytogenetics and Cell Genetics* **21**, 19–32.

Maio, J. J., Brown, F. L. & Musich, P. R. (1977). Subunit structure of chromatin and the organization of eukaryotic highly repetitive DNA: Recurrent periodicities and models for the evolutionary origins of repetitive DNA. *Journal of Molecular Biology* **117**, 637–655.

Manuelidis, L. & Wu, J. C. (1978). Homology between human and simian repeated DNA. *Nature* **276**, 92–94.

Matsumoto, Y., Yasuda, H., Mita, S., Marunouchi, T. & Yamada, M. (1980). Evidence for the involvement of H1 histone phosphorylation in chromosome condensation. *Nature* **284**, 181–183.

Maynard-Smith, J. (1977). Why the genome does not congeal. *Nature* **268**, 693–696.

Miklos, G. L. G. & John, B. (1979). Heterochromatin and satellite DNA in man: properties and prospects. *American Journal of Human Genetics* **31**, 264–280.

Miklos, G. L. G., Willcocks, D. A. & Baverstock, P. R. (1980). Restriction endonuclease and molecular analyses of three rat genomes with special reference to chromosome rearrangement and speciation problems. *Chromosoma (Berlin)* **76**, 339–363.

Miklos, G. L. G. & Gill, A. C. (1981). The DNA sequences of cloned complex satellite DNAs from Hawaiian *Drosophila* and their bearing on satellite DNA sequence conservation. *Chromosoma* **82**, 409–427.

Miklos, G. L. G. (1981). Sequencing and manipulating highly repeated DNA. Systematics Association Special Volume Series. Academic Press (In the Press).

Mitchell, A. R., Beauchamp, R. S. & Bostock, C. J. (1979). A study of sequence homologies in four satellite DNAs of man. *Journal of Molecular Biology* **135**, 127–149.

Mullins, J. I. & Blumenfeld, M. (1979). Satellite 1 c: a possible link between the satellite DNAs of *D. virilis* and *D. melanogaster*. *Cell* **17**, 615–621.

Musich, P. R., Brown, F. L. & Maio, J. J. (1977). Subunit structure of chromatin and the organisation of eukaryotic highly repetitive DNA: Nucleosomal proteins associated with a highly repetitive mammalian DNA. *Proceedings of the National Academy of Sciences (USA)* **74**, 3297–3301.

Musich, P. R., Brown, F. L. & Maio, J. J. (1980). Highly repetitive component α and related Alphoid DNAs in man and monkeys. *Chromosoma* **80**, 331–348.

Nagl, W. (1978). Endopolyploidy and polyteny in differentiation and evolution. Amsterdam: North-Holland.

Nature. News and Views. (*a*) How selfish is DNA (1980) **285**, 617–620. (*b*) Selfish DNA (1980) **288**, 645–648.

Olmo, E. & Morescalchi, A. (1978). Genome and cell sizes in frogs: a comparison with salamanders. *Experientia* **34**, 44–46.

Orgel, L. E. & Crick, F. H. C. (1980). Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.

Orgel, L. E., Crick, F. H. C. & Sapienza, C. (1980). Selfish DNA. *Nature* **288**, 645–646.

Pech, M., Igo-Kemenes, T. & Zachau, H. G. (1979*a*). Nucleotide sequence of a highly repetitive component of rat DNA. *Nucleic Acids Research* **7**, 417–432.

Pech, M., Streeck, R. E. & Zachau, H. G. (1979*b*). Patchwork structure of a bovine satellite DNA. *Cell* **18**, 883–893.

Petes, T. D. (1980). Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. *Cell* **19**, 765–774.

Poschl, E. & Streeck, R. E. (1980). Prototype sequence of bovine 1.720 satellite DNA. *Journal of Molecular Biology* **143**, 147–153.

Rees, H. & Dale, P. J. (1974). Chiasmata and variability in *Lolium* and Festuca populations. *Chromosoma* (*Berlin*) **47**, 335–351.

Roizes, G., Pages, M. & Lecou, C. (1980). The organization of the long range periodicity calf satellite DNA I variants as revealed by restriction enzyme analysis. *Nucleic Acids Research* **8**, 3779–3792.

Rosenberg, H., Singer, M. & Rosenberg, M. (1978). Highly reiterated sequences of Simian simian simian simian simian. *Science* **200**, 394–402.

Rubin, C. M., Deininger, P. L., Houck, C. M. & Schmid, C. W. (1980). A dimer satellite sequence in bonnet monkey DNA consists of distinct monomer subunits. *Journal of Molecular Biology* **136**, 151–167.

Salser, W., Bowen, S., Browne, D., El Adli, F., Federoff, N., Fry, K., Heindell, H., Paddock, G., Poon, R., Wallace, B. & Whitcome, P. (1976). Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Federation Proceedings* **35**, 23–35.

Scherer, S. & Davis, R. W. (1980). Recombination of Dispersed repeated DNA sequences in yeast. *Science* **209**, 1380–1384.

Selsing, E. & Arnott, S. (1976). Conformations of A,T-rich DNAs. *Nucleic Acids Research* **3**, 2443–2450.

Schmookler-Reis, R. J. & Biro, P. A. (1978). Sequence and evolution of mouse satellite DNA. *Journal of Molecular Biology* **121**, 357–374.

Skinner, D. M. (1977). Satellite DNA's. *Bio Science* **27**, 790–796.

Southern, E. M. (1970). Base sequence and evolution of guinea-pig α satellite DNA. *Nature* **227**, 794–798.

Southern, E. M. (1975). Long range periodicities in mouse satellite DNA. *Journal of Molecular Biology* **94**, 51–69.

Stanley, S. M. (1975). A theory of evolution above the species level. *Proceedings of the National Academy of Sciences* (*USA*). **72**, 646–650.

Szostak, J. W. & Wu, R. (1980). Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* **284**, 426–430.

Thayer, R. E., Singer, M. F. & McCutchan, T. F. (1981). Sequence relationships between single repeat units of highly reiterated African green monkey DNA. *Nucleic Acids Research* **9**, 169–181.

Timberlake, W. E. (1978). Low repetitive DNA content in *Aspergillus nidulans*. *Science* **202**, 973–975.

Varley, M. J., Macgregor, H. C. & Erba, H. P. (1980). Satellite DNA is transcribed on lampbrush chromosomes. *Nature* **283**, 686–688.

Walker, P. M. B. (1979). Genes and non-coding DNA sequences. In *Human Genetics: possibilities and realities. Exerpta Medica*, pp. 25–45. Amsterdam.

Will, H. & Bautz, E. K. F. (1980). Immunological identification of a chromocenter – associated protein in polytene chromosomes of *Drosophila*. *Experimental Cell Research* **125**, 401–410.

Wheeler, L. L., Arrighi, F., Cordeiro-Stone, M. & Lee, C. S. (1978). Localization of *Drosophila nasutoides* satellite DNAs in metaphase chromosomes *Chromosoma* (*Berlin*) **70**, 41–50.

Wu, J. C. & Manuelidis, L. (1980). Sequence definition and organisation of a human repeated DNA. *Journal of Molecular Biology* **142**, 363–386.

Yamamoto, M. & Miklos, G. L. G. (1977). Genetic dissection of heterochromatin in *Drosophila*: The role of basal *X* heterochromatin in meiotic sex chromosome behaviour. *Chromosoma* (*Berlin*) **60**, 283–296.

Yamamoto, M. & Miklos, G. L. G. (1978). Genetic studies on heterochromatin in *Drosophila melanogaster* and their implications for the functions of satellite DNA. *Chromosoma* (*Berlin*) **66**, 71–98.

Yamamoto, M. (1979*a*). Cytological studies of heterochromatin function in the *Drosophila melanogaster* male: autosomal meiotic pairing. *Chromosoma* (*Berlin*) **72**, 293–328.

Yamamoto, M. (1979*b*). Interchromosomal effects of heterochromatic deletions on recombination in *Drosophila melanogaster*. *Genetics* **93**, 437–448.