

The Evolution of Satellite DNAs in the genus Apodemus

H. J. Cooke



Thesis submitted for the degree of  
Doctor of Philosophy  
University of Edinburgh 1974



### Acknowledgements

I should like to thank the following people for help and encouragement in many forms. Their support has made the task of working towards, and writing, this thesis much less difficult than it would otherwise have been.

|                |               |
|----------------|---------------|
| J. Allan       | J. McKay      |
| A. Biro        | R.D.G. McKay  |
| A.J. Buchanan  | K. Murray     |
| R.P. Clayton   | M. Philp      |
| J. Cooke       | B. Rigal      |
| A.H. Denniss   | G. Roizes     |
| D. Fletcher    | P.G. Sealey   |
| D. Gould       | J.M.W. Slack  |
| K.M. Gummerson | A.M. Southern |
| P. Hamilton    | P.M.B. Walker |
| C. Herzberg    | M. White      |
| P.F.R. Little  |               |
| D. Long        |               |

In addition I am grateful to British Olivetti Ltd., for the loan of an electric typewriter. I am required to acknowledge the financial support of the Medical Research Council in the form of a training award.



## ABSTRACT

DNAs from mice in the genus Apodemus show bands when digests of total DNA with a variety of restriction enzymes are analysed by agarose gel electrophoresis. From the observed patterns a phylogenetic tree can be drawn.

In two species, A. sylvaticus and A. flavicollis, the bands with Hind III are integral multiples of a size of 370 base pairs and derive from satellite DNAs. Divergence in the satellite DNA sequence can be estimated from the proportions of the monomer, dimer and trimer and from the thermal stability of the DNAs after reassociation. There is a marked difference between the melting behaviour of A. sylvaticus and A. flavicollis DNAs and it is suggested that this is due to the presence of a number of related families of sequences in A. sylvaticus satellite DNA but only one family in A. flavicollis. The observation that these different families of repeated sequences have the same periodicity is consistent with the view that unequal crossing over is responsible for their evolution into separate families.



## INDEX

### a) Text

|           |   |    |
|-----------|---|----|
| Chapter 1 | INTRODUCTION  |    |
|           | Repeated sequence DNA   | 2  |
|           | Reassociation kinetics  |    |
|           | Rate  | 4  |
|           | Complexity  | 4  |
|           | Mismatch  | 5  |
|           | Temperature of reassociation  | 6  |
|           | Isolated satellites   |    |
|           | Sequence analysis   | 8  |
|           | Evolution of satellite  | 10 |
|           | Function of repeated sequences  | 12 |
|           | Localisation of repeated sequences  | 15 |
|           | Restriction endonucleases   | 17 |
| Chapter 2 | MATERIALS & METHODS   |    |
|           | DNA preparation   | 20 |
|           | Density gradient centrifugation   | 20 |
|           | Melting and reassociation   | 21 |
|           | Gel electrophoresis   | 22 |
|           | Recovery of DNA from gels   | 23 |
|           | Preparation of restriction enzymes  | 24 |
|           | Iodination of DNA   | 25 |
| Chapter 3 | RESULTS & DISCUSSION  |    |
|           | Gel electrophoresis   | 26 |
|           | Isopycnic centrifugation  | 36 |
|           | Banding patterns of <u>Apodemus</u> DNAs  | 39 |
|           | Restriction of <u>A. sylvaticus</u> satellite   | 47 |
|           | Estimation of divergence at the<br>restriction sites                                  | 47 |
|           | Formation of $\frac{1}{2}$ mers of the repeat   | 51 |
|           | Thermal denaturation of <u>A. sylvaticus</u><br>and <u>A. flavicollis</u> band I DNAs | 54 |



|  |        |
|--|--------|
| Reassociation of restriction band DNA  | 56     |
| Thermal stability of reassociated<br>band DNA from <u>A. sylvaticus</u> and <u>flavicollis</u>       | 59     |
| The significance of differences between<br>measured divergence of <u>A. sylvaticus</u><br>band I DNA | 61     |
| Cross reassociation between families<br>in <u>A. sylvaticus</u> satellite DNA                        | 65     |
| Melts of individual DNA  | 67     |
| Reassociation register of restriction<br>fragments   | 70     |
| Divergence between total DNAs  | 72     |
| Derivation of a phylogenetic tree  | 75     |
| <br>Chapter 4 CONCLUSION   | <br>82 |
| <br>References   | <br>86 |



b) Plates

|         |   |    |
|---------|---|----|
| Plate 1 | Restriction pattern of<br>$\lambda$ .C DNA with Hin | 27 |
| Plate 2 | Restriction patterns of<br><u>Apodemus</u> DNAs     | 41 |

c) Tables

|         |   |    |
|---------|---|----|
| Table 1 | Properties of <u>A. sylvaticus</u><br>and <u>A. flavicollis</u> satellite<br>DNAs | 37 |
| Table 2 | Sizes of restriction bands<br>in <u>Apodemus</u> DNAs                             | 42 |
| Table 3 | Properties of <u>A. sylvaticus</u><br>and <u>A. flavicollis</u> band I<br>DNAs    | 60 |

d) Figures

|        |  |    |
|--------|--|----|
| Fig 1  | Tracing of $\lambda$ .C restriction<br>pattern                                       | 28 |
| Fig 2  | Mobility vs. molecular<br>weight   | 29 |
| Fig 3  | Calibration plot vs. SV40  | 31 |
| Fig 4  | Partial digest of mouse<br>satellite with Eco RII                                    | 32 |
| Fig 5  | Preparative gels   | 34 |
| Fig 6  | Effect of perchlorate on<br>DNA recovered from gels                                  | 35 |
| Fig 7  | Analytical CsCl traces of<br><u>Apodemus</u> DNAs                                    | 38 |
| Fig 8  | CsSO <sub>4</sub> centrifugation of<br><u>Apodemus</u> DNAs                          | 40 |
| Fig 9  | Traces of restricted<br><u>A. sylvaticus</u> and <u>A. flavicollis</u><br>total DNAs | 43 |
| Fig 10 | Partial and complete digests<br>of <u>A. microps</u> DNA                             | 45 |



|        |   |    |
|--------|---|----|
| Fig 11 | Divergence in a tandemly repeating sequence                           | 46 |
| Fig 12 | Digestion of satellite DNA  | 48 |
| Fig 13 | Measurement of divergence in the restriction site                     | 50 |
| Fig 14 | Unequal crossing over   | 53 |
| Fig 15 | Thermal denaturation of band I DNAs native and reassociated           | 55 |
| Fig 16 | Reassociation of band I DNAs  | 57 |
| Fig 17 | Two populations of duplexes or two regions                            | 62 |
| Fig 18 | Snap back of duplexes   | 64 |
| Fig 19 | Cross reassociating families  | 66 |
| Fig 20 | Multiple related families   | 68 |
| Fig 21 | Thermal denaturation of DNA from individual mice                      | 69 |
| Fig 22 | Size of reassociated band I duplexes                                  | 71 |
| Fig 23 | Size distribution of band I duplexes                                  | 73 |
| Fig 24 | Cross reassociation of <u>A. sylvaticus</u> and <u>A. flavicollis</u> | 76 |
| Fig 25 | Cross reassociation of <u>A. sylvaticus</u> and <u>A. microps</u>     | 77 |
| Fig 26 | Stages in the evolution of satellite in the genus <u>Apodemus</u>     | 78 |
| Fig 27 | Phylogenetic tree based on restriction patterns                       | 81 |



## INTRODUCTION

The genome of E. coli consists of  $4 \times 10^{-3}$  pg of DNA, sufficient to code for about 3000 average size proteins (Massie & Zimm, 1965). Haploid DNA contents of higher organisms range between  $22 \times 10^{-3}$  and 84 pg (Mirsky & Ris, 1951) which, if assessed on the basis of E. coli, would be capable of coding for between  $1.6 \times 10^4$  and  $6.3 \times 10^7$  different proteins. There is no direct evidence on the total numbers of proteins which can be present in either E. coli or a higher organism so the possible interpretation that higher organisms use all their DNA for protein synthesis (or at least the same proportion as E. coli) cannot be simply excluded.

A number of pointers suggest that this is not the case: certain necturids, which are evolutionarily closely related, have c values which vary over a ten fold range whilst the animals themselves remain very similar. From observations on the rate of mutations observed in protein sequences Kimura (1968) has pointed out that there would be an intolerable genetic load on the average mammal unless most mutations were selectively neutral. King & Jukes (1969) argued that only a small proportion of mutations were neutral and therefore the proportion of the DNA which could carry detectable mutations was around 1% in mammals.

There are many arguments along these lines but the two basic hypotheses remain: either higher organisms use all their DNA in the same way as E. coli (although this is not well understood) or only a relatively small fraction of the DNA is used in this fashion.

If the latter supposition is correct then there is DNA in the higher organism genome which is not part of an operon organisation. Compared to a prokaryotic DNA, higher organism DNAs are heterogeneous in physical and chemical properties such as buoyant density and rate of reassociation. This does suggest that there are different classes of DNA in eukaryotes which differ from each other in properties, and hence presumably in function. But if these DNAs have



functions different from the DNA in E. coli what functions do they have?

Our ideas of DNA function are in fact severely limited by the model system presented by E. coli; to escape from this situation and to become able to postulate and confirm the functions of DNA in a higher organism requires that more should be known about the DNA of these organisms. It is only recently that sufficient data has accumulated to begin the process of showing eukaryotic DNA to be qualitatively different from prokaryotic DNA.

Since the higher organism genome is so large, studies of the DNA as a whole have underlined chiefly the heterogeneity mentioned above. The heterogeneity shown by re-association kinetics has been used to provide an arbitrary division of eukaryotic DNAs into fast, intermediate and slow classes (Britten & Kohne, 1967).

#### REPEATED SEQUENCE DNA.

Repeated sequence DNA is that fraction of the genome which is composed of sequences present in more than a few copies. It can also be defined as the DNA which reassociates at a rate faster than would be expected if the whole genome was composed of unique sequences. Satellite DNAs are a subset of this general group of DNAs which as well as being repeated, differ sufficiently in base composition from the bulk of the DNA to be separable from it by CsCl density gradient centrifugation (Walker, 1971). This definition of satellite DNA had to be expanded since the finding that sequences in the DNA complex with different quantities of heavy metal ions such as  $Ag^+$  or  $Hg^{2+}$  and may then give a uniform band with a density difference from the bulk DNA.

Certain classes of DNA with known function are repeated to a greater or lesser extent in all organisms studied. For example DNA coding for ribosomal RNA has repetition frequencies, measured by RNA excess hybridisation, between ten in E. coli (Zehavi et al, 1966) and 400-800 in Xenopus (Birnstiel et al, 1968; Brown & Weber, 1968). The repetition



of sequences which are transcribed but not translated probably reflects the need of cells for large quantities of transcription products for protein synthesis where protein synthesis itself can-not be used to increase the levels of the components in the cell. Multiple copies of histone genes (Pardue et al 1972) and polyteny are also probably means of stepping up RNA synthesis.

In general the repetition of these sequences is low (with some notable exceptions, e.g. ribosomal cistrons in Xenopus oocytes (Dawid et al, 1970)) and they represent only a small fraction of the DNA. In higher organisms, repetitious DNA, defined simply by its reassociation rate, is present in almost all examples studied, in many cases as a large proportion of the genome (see Bostock, 1972, table 1). The reassociation rate of some of these repeated DNAs is very high, meaning that the sequence complexity is low. This in turn means that they are more easily studied than other classes of DNA. For this reason, and because they are the most obvious exception to the E. coli paradigm, they have been the most extensively studied fraction of eukaryotic DNAs.

In writing an introduction to a thesis concerned mainly with the properties of satellite DNA the immediate difficulty is that there is no known function for this class of DNA. There is a mass of experimental work on structure, evolution and distribution of satellite DNA which, although relevant to the problem of function, is complex and difficult to interpret because of the incomplete understanding of the precise physical chemistry of the methods used. These uncertainties may significantly affect the conclusions drawn from the results and hence their functional significance. ✓

I have reviewed the field of repeated DNA with an emphasis on satellite DNA in order to relate the work in the thesis to what is known already and also to bring out the functional significance, if any, of a variety of approaches.



## REASSOCIATION KINETICS

### a) Rate

The most useful and widely applied method for the study of repeated sequences in higher organism DNA is that of DNA/DNA reassociation in solution. The rate of reassociation can be followed because of differences in physical properties of single and double strand DNA such as ultra-violet absorption, optical rotation and ability to bind to hydroxylapatite. When the rate of reassociation of DNA from a number of DNA from a number of prokaryotes and viruses was measured it was found that there was an inverse linear relationship between the size of the genome and the rate of reassociation (Britten & Kohne 1967) and that a single rate constant could describe the curve.

Britten & Kohne (1967) expressed their data as extent of reaction v.s. a product of the initial single strand concentration and time ( $C_0 t$ ). For a simple second order reaction the rate constant is given by the following equation;

$$C_0 t_{1/2} = 1/k_2$$

When eukaryotic DNAs were examined in this way it was found that, in calf for example, there was a spread of reassociation rates with two major components - one with a  $Cot_{1/2}$  of 0.03 and one with a  $Cot_{1/2}$  of  $3 \times 10^3$ . This type of reassociation behaviour has been found in large numbers of eukaryotes and has led to the arbitrary division into kinetic classes mentioned above:- rapidly reassociating or repetitive DNA is that which reassociates between a  $Cot$  of  $10^{-4}$  and 1, intermediate DNA reassociates between  $Cot$  1 and 100 and the unique or slow fraction which reassociates with a  $Cot$  above 100 (Walker, 1971).

### b) Complexity

Wetmur & Davidson (1968) have derived an equation relating reassociation rate to the sequence complexity ( $n$ ) of DNA of a given length ( $l$ ) which agrees with a model of the reaction mechanism in which nucleation of complementary



sequences is the rate limiting step followed by a rapid zipping of the nucleated strands.

$$k_2 = 3.5 \times 10^5 l^{0.5}/n$$

This expression, with the Britten & Kohne relationship between complexity and genome size, has been used in a number of cases to estimate the sequence complexity and multiplicity of a kinetic fraction based on its rate of reassociation and the percentage in the genome. Thus mouse satellite DNA with a kinetic complexity of 300 base pairs and comprising 10% of the mouse genome must be present in  $10^6$  copies per cell (Waring & Britten, 1966). Spectrograms of frequency of repetition of nucleotide sequences in DNA of calf and mouse show similar patterns, in both the bulk of the DNA reassociates as single copy DNA but a high proportion reassociates at the rate expected of DNA sequences present in  $10^4$ - $10^5$  copies. In neither organism is there detectable DNA with a repetition frequency of 2 -10.

### c) Mismatch

The relationship between sequence complexity and rate of reassociation was derived from viral and bacterial DNAs which form perfect duplexes when reassociated. When rapidly reassociating nuclear DNA from eukaryotes is melted after reassociation, the temperature at which half of the duplex has melted is much lower than the native  $T_m$  of the DNA because of mismatching between diverged copies. Estimates of the effect of mismatch on the reduction of thermal stability vary from 0.7 to 1.5% mismatch per  $1^\circ\text{C}$  difference between the  $T_m$ s of native and reassociated duplex (Laird, McConnaughty & McCarthy, 1970; Wang & Kallenbach, 1971; McCarthy & Farquhar, 1972)

It has been suggested that mismatching in the reassociated duplex may cause a reduction in the rate of reassociation and hence an overestimate of the complexity of repeated sequences.



There is much disagreement about the magnitude of the effect of mismatch; this is a point of importance since a large effect would dramatically reduce the estimated complexity of sequences such as mouse satellite which show differences between the native and reassociated duplex melting temperatures of five degrees.

Southern (1971) suggests that mismatched bases have a large effect on the rate of reassociation if a duplex region in the process of zippering meets a mismatched base pair and is thus unable to reach a stable length. This process would reduce the apparent concentration of nucleation sites, and hence the observed rate of reaction, by factors around ten fold for the degree of mismatch corresponding to a  $T_m$  of 10 degrees and stable lengths of 10-20 base pairs. The correction thus decreases the sequence complexity of the repeated sequence and increases the number of copies present.

Data confirming this proposal has been published by Sutton & McCallum (1971) who examined the reassociation rate of mouse satellite DNA fractionated according to extent of divergence. Other authors (McCarthy & Farquhar, 1972; Bonner et al, 1973) find a much smaller reduction in rate of two to three fold for a  $10^{\circ}\text{C}$   $T_m$ . This effect of mismatch on rate of reassociation may be most significant when the reassociating molecules are randomly sheared repeated sequences which can reassociate to form many different mismatched duplexes. McCarthy & Farquhar (1972) used modified bacterial DNAs which, although forming mismatched duplexes, had only one possible reassociation register. Southern (1974) has confirmed the presence of the 120 base pair repeat predicted by Sutton & McCallum (1971) from the effect of mismatch on rate in the mouse satellite system.

#### d) Temperature of reassociation

The maximum rate of reassociation is reached at a temperature  $25^{\circ}\text{C}$  below the  $T_m$  of the native DNA. This is true for most DNAs and the effect is such that over a range of temperature between  $16$  and  $32^{\circ}\text{C}$  below the  $T_m$  there is



little difference in the rate of reassociation. When the temperature is outside these limits then the rate is slowed. Above this range the minimum stable length for nucleation is increased and hence the rate of reassociation and the extent of mismatch in the duplex are decreased. Below this range the rate of reassociation is reduced because of kinetic effects of the lower temperature, the duplex formed will be less well matched, because of the reduction of the minimum stable length, and hence the kinetics will reflect repeats not detectable at higher temperatures. Sutton & McCallum (1972) have detected a diverged repeat length of between 8 and 20 base pairs from measurements of the rate of cross reaction of Mus satellite DNAs at low temperatures.

Reassociation kinetics and thermal stability studies are equally applicable to total DNA and to isolated satellite DNA; other than the kinetic complexity they give little information on the structure of the repeated sequences. the advantage of these methods is that they can be used where the repeated sequences are not separable in native form from the bulk of the DNA and to give an overall view of the distribution of DNA into different kinetic classes (Laird, 1971; Davidson & Hough, 1968).

Perhaps the most interesting use of reassociation is in the comparison of sequences present in related species. Rice (1972) has compared sequences reassociating by a  $Cot$  of about fifty in a number of species by allowing a trace of one labelled DNA to reassociate with driver quantities of another DNA labelled with a different isotope. The stabilities of the hetero- and homoduplexes were then measured by thermal elution from hydroxylapatite. From a number of experiments of this type Rice (1972) has concluded that the number and relatedness of repeated sequences common to a pair of rodent species declines as their evolutionary separation increases.

The conditions of reassociation used in this type of experiment contribute significantly to the amount of



DNA found to be related between two species. Increased salt concentration or reduced temperature makes badly matched duplexes stable and hence detectable by hydroxyl-apatite chromatography (Sutton & McCallum, 1972). In functional terms the striking feature of this type of approach is the variation found between species for satellite and repeated DNA content in very similar species (Hennig & Walker, 1970).

#### ISOLATED SATELLITES

When a satellite DNA is isolable a number of other approaches become possible, due largely to the presence of only one or a few families of repeated sequences in the separated satellite. A number of satellite DNAs can be strand separated since differences in G+T content between the strands results in their having different buoyant densities in alkaline CsCl (see Walker, 1971). If strand separation of a group of related satellites is possible then the reassociation behaviour and thermal stabilities of the heteroduplexes can be determined in a well defined manner (Blumenfeld, 1973).

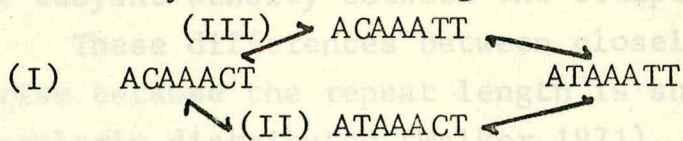
#### a) Sequencing

More importantly the ability to separate the strands of a homogeneous DNA fraction doubles the amount of information available from pyrimident tract analysis (Burton & Peterson, 1960). As Southern (1973) has pointed out, when a repeated sequence contains only one pyrimidine tract on each strand the complete sequence can be derived from the pyrimidine tracts of the two separated strands. Using this approach Southern (1970) was able to derive a basic repeating hexamer for the guinea pig satellite from which most of the other pyrimidine tracts found on the finger prints could be derived by single transitions or transversions.

Gall & Atherton (1974) have sequenced each strand of the three satellites in D. virilis which make up 45% of the genome. All of the sequences are related; those of



satellites II and III can be derived from that of satellite I by one base change.



In all three cases at least 90% of the material was found to be in sequences corresponding to the repeating heptamer. This corresponds well with the finding that the  $\Delta T_m$  of these satellites is about  $1^\circ\text{C}$  (Blumenfeld, 1973) and suggests that they have been formed recently.

Mouse satellite DNA has not been sequenced fully because the sequence is more complex than that of the guinea pig satellites (Biro, Carr-Brown, Southern & Walker, 1974) but the basic sequence is probably about 12 base pairs with longer overlying repeats.

The known sequences for satellite DNAs all suggest short repeats which contrast with the longer repeat lengths suggested by kinetic data. This could be due to the effect of mismatch on rate of reassociation mentioned above or to the reassociation register consisting of only the longer repeats perhaps because of extensive base changes which have destroyed the ability of the shorter repeats to reassociate (Britten, 1971).

The properties of the *D. virilis* satellites illustrate the dangers involved in interpretation of thermal stability data. Blumenfeld (1973) has shown that isolated strands of satellites I and II, and I and III, form hybrid duplexes which have  $\Delta T_m$ s of approximately  $17^\circ$  and  $30^\circ\text{C}$  (IH + IIL and IH + IIIL,  $\Delta T_m$   $17^\circ$ ; IL + IIH and IL + IIIH,  $\Delta T_m$   $30^\circ$ ). The sequences of I and II, and I and III are 85% homologous, according to Gall and Atherton (1974), those of II and III are 71% homologous and do not form stable duplexes under the conditions used. These figures are remarkable since the sequences are closely related and also because of the differences between reciprocal hybrids due to the differences



Direct evidence on the evolution of repeated sequences is available only for satellite DNA because of the relative ease with which they can be sequenced; repeated but non satellite DNA is tacitly assumed to have evolved by the same mechanisms. The occurrence of base changes in the sequence of guinea pig ~~a~~ satellite differs considerably from random; Southern (1970) suggests that this is incompatible with a one step saltation event unless there is selection for some mutations but not for others. To explain this distribution of mutation without invoking selection he suggests that if an initial multiplication event was followed by random divergence a subsequent multiplication of randomly chosen sequences might result in a nonrandom occurrence of base changes in the product sequences.

The same scheme can be applied to the D. virilis satellites. To derive the sequences from each other would also require a mechanism in which at some time one mutated sequence from satellite I must have been multiplied to be present in many copies to form satellite II; satellite III must also have been generated in this way. This type of scheme has also been suggested by Southern (1972) for the guinea pig satellites.

Clearly the multiplication process must require a special mechanism, but the nature of this is at present unknown. Localised duplications of genetic material are known to occur in plants and amphibians (Rothfels et al 1966; Ullerich 1966) but the extent of these does not suggest a parallel with the massive multiplications found for satellites. The production of large numbers of copies of extra-chromosomal DNA during oocyte maturation in Xenopus (Brown & Dawid, 1968) provides a quantitatively closer analogy. The rolling circle model for the production of viral DNAs (Gilbert & Dresler, 1968) has been used as a model for replication of repeated DNA in an indefinite number of copies from a circular template. Bird et al (1973) and Hourcade et al (1973a,b) have observed structures compatible with this in the amplifying DNA in Xenopus oocytes.



This is interesting in view of the genetic evidence from prokaryotes reviewed by Doerman (1973) suggesting strongly that  $T_4$  replication does not involve a rolling circle mechanism.

#### Function of repeated sequence DNAs.

The lack of a known function for DNA fractions which represent a large portion of the genome in many organisms is a severe drawback to understanding the organisation of the higher organism genome. Many suggestions have been made to try to fill this gap. Hypotheses about the function of repetitive DNA usually consider either non-satellite repetitive DNA and satellite DNA separately; repetitive DNA is found in all higher organisms in varying quantities whereas satellites vary greatly in their occurrence even between closely related species. There is also more information available about satellite DNAs than about repetitive DNAs which do not form satellites.

Satellite DNA sequences are not found to be transcribed into RNA (Flamm et al, 1969) although they should be simple to detect because of their multiplicity. In situ hybridisation confirms the findings of Maio & Schildkraut (1969) and Yasmineh & Yunis (1970) that satellite DNA is associated with the heterochromatin and is preferentially localised near the centromeres of chromosomes. Because of this generally better understanding of satellite, and in particular because of the observations of centromeric/heterochromatic localisation and lack of transcription, a number of possible functions can be ruled out for satellite DNA which must be considered for repeated DNA.

Bostock (1972) has reviewed possible functions for repetitive DNA comprehensively. They can be divided into two classes: control functions which are distributed over many sequences coding for proteins and RNA, and the second class - functions in chromosomes affecting either structure or behaviour. Satellite DNAs seem to be so repetitive that there are insufficient differences between sequences to allow



differential recognition, by proteins or RNA, of individual repeats. The evidence that they are present in a few large blocks and that they are associated with heterochromatin would also seem to exclude possible functions in control of replication or transcription. Repeated sequence DNA, however, seems to have the interspersion characteristics compatible with the idea that a region of repetitive DNA adjacent to a structural gene might be responsible for control of its expression and that the observed repetition frequencies might reflect the number of control elements for co-ordinate expression of overlapping groups of genes (Britten & Davidson 1969; Georgiev 1969).

Functions suggested for satellite DNA generally relate to some aspect of chromosome behaviour. Walker (1971<sup>b</sup>) has suggested that satellite is involved in control of the various folding patterns which chromosomes undergo. Sutton (1972<sup>b</sup>) has suggested that once basic proteins have neutralised the negative charges on the DNA backbone similar sequences will tend to pack together in a crystalline manner and that these crystalline regions become heterochromatic.

Since the satellite sequences seem to be confined to the heterochromatin adjacent to the centromeres it is difficult to see how they could affect the structure of the chromosomes outside these areas. The presence of satellite DNA sequences in other regions of the chromosome can not be ruled out by the in situ hybridisation results and other repeated sequences do not seem to show centromeric localisation (Hearst et al 1974).

At the same time it should be recognised that centromeric functions are probably extremely important to the cell since these are the sites of attachment of spindle proteins, are involved in pairing of sister chromosomes at meiosis and are the sites at which sister chromatids are held together during meiosis. Satellite DNA has been suggested to be involved in promoting association between chromosomes in Drosophila (Walker, 1971<sup>b</sup>), X chromosomes with large deletions in heterochromatin do not pair with the Y chromosome resulting in frequent loss of the chromosomes;



attached XY chromosomes which have extra heterochromatin at either side of the centromere are lost less often (Sandler & Braver, 1954).

Repetitive DNA could be involved in pairing of chromosomes generally. Pairing of homologous chromosomes occurs in both meiosis and mitosis; the mechanism is unknown but is generally agreed to result in precise interchange of genetic information during recombination. The most obvious way of achieving this is by pairing at the nucleotide level and there has been much speculation on this basis (Holliday, 1964; 1968; Sobell 1973). In vitro, satellite and repeated DNAs have properties which suggest that they could be involved in this pairing at the nucleotide level; they reassociate rapidly and in a number of different registers. Their centromeric localisation could reflect the occurrence of processes analogous to recombination; somatic genetic exchange (Walen, 1964), sister chromatid exchange (Marin & Prescott, 1964) and Robertsonian fusion all occur at or near the centromeres. Arguing that repeated sequences are involved in processes of chromosome or chromatid association, crossing over could be promoted by regions of repeated DNA, perhaps repetition frequency is related to the frequency with which chromosomal interactions resulting in rearrangements of DNA molecules occur.

This speculation is symptomatic of the basic problem facing eukaryotic molecular biology at the present; we have insufficient knowledge of the functions which might be necessary to higher organisms to be able to see the role of the different classes of DNA and their organisation. Even knowing the large amount that we do about satellite DNA we have no firm idea of a class of functions which might be filled by this DNA. Ideas such as 'chromosomal house keeping' or packing of DNA into chromosomes are attempts to fill this conceptual gap but they must be refined to become well defined hypotheses if they are to be useable as a guide for further research.



Perhaps the most significant difference (in terms of function) between satellite DNA and the repeated sequences which do not form satellites, cryptic or otherwise, lies in their arrangement with other classes of DNA. Satellite DNA can be isolated as high molecular weight molecules; Gall & Atherton (1974) examined their preparations of satellites from D. virilis in the electron microscope and found molecules up to  $6 \times 10^7$  daltons in DNA which was composed of 95% of the repeated septamer. This implies that up to  $10^4$  copies of the sequence are tandemly linked. Southern (1974) has found that mouse satellite is composed of tandemly arranged 240 base pair repeats which can be detected linked in multiples of up to about 25 copies.

The in situ hybridisation technique of Pardue & Gall (1970) and Jones (1970) supports the finding that mouse satellite is present in the genome as a few short blocks, in this case located near the centromeres of the mouse chromosomes. Similar results have been obtained with a number of Drosophila satellites (Hennig, 1972) but one satellite DNA from D. pseudoneohydei showed no localisation (Hennig, 1970). The limitations of the in situ hybridisation method mean that only high concentrations of a sequence can be detected and so it remains possible that sequences of satellite or related sequences are found elsewhere on the chromosomes even where they have so far been found only at the centromeres.

Repetitive non-satellite DNA shows little if any localisation by the in situ method (Arrighi et al, 1970; Botchan et al, 1971). Since isolation by reassociation methods involves shear treatment of DNA to avoid network formation and since material isolated in this way must almost always be a heterogeneous mixture of families it has not been possible to measure the length of the repeated units or their homogeneity in the way used for satellites. Davidson et al (1973) have studied the interspersion of repeated DNA by using trace amounts of DNA of different lengths and reassociating it to a Cot of about 50. From the percentage



of the DNA which becomes partially double stranded under these conditions, and from the behaviour of these molecules when thermally denatured or treated with a single-strand specific nuclease they concluded that the molecules have a double stranded region of about 300 base pairs which is repetitive; interspersed with this there are longer, non-repeated regions. They also concluded that more than 80% of the DNA fragments contained repeated sequences at a size of 4500 nucleotides. Wu et al (1972) found that DNA with a Cot between 0.1 and 10 from Drosophila has a family size of between 40 and 100 with about 3000 families present; they concluded from electron microscope studies that the repeated DNA is 150 base pairs long, separated from each other by non-repetitive regions 750 base pairs long. In contrast to these results Davidson (1974) has found a much longer range interspersion.

Thomas and his co-workers have shown in a number of papers (1973a, 1973b) that fragments of eukaryotic DNA can form rings after random shearing followed by exonuclease treatment (resection) and reannealing. They found that Drosophila polytene DNA (in which repetitive DNA is under replicated) formed about 15% thermally stable rings, Necturus DNA formed 35% rings and salmon sperm DNA 20%. This was interpreted as evidence for a high degree of tandem repetition but this conclusion contradicts the available evidence from kinetic studies which shows that most higher-organism DNAs have a significant proportion reassociating as single copies. A more serious criticism of their work is that when the circles are isolated after enzymatic closure they have buoyant densities, reassociation kinetics and chromosomal locations identical to those of satellite DNAs (Schachat & Hogness, 1973; Peacock et al, 1973).



RESTRICTION ENDONUCLEASES

The discovery of bacterial restriction enzymes (Smith & Wilcox, 1970; Yoshimuri, 1971; Sharp et al, 1973) has made possible a number of new approaches to problems of DNA organisation and function. These enzymes, which recognise specific short symmetrical sequences in the DNA and make a double strand break in the sequence, have been used to study a number of viral DNAs (Danna & Nathans, 1973; Lee & Sinsheimer, 1974 ), to provide a physical map of the genome which is being used to map different viral functions to different DNA fragments. Fragments produced by these enzymes have been used for fine structural studies of the lambda operator (Maurer, Maniatis & Ptashne 1974; Maniatis & Ptashne, 1973) and the sites of initiation of RNA synthesis on lambda and T<sub>7</sub> DNAs.

In larger genomes where the distribution of restriction sites is near random these enzymes will produce a spread of fragments which will have an average size related to the length of the specific sequence recognised by the enzyme, e.g. an enzyme such as Hae III which recognises a tetramer (K. Murray, personal communication) will make a break on average once every 256 base pairs in DNA with a random sequence.

Since repeated DNA has a non-random sequence there are two distinct classes of products possible from restriction endonuclease digestions depending on whether or not the restriction sequence is present in the repeat. When the sequence is not present the repeated sequences will be high molecular weight after restriction, the probability of the repeated DNA containing the restriction sequence being inversely proportional to the repeat length; this type of behaviour has been observed by Botchan & McKenna (1973) and by Southern & Roizes (1973).

When the restriction sequence is part of a repeated sequence then restriction enzymes will produce fragments of a uniform length from the repeated sequence and the



length of these fragments will be the length of the repeat unit. In a eukaryotic DNA digested with a restriction enzyme which breaks a repeated sequence there will be a large number of fragments of this length superimposed on a length distribution composed of larger than average fragments of repeated DNA which does not contain the restriction site and of DNA with a random distribution of restriction sites. Those fragments of uniform length produced from the repeated sequence will be found as a peak in the size distribution.

When restriction fragments from total eukaryotic DNAs are analysed by gel electrophoresis many then show the bands expected of repeated sequences. The observation of bands in a digest of a total DNA analysed in this way provides a rapid way of detecting repetitive DNA; however not all repeated sequences need contain a restriction site and one enzyme will only show those repeated sequences which contain that restriction site. As a means of detection the method is limited but as a means of estimating the repeat length of a fraction defined in some other way it is extremely useful. Restriction enzymes make possible the isolation in native form of single families of repeated sequences which do not form satellites and so should make evolutionary, structural and functional studies possible in the same way as for satellites.

In the case of the genus Apodemus, I have used restriction endonucleases to show that a common 370 base pair repeat length occurs in three of five species and that in two species the satellite DNA is composed entirely of DNA with this repeat length. Over relatively long sequences of molecular weight greater than  $2.6 \times 10^6$  daltons the divergence is distributed randomly. This confirms the picture already established for satellite DNA of a tandemly repeating collection of sequences which are related to, and probably derived from each other.

Additionally the kinetics of reassociation and extent of mismatch in the duplexes formed by isolated repeat units suggest that multiplication events take place at a number



of stages in the evolution of the satellite. The effect of mismatch on rate in this system suggests that there is no great effect of mismatch on the rate of reassociation when the ends of the sequence are defined.

Although the overall rate of accumulation of mutations is the same as the mutation rate at individual sites in the sequence (the restriction sites), the repeat length is maintained over a considerable span of evolutionary time.



### Materials

All chemicals used were analytical reagent grade. Agarose was electrophoresis grade from Sigma Chemical Co., London; hydroxyapatite was from Bio Rad Laboratories California, USA, HTP grade. Sources of all other chemicals and materials, other than these are given in the text.

### Mice

Apodemus sylvaticus from Midlothian (Scotland), A. flavicollis, from Yugoslavia, A. mystacinus and A. microps from Czechoslovakia were maintained as laboratory colonies.

### Bacterial strains

Haemophilus influenzae was a non encapsulated strain Rd 123 from S. Glover. H. aegyptius was a gift from D. Gould; H. parainfluenzae was a gift from R. Old.

### Methods

#### DNA preparation

DNA was prepared from liver, kidney, testis, and spleen by extraction with cold phenol and sodium dodecyl sulphate as described by Walker & McLaren (1965). DNA was routinely further purified by iso-propanol precipitation; 0.58 volumes of iso-propanol were added dropwise to one volume of DNA solution made 0.3M in sodium acetate. The DNA was spooled onto a glass rod, washed in 70% ethanol, air dried and redissolved in distilled water.

Bacteriophage DNA was a gift from K. Murray.

#### Preparative density gradient centrifugation.

Preformed CsCl density gradients (Brunk & Leich, 1969) were used to isolate satellite DNA. A solution of CsCl at a density of 1.62 gm/ml containing 400 to 500 microgrammes of DNA was layered onto an equal volume of CsCl solution at a density of 1.80 gm/ml. The tube was filled with paraffin oil and centrifuged at 26,000 rpm for 70 hours at 25°C in a MSE 8x40ml angle rotor (Flamm et al 1966).



Fractions of 0.2 ml were collected by pumping from a perspex float lowered onto the surface of the solution. The u.v. absorbance of the fractions was measured on a Beckman DB-G spectrophotometer at 260nm. Satellite fractions were pooled then diluted to a density of 1.62 gm/ml and layered onto CsCl at a density of 1.8 as above. Three cycles of centrifugation were found to be sufficient for purification of A. sylvaticus and A. flavicollis satellite DNAs.

#### Analytical density gradient centrifugation.

1-2 microgrammes of the DNA to be analysed was added to 0.5 ml of CsCl solution (Merk Ultrapur) with a density of 1.69 to 1.70 gm/ml. In some cases M. lysodeicticus DNA at a concentration of 0.2 to 0.5 microgrammes/ml was added as a density marker. ✓

Cells were centrifuged at 45,000 rpm for at least 17 hours in either a MSE analytical ultracentrifuge or a MSE centriscan 75. In the former case the cells were photographed onto Ilford N4E. 50 line film and scanned in a Joyce-Lobel recording microdensitometer. In the centriscan cells were scanned directly. ✓

#### Recovery of DNA from CsCl gradients.

High molecular weight DNA from gradients was diluted with distilled water and centrifuged at 40,000 rpm overnight in a MSE 10 x 10 ml titanium rotor. The supernatant was discarded and the pellet washed in 70% ethanol to remove traces of CsCl. The dehydrated pellet was dissolved in distilled water.

#### Melting and reassociation experiments.

DNA for melting and reassociation was dialysed overnight against 0.01M NaCl, 0.001M Tris/HCl 0.0001 M EDTA pH 8.0 then degassed in a vacuum desiccator and centrifuged briefly to remove dust. A Zeiss PMQ II spectrophotometer with a thermostated cell holder, automatic slit control and sample changer equipped with digital print out of ✓



temperature and optical density was used in all experiments.

Melting was carried out with the temperature increasing at  $0.5^{\circ}\text{C}$  per minute, absorbance readings at 260 nm were corrected for thermal expansion. After melting the solutions were cooled in the spectrophotometer to 10 to  $15^{\circ}$  below the temperature at which they were allowed to reassociate. The cell holder was then equilibrated to the reassociation temperature and the solution made 0.04 M NaCl with the appropriate volume of 5M NaCl, mixed rapidly and the reassociation followed. Cot values in this buffer were corrected to 0.18 M  $\text{Na}^{+}$  using the tables of Britten & Smith (1970).

### Gel Electrophoresis

Agarose slab gels 19 x 19 x 0.3 cms were cast between glass plates separated by perspex spacers 0.3 cms thick. During pouring the bottom of the gel compartment was formed by a perspex slot former. A variety of sizes were used, formers with 40 slots 0.1 x 0.2 x 1.0 cms were used to make gels for assay of restriction activity in column effluents; formers with 10 slots 0.1 x 1.0 x 1.0 cms and with 15 slots 0.1 x 1.0 x 0.7 cms were used for analytical gels. Gels with three slots 0.1 x 1.0 x 6.0 cms were used for preparation of restricted material.

Agarose solutions of 0.6 to 2% were made by boiling the required quantity of agarose in E buffer (Loening 1967) until the agar was dissolved. The solution was cooled with stirring to  $60^{\circ}\text{C}$ , ethidium bromide (Sigma Chemical Co.) added to 0.5 microgramme/ml and poured into the gel former. When the gel had set the slot former was removed.

Gels were run in E buffer with 0.5 g/ml ethidium bromide at room temperature. The bottom of the gel was placed in a small tank containing 500ml of buffer and one electrode, the other buffer compartment was connected to the top of the gel by a wick made from four thicknesses of Whatman 3MM filter paper. Buffer was circulated between the compartments and was also dripped slowly onto the wick. ✓



Samples containing between 1  $\mu$ g and 1mg of DNA, depending on the size of the slot, were loaded in 5% glycerol and run into the gel at 50 volts for 20 minutes. Gels were run at about 200 volts, 60 to 80 m.a. for four to six hours, the plates were separated and the gel examined under short wave ultra-violet light.

Gels were photographed under short wave U.V. through a red filter (Hoya 25A) onto Ilford FP4 sheet film which was developed in Kodak D76 developer. The negatives were scanned with a Joyce-Lobel recording microdensitometer; areas under peaks were measured with a planimeter after a background level was measured between the sample slot and an adjacent empty slot.

#### Recovery of DNA from gels

Gel containing DNA was cut from the gel slab then passed through a 19G syringe needle and weighed, an equal volume of 5M NaClO<sub>4</sub> to the weight of the gel was added and the mixture heated to 60° and held at that temperature during subsequent steps in the procedure. When the agar was completely dissolved, 0.5 to 1ml of Bio Rad HAP was added and mixed into the solution at 60°. After a few minutes the HAP was centrifuged down and washed several times with 0.05 M PB. Over 95% of the DNA was recovered in this way, estimated by using <sup>32</sup>P labelled DNA.

#### Growth of bacteria

All strains of Haemophilus were grown in brain heart infusion (Difco) supplemented with NAD at 2 g/ml. H. influenzae and H. aegyptius also required hemin at 10 g/ml for growth. Cells were grown in a Churchill laboratory fermenter and harvested in an Alpha Laval solids retaining separator when in end log phase. Yields of bacteria were around 3gm wet weight of cells per litre of culture.



### Preparation of restriction enzymes

Restriction enzymes Eco RI, Eco RII, Hpa I and II were gifts from K. Murray. Hin restriction enzyme and its two component enzymes Hind II and Hind III were prepared according to Smith (personal communication).

Restriction enzymes from H. aegyptius (Hae I, II, and III) were prepared by a modification of a procedure of R. Roberts (personal communication). Ten grammes of cells were ground to a sticky paste with 10gm of alumina (Alcoa Chemicals), this was then suspended in 10ml of 0.01 M Tris/HCl 0.01 M mercaptoethanol, spun at 5000 rpm in a Sorval SS-33 rotor at 0° for 15 minutes; the viscous DNA pellet was removed with the alumina and the viscous supernatant sonicated for 4 x 15 seconds in a salt ice bath (this step reduces the viscosity to improve running on the columns).

The lysate was made 1.0 M NaCl and loaded onto a Biogel A 0.5m 200 mesh column 50 x 2.5 cms equilibrated with 1.0M NaCl, 0.01 M Tris/HCl, 0.01 M mercaptoethanol pH 7.9. The exonuclease peak was pooled and solid  $(\text{NH}_4)_2\text{SO}_4$  added slowly to 0.5 gm/ml with 0.005 ml 1 N NaOH per ml to maintain the pH of the solution. The precipitate was dissolved in the smallest possible volume of 0.01 M  $\text{KPO}_4$ ,  $10^{-4}$  M EDTA, 0.01 M mercaptoethanol, 10% glycerol and dialysed against the same buffer. This was applied to a 25 x 0.9 cm column of Whatman P 11 cellulose which had been precycled twice and equilibrated with the same buffer. The column was eluted with a linear gradient of KCl from 0 to 0.75 M KCl, 100 ml total volume.

The three activities were well separated; Hae I and Hae II were dialysed against the P 11 buffer and applied to Whatman DE 52 columns equilibrated with the same buffer. Elution was with 0 to 0.3 M KCl. This step removes exonuclease but at the expense of a considerable loss of activity.

All restriction enzymes were made 50% in glycerol and stored either at 0 or -20°. under these conditions Hind III was stable for at least six months as was Hae III, Hind II was never as stable as Hind III made in the same preparation



and was even less stable when purified from Hind III.

All enzymes were assayed viscometrically according to Smith & Wilcox (1970) or by analysing the digestion of phage DNA by gel electrophoresis.

#### Iodination of DNA

Iodination of DNA was carried out according to Commerford (1971); for some DNAs HAP was used to remove unbound iodine, rather than exclusion chromatography. Only preparations of  $^{125}\text{I}$  DNA in which 90% or more of the counts were acid precipitable and DNase sensitive were used.

#### Electron microscopy

Electron microscopy of DNA samples was carried out according to Southern (1974).



## RESULTS AND DISCUSSION

### Gel Electrophoresis

Agarose gel electrophoresis is used through this thesis as a method of analysis and also as a preparative method for material of a given size class. These gels are used to estimate the size of a DNA molecule and to estimate relative proportions of DNA in different size classes by quantifying the fluorescence of ethidium bromide bound to DNA in the gel. Plate 1 is a photograph of a complete digest of two microgrammes of  $\lambda$ .C (wild type) DNA produced by *Hin* restriction endonuclease and analysed on a 2% agarose gel as described in the materials and methods section. Fig. 1 is a microdensitometer trace of the same negative.

The areas under the individual peaks in this tracing were measured with a planimeter and the log of the percentage of the total area in each peak plotted against the distance run. A logarithmic relationship between mobility and molecular weight should, when plotted in this way, be linear since stoichiometric yields of each fragment should mean that the relative fluorescence in any peak is proportional to the size of the DNA in the peak. This also depends on the ethidium bromide fluorescence being directly proportional to the amount of DNA in the peak. Fig 2 (a plot of the areas of the peaks in Fig. 1, plotted in this way) shows that this is the case and that the mobility-molecular weight relationship is approximately log-linear over almost all of this range of molecular weights. This agrees well with the results of Sharp et al (1973). There is an uneven background level of fluorescence in the negative due to a number of factors. These include: a) the presence of DNA which is either incompletely digested or from a contamination of the  $\lambda$ .C DNA with other DNA (probably *E. coli*), b) uneven illumination of the gel by the ultra-violet source. This uneven background makes



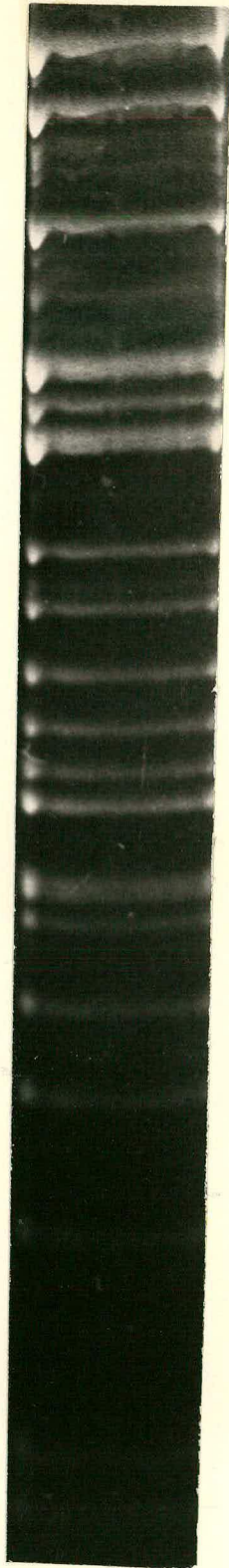


Plate 1 Gel electrophoresis of  $\lambda$ .C restricted with Hin restriction endonuclease.



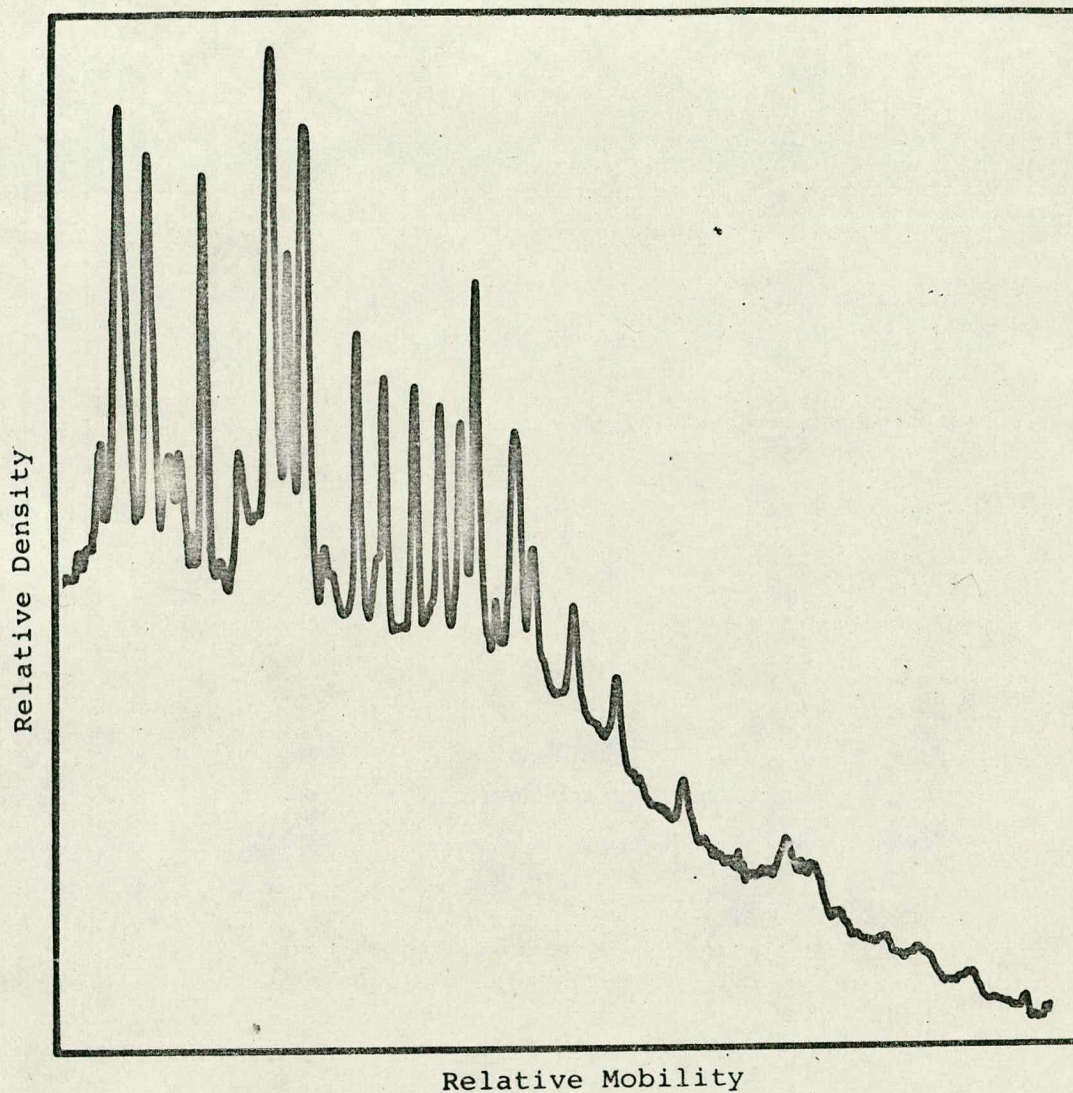


Fig. 1 Microdensitometer trace corresponding to plate 1. 2 microgrammes of  $\lambda$ .C DNA digested with 0.2 units of Hin restriction endonuclease for 18 hours at 37°. Electrophoresis and photography as described in text.



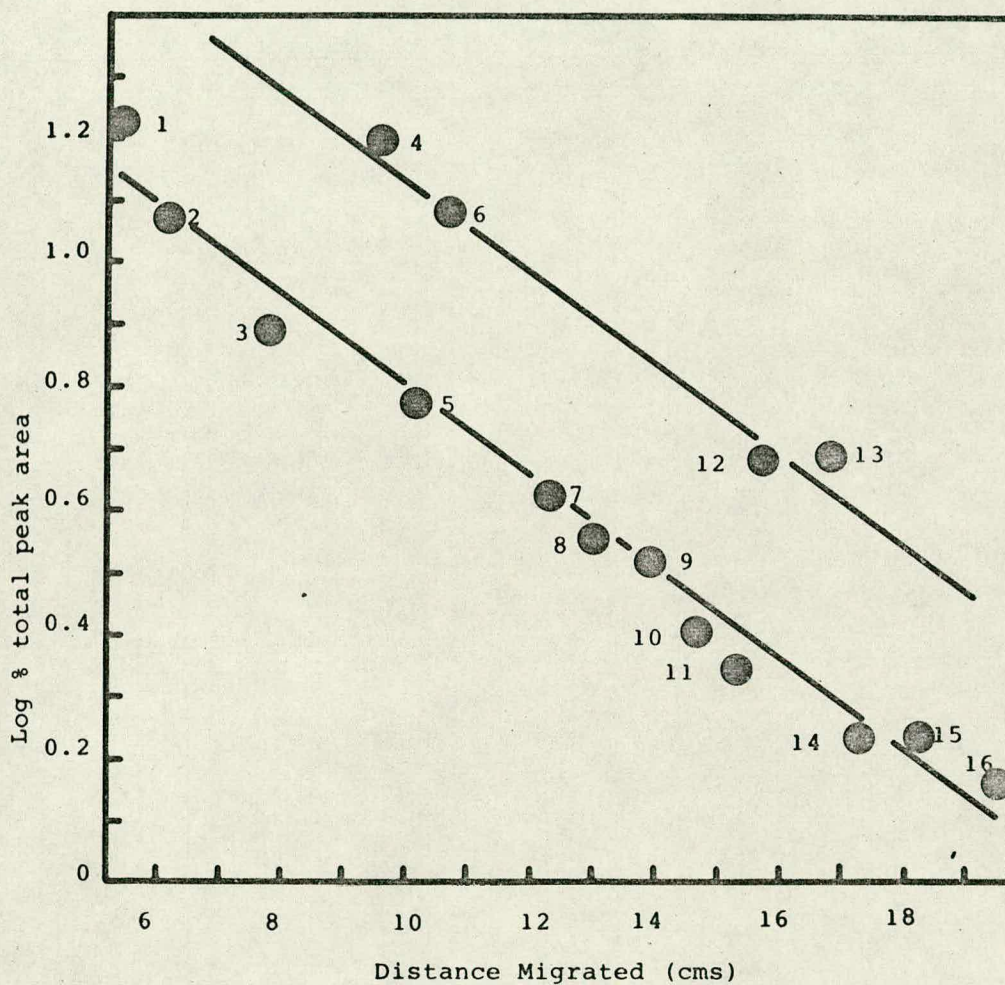


Fig. 2 Log % total area of peaks in fig. 1 plotted against distance measured on microdensitometer scans. Peaks 4, 6, 12 and 13 fall on a line parallel to and log 2 above the line of the majority of points due to the presence of two different fragments.



measurement of small peak areas difficult. The smallest visible band in the tracing represents approximately 0.5% of the input DNA or 0.01 microgramme. In less concentrated gels the background due to light scattering is lower and longer exposures, which in this case would have lead to over-exposure of the intense bands, makes it possible to see even smaller quantities of DNA. Smith (personal communication) has estimated the lower limit of sensitivity of this method to be around 2 picogrammes of DNA.

Gel electrophoresis thus provides a rapid, cheap, accurate, and highly sensitive method of analysis of DNA of a wide range of sizes. To eliminate the necessity of close control over the running conditions in the electrophoresis it is convenient to use marker DNAs. The size of the fragments from  $\lambda$ .C DNA was estimated by electrophoresis against fragments of SV 40 DNA of known molecular weight produced by digestion with Hind III (Danna & Nathans, 1973) Fig.3. The fragments were used as markers for early experiments. Mouse satellite DNA partially restricted with Eco RII gives a series of bands which are integral multiple lengths of 240 base pairs (Southern, 1974) Fig.4. The mouse fragments were used in later experiments. The size estimates for the Apodemus bands based on these two independent markers, both of which are based on electron microscopy, were 370 base pairs and 340 base pairs. Both these values are within 10% of each other but the larger value was taken, as this was the marker used for most of the work. Southern (1974) has estimated the error in his measurement of the size of the mouse satellite band to be 10%; this is the largest error in the estimation of the Apodemus bands and so the sizes determined using this marker are subject to a + or - 10% error.

Agarose gels do not seem to be very sensitive to the quantity of DNA loaded; when used preparatively up to two mg. of DNA could be separated with a slot area of 1.8 cms.<sup>2</sup>



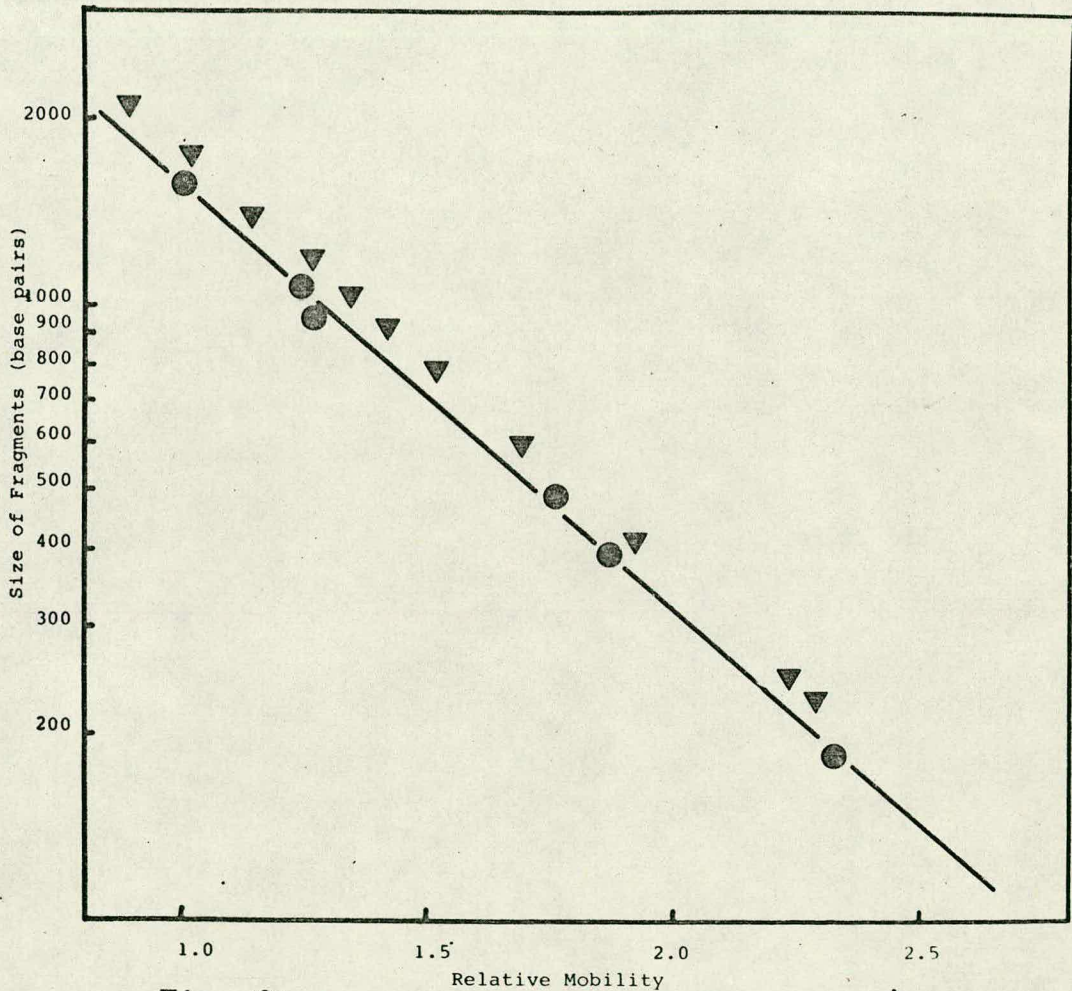


Fig. 3 Estimation of the size of DNA fragments from  $\lambda$  against SV40 fragments of known size. ● SV40 fragments  
 ▼  $\lambda$  fragments



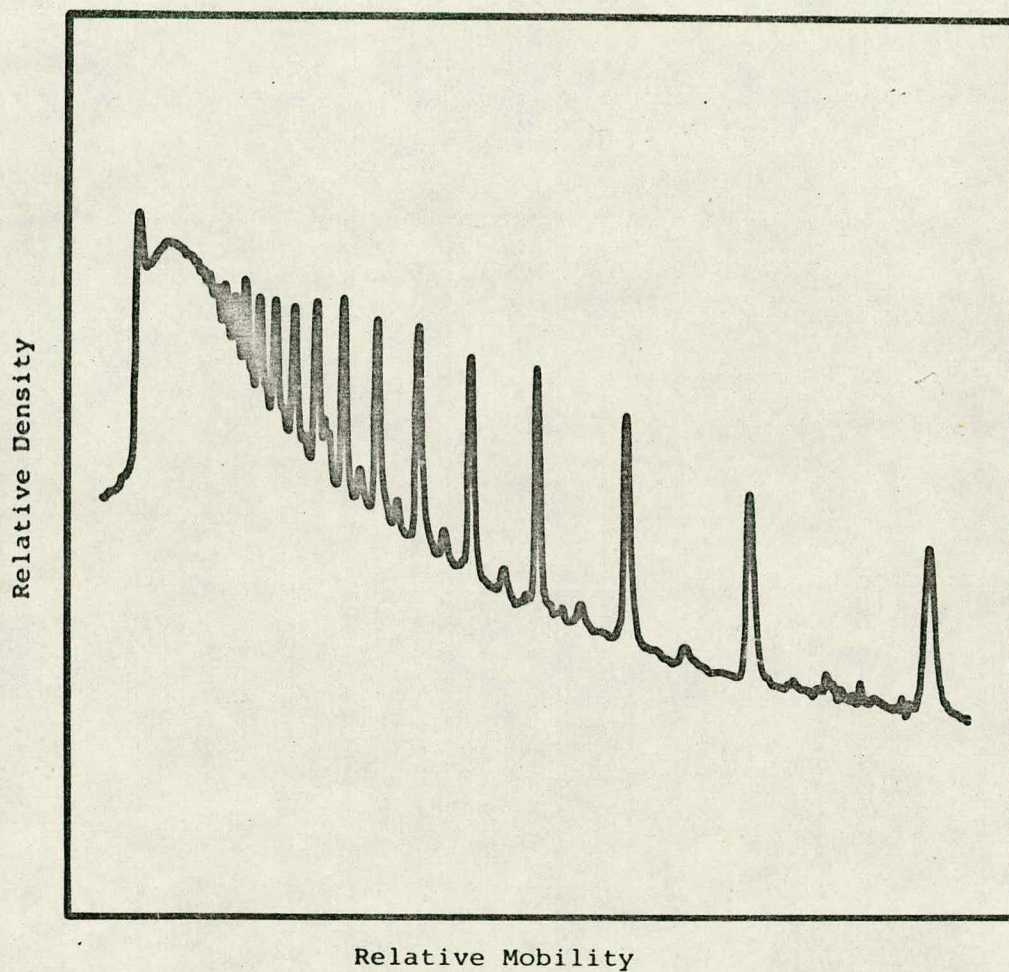


Fig. 4 Mouse satellite DNA digested with Eco RII, 1 microgramme of DNA digested with 0.02 units of Eco RII for 2 hours.



Figure 5a is an example of such a slab gel on which two mg. of Apodemus sylvaticus DNA, in the form of restriction fragments produced by Hin restriction endonuclease, was separated. High concentrations of DNA in the gel are capable of binding more ethidium bromide than is present in the gel under the conditions normally used for electrophoresis. This results in the uneven fluorescence seen in this gel. The effect can be overcome if necessary by staining the gel, after electrophoresis, for six to eight hours in E buffer containing 0.5 microgrammes per ml. of ethidium bromide.

The photograph in Fig. 5a demonstrates that even at this high loading of DNA the separation and resolution of the gel is good. When the smallest strong band is cut from the gel, cast into another gel and then re-run the band becomes slightly broader but there is little DNA moving very much faster or slower than the peak (fig. 5b). If overloading had caused retention of smaller molecules by larger molecules the effective dilution of the sample by rerunning the material would reduce the effect and the gel would show a trail of smaller DNA.

Recovery of DNA from a gel slice by the perchlorate/hydroxylapatite method could possibly lead to partial denaturation; 2.5 M  $\text{NaClO}_4$  would be expected to lower the melting temperature of a DNA by  $20^\circ$  (Wetmur & Davidson, 1968). To check if the recovery procedure was having an effect on the physical properties of the DNA, E. coli DNA sheared to a length of 500-1000 base pairs was mixed with perchlorate and agar at  $60^\circ\text{C}$  in the same quantities used for recovery from gels. A sample of the same DNA was heated to  $60^\circ$  in 0.05 M PB and both DNA samples were recovered by the addition of HAP. When thermally denatured the  $T_m$  of both samples of DNA, measured optically, was  $70.5^\circ\text{C}$  showing that the DNA was unaffected and that all the perchlorate was removed. There was a small reduction in total hyperchromicity for the perchlorate treated DNA, about 2%, but this effect was insignificant (fig. 6).



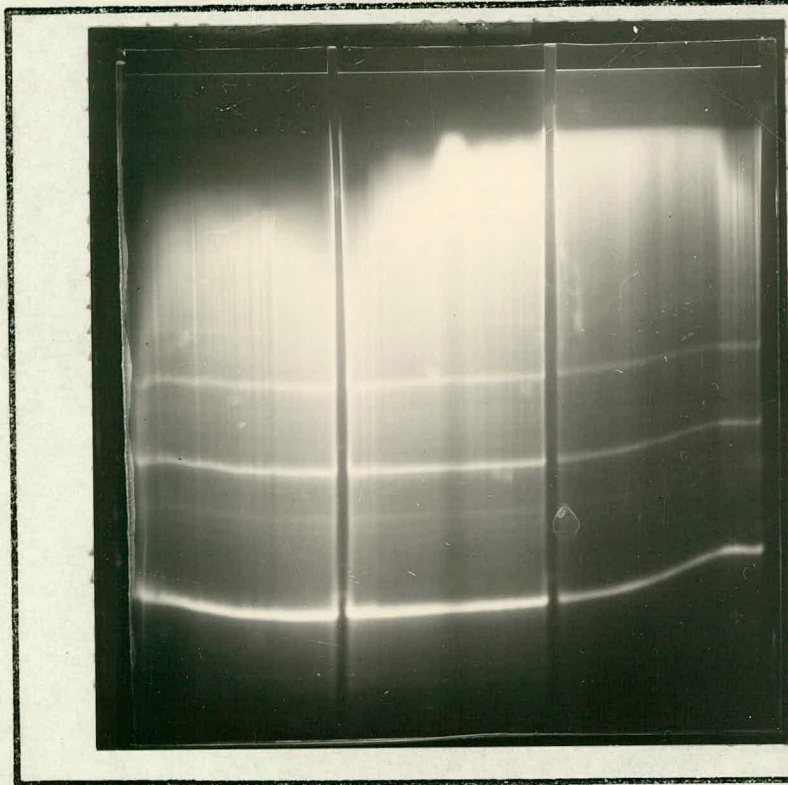


Fig. 5a Photograph of preparative gel loaded with 2mg. of A. sylvaticus restriction fragments after electrophoresis

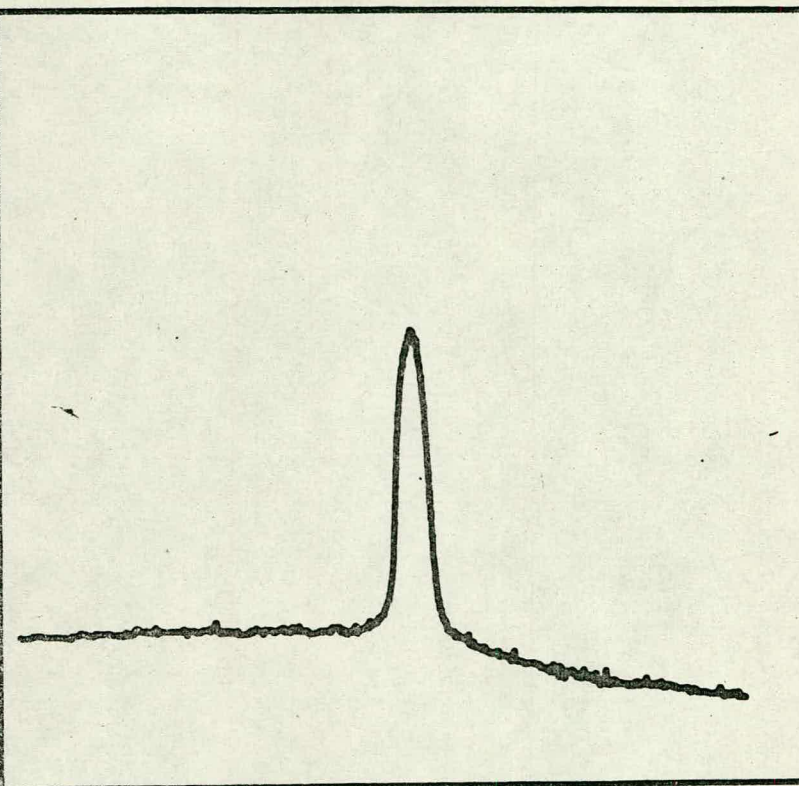


Fig. 5b Major band cut from preparative gel and re-run (tracing of photograph)



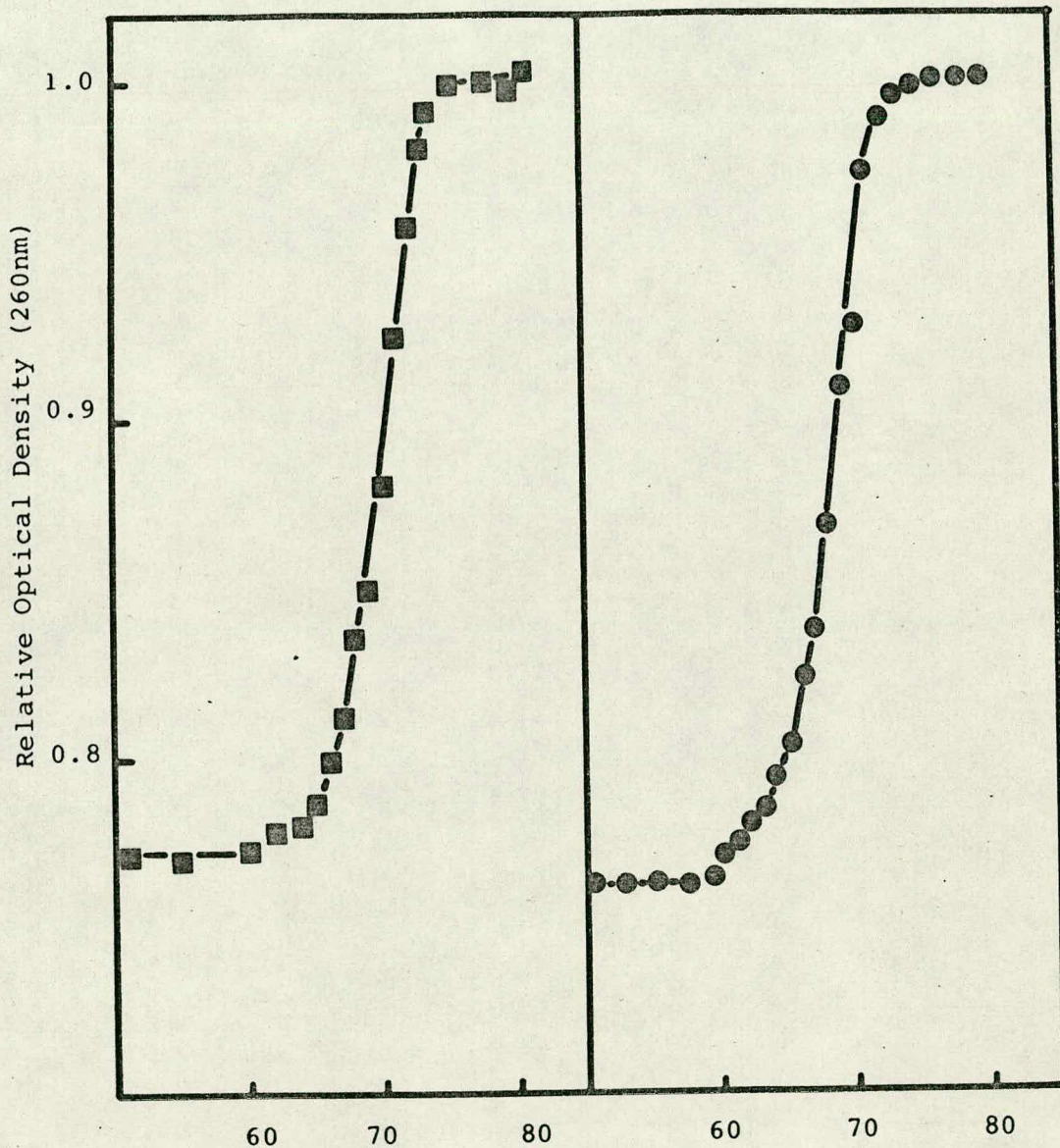


Fig. 6 Effect of perchlorate on thermal denaturation of *E. coli* DNA. —■— recovered from agarose by HAP and sodium perchlorate, —●— eluted from HAP without sodium perchlorate.



The yield of DNA from the gels by this method was better than all other methods tested provided that a relatively large volume of HAP was used to bind the DNA, when 1ml packed volume of HAP per 0.1 mg of DNA was used recovery of DNA was around 90%. In contrast elution from the gel by shaking with buffer or extraction with phenol gave recoveries of 30-50%. The large volume of HAP is probably needed because the surface of the HAP crystals become coated with agar and are then unable to bind DNA.

### b) Isopycnic Centrifugation

#### CsCl Analytical Centrifugation

Hennig & Walker (1970) have described the analysis of A. sylvaticus, A. flavicollis, A. mystacinus, and A. agrarius DNAs in CsCl density gradients. A. sylvaticus and A. flavicollis DNAs both contain light satellite DNAs which amount to 8% and 6% of the total DNAs respectively (table 1). A. agrarius also has a well defined satellite DNA which differs from those from A. sylvaticus and A. flavicollis in buoyant density, reassociation rate, thermal stability of the reassociated duplex, and sequence complexity as estimated by partial sequence analysis (Allan 1974). A. microps and mystacinus show slight skewing of the single peaks to the heavy side of the gradient but do not have any distinguishable satellite peaks (figure 7). On the basis of similarities between the distribution of DNA of each species in density gradients, A. sylvaticus, A. flavicollis and A. agrarius would be grouped together as being more similar to each other than to A. microps and A. mystacinus which would also form a related group.

#### Ag<sup>+</sup>/CsSO<sub>4</sub> centrifugation

Allan (1974) has reported that analytical Ag<sup>+</sup>/CsSO<sub>4</sub> gradient centrifugation gives separation of total A. sylvaticus DNA into two or more peaks, if these peaks rep-



Table 1 Properties of A. sylvaticus and A. flavicollis satellite DNAs

|                       | <u>A. sylvaticus</u> | <u>A. flavicollis</u> |
|-----------------------|----------------------|-----------------------|
| Relative quantity     | 8%                   | 6%                    |
| Native density        | 1.695 g/ml           | 1.693 g/ml            |
| Native T <sub>m</sub> | 67.5°C               | 67.5°C                |
| ΔT <sub>m</sub>       | 11.0°C               | 6.0°C                 |

From Allan (1974)

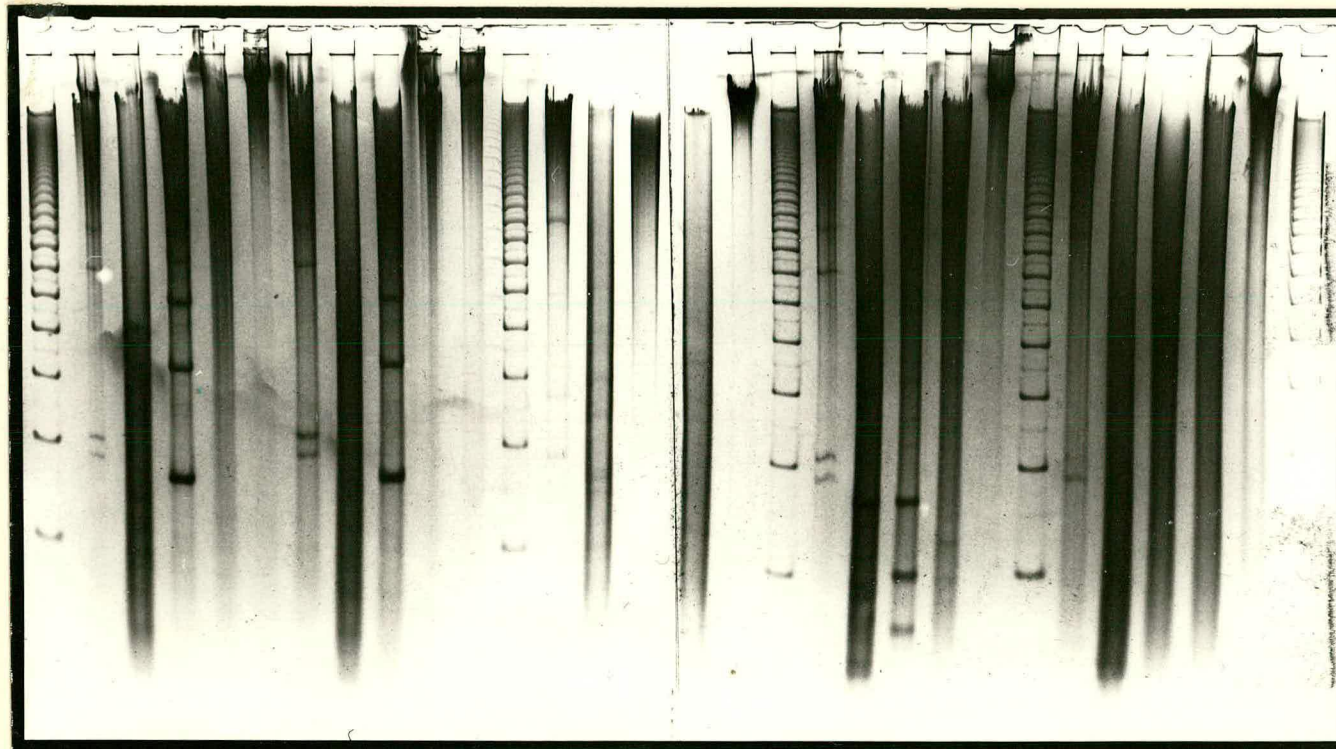




Fig. 7 Analytical CsCl centrifugation of 1 A. flavicollis, 2 A. sylvaticus, 3 A. agrarius, 4 A. microps, 5 A. mystacinus. The tracings are aligned on the M. lysodecticus density marker.



Plate 2 Restriction patterns of  
Apodemus DNAs



1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5  
A. sylvaticus-A. flavicollis A. agrarius A. microps A. mystacinus

Enzymes used: 1, Eco RI. 2, EcoRII. 3, Hin. 4, Hae II. 5, Hae III



Table 2 Band sizes in Apodemus DNAs

| Enzyme   | <u>A. sylvaticus</u> | <u>A. flavicollis</u> | <u>A. agrarius</u> | <u>A. microps</u> | <u>A. mystacinus</u> | Band Size |
|----------|----------------------|-----------------------|--------------------|-------------------|----------------------|-----------|
| Hind III | -                    | -                     | -                  | +                 | -                    | 180       |
|          | -                    | -                     | -                  | +                 | -                    | 240       |
|          | +                    | +                     | -                  | +                 | -                    | 370       |
|          | +                    | +                     | -                  | +                 | -                    | 740       |
|          | +                    | +                     | -                  | +                 | -                    | 1110      |
| Eco RI   | +                    | +                     | +                  | +                 | +                    | 430       |
|          | +                    | +                     | -                  | +                 | -                    | 500       |
|          | +                    | +                     | -                  | +                 | -                    | 1440      |
|          | +                    | +                     | +                  | +                 | +                    | 1850      |



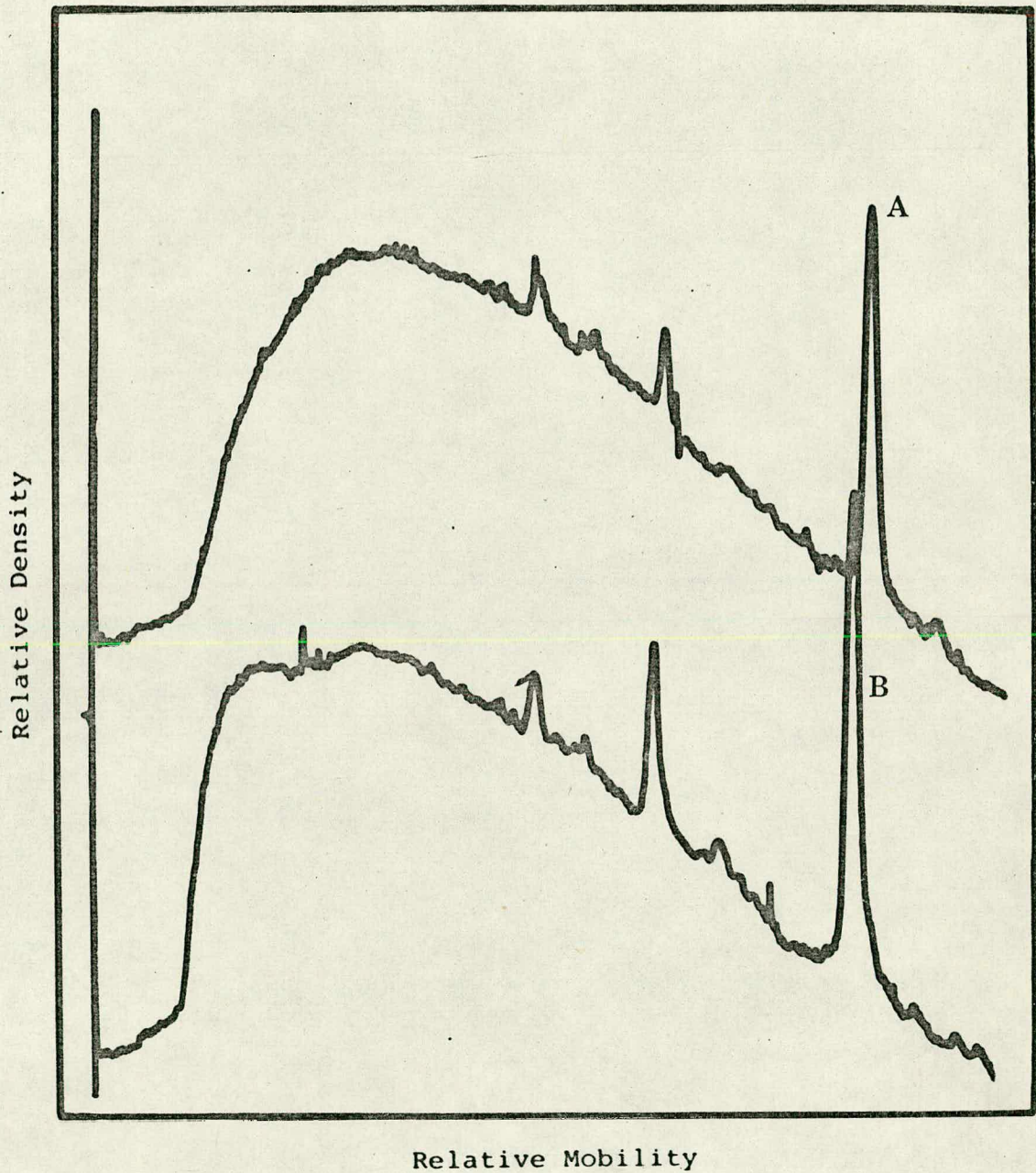


Fig. 9 *A. sylvaticus* and *A. flavicollis* total DNAs after digestion with Hin restriction endonuclease and electrophoresis on 2% agarose gels. A) *A. sylvaticus* ,B) *A. flavicollis*



1110 base pairs in the case of both A. sylvaticus and A. flavicollis. A. microps shows a more complex banding pattern in which there is a strong band 240 base pairs long with weaker bands at 185 and 370 base pairs: in partially digested A. microps total DNA bands are visible with sizes of 740 and 1110 base pairs (figure 10), in addition there are very weak bands with estimated lengths of 550 and 600 base pairs.

The simplest explanation for the occurrence of bands which are integral multiples of one size in a complete digest of a DNA is that mutation has occurred in a tandemly repeated sequence and has removed a number of restriction sites generating multiples of the basic repeat length (see figure 11). This mechanism will explain the patterns found in A. sylvaticus and A. flavicollis, but in A. microps the situation is different in that the predominant band in the series is the 180-185 base pair fragment. There are also relatively large quantities of repeated sequences 240 base pairs long which do not seem to be tandemly repeated since integral multiples of this length are not seen in partial digests.

If these repeated sequences, which are 240 base pairs long, are short interspersed repeated sequences it is unlikely that they have evolved from a tandemly repeated sequence in the way suggested for intermediate and single copy DNA (Sutton & McCallum, 1971). It seems more likely that this type of repeat would be formed by selection of the sequence to carry out a function at many places in the DNA and that this function would be sequence dependent.

Although the A. microps pattern is different the precursor/product type of behaviour shown in figure 10 (where increasing times of digestion decreases the proportion of the band material in peaks which are multiples of 180 base pairs and at the same time increases the quantity of material in the 180 base pair peak) strongly implies that these peaks can be related to the series of peaks found in Hin digests of A. sylvaticus and A. flavicollis.



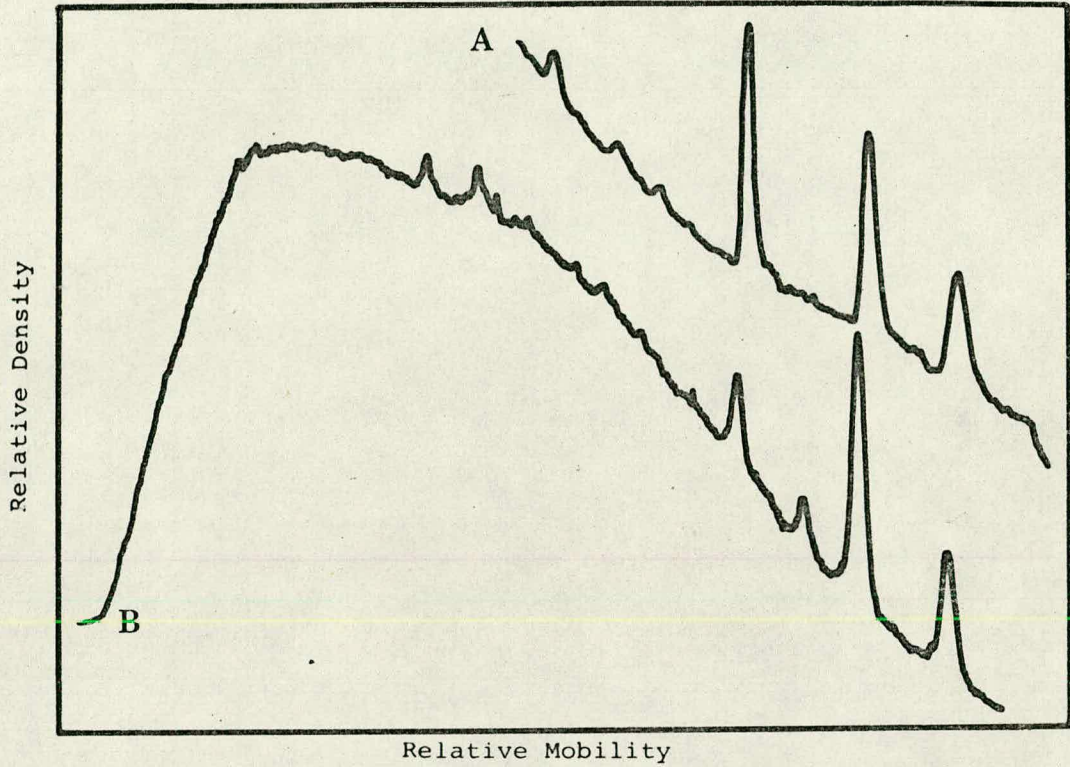


Fig. 10 Gel electrophoresis patterns of *A. microps* total DNA after digestion with *Hin* restriction endonuclease A) partial digest, B) complete digest.



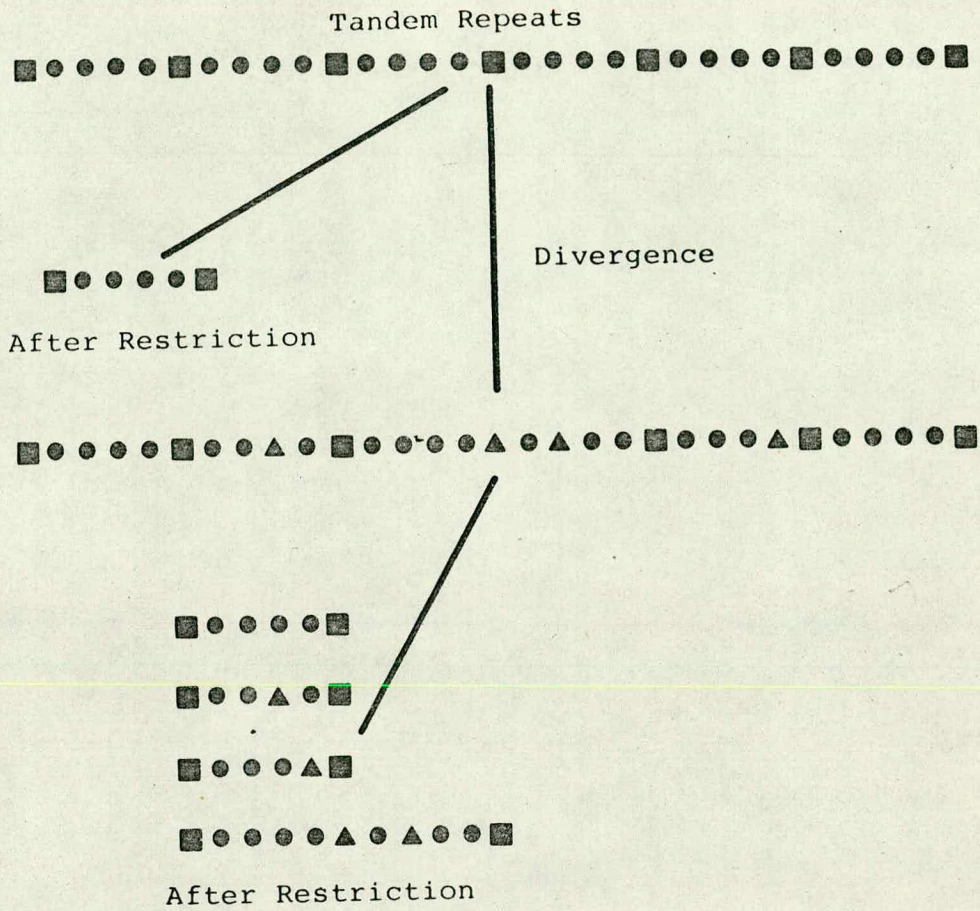


Fig. 11 The effect of divergence on the restriction endonuclease products from a tandemly repeating sequence. Restriction site  $\blacksquare$ , other sequences  $\bullet$ , sequences altered by divergence  $\blacktriangle$ .



### Restriction of purified *A. sylvaticus* satellite DNA

When *A. sylvaticus* light satellite DNA, purified by three cycles of CsCl centrifugation is digested with Hin restriction endonuclease more than 90% of the DNA is found in bands of the same sizes as the strong bands observed in digests of total DNA from *A. sylvaticus* or *A. flavicollis*.

There are two enzyme activities found in Hin restriction endonuclease (Smith & Wilcox, 1970) described as Hind II and Hind III. The bands in digests of *A. sylvaticus* and *A. flavicollis* are produced by Hind III. This can be shown either by digesting the DNAs with the separated enzymes or by digesting with a preparation of Hin enzyme which has lost Hind II activity (this can be tested by digesting  $\lambda$ .C DNA) when the band pattern given by Hin and *A. sylvaticus* and *A. flavicollis* is unchanged.

Figure 12b shows a partial digest of *A. sylvaticus* with Hin in which multiples of up to 10 of the 370 base pair repeat length are detectable. The presence of these multiples in the partial digest shows that they are tandemly linked in the native satellite, the low proportion of material which, after restriction, has a length which does not correspond to a multiple of 370 base pairs in both partial and complete digests suggests that as isolated the satellite is composed of tandemly repeated 370 base pair units which are diverged from each other with little if any interspersions of other sequences.

Un-restricted satellite has a molecular weight of greater than  $3 \times 10^6$  daltons, i.e. about 5000 base pairs, since it is excluded from agar gels which do not exclude DNA below this size. This means that at least 14 and most probably many more than 14 of the 370 base pair repeats are tandemly linked within the isolated molecules.

### Estimation of divergence at the restriction sites

Slack (1974) has derived an equation which can be used



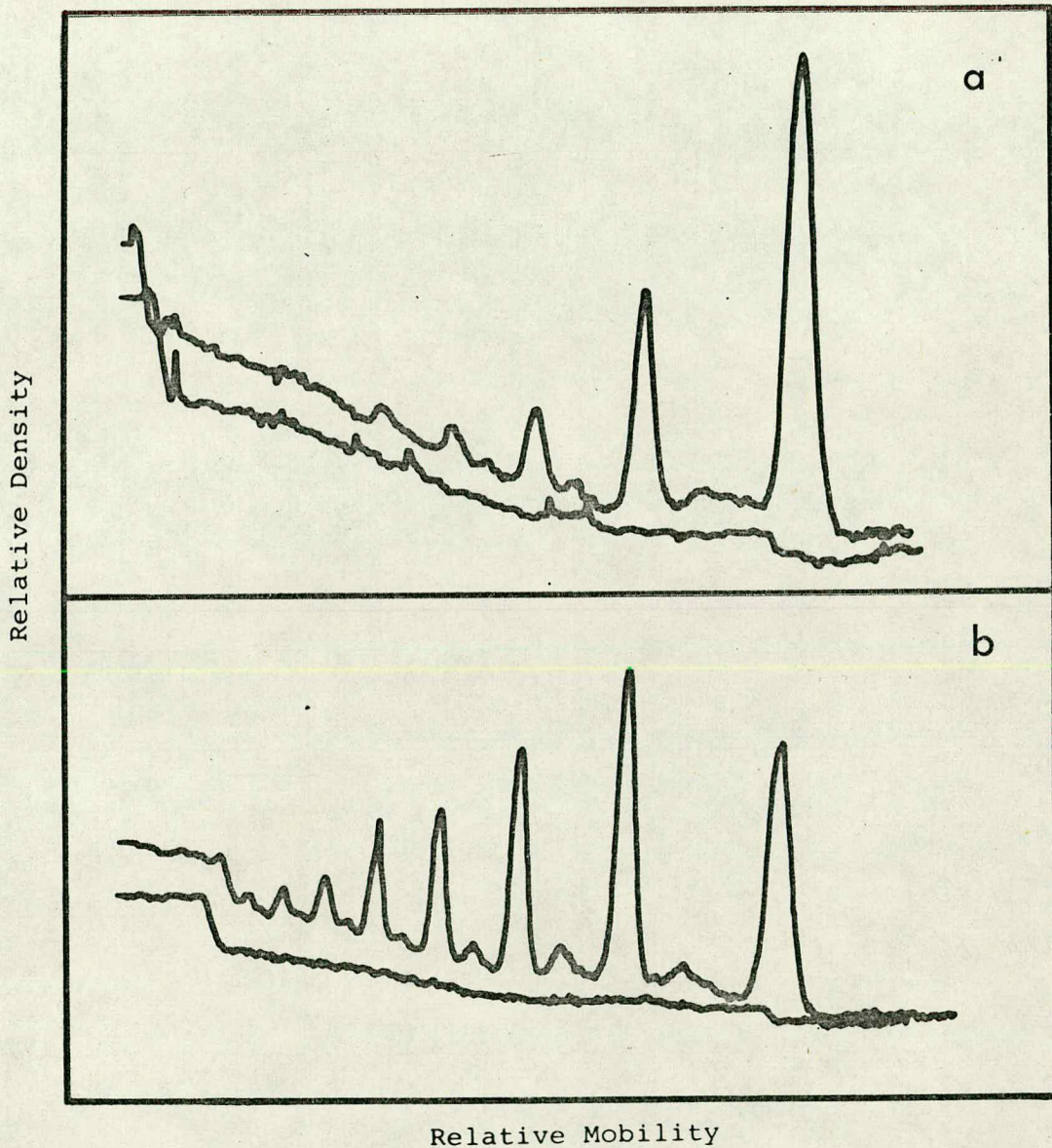


Fig. 12 *A. sylvaticus* satellite DNA after a) complete digestion, 0.5 microgrammes of DNA digested with 0.05 units of *Hin* restriction endonuclease for 18 hours, b) partial digestion of 1.0 microgrammes of DNA with 0.005 units of enzyme for 2 hours.



to calculate the percentage of the restriction sites which have been lost by mutation from the proportion of the DNA in the monomer, dimer, trimer, and higher integral multiple peaks. For multiples of the peaks arising by random divergence in the restriction sites then the fraction  $(a)_n$  of the DNA in a size class  $n$  is given by

$$(a)_n = nP(n-1)(1-P)^2$$

or

$$\log \frac{(a)_n}{n} = (n-1) \log P + 2 \log(1-P)$$

where  $P$  is the proportion of the sites which have been modified. Figure 13 is a plot of  $\log \frac{(a)_n}{n}$  v.s.  $n-1$  and thus has a slope of  $\log P$  and intercept  $2 \log 1-P$ .

If  $(1-P)$  is the proportion of unmodified sites, the divergence per base pair in the restriction site can be calculated if it is assumed that any base change in the site is effective in modifying the site so that the enzyme no longer recognises it. A further assumption is that all nucleotide pairs have an equal probability of a base change occurring in them. In this case the divergence per base pair  $p$  in the sequence is

$$(1-p)^n = 1-P$$

where  $n$  is the number of base pairs in the restriction site.

The sequence recognised by Hind III is six base pairs long (Old, Murray and Roizes, 1974). The value of  $p$  calculated on the basis of  $P$  equal to 18% and  $n$  equal to 6 is 3.2%.

A further problem with this method for estimating divergence is that incomplete digestion increases the amount of DNA found in the dimer and trimer peaks (see figure 12a and b and figure 13), for this reason the value obtained for the divergence in the restriction site is an upper limit although in these cases complete digestion was thought to have been reached since a number of digestions gave a constant value for  $P$ .



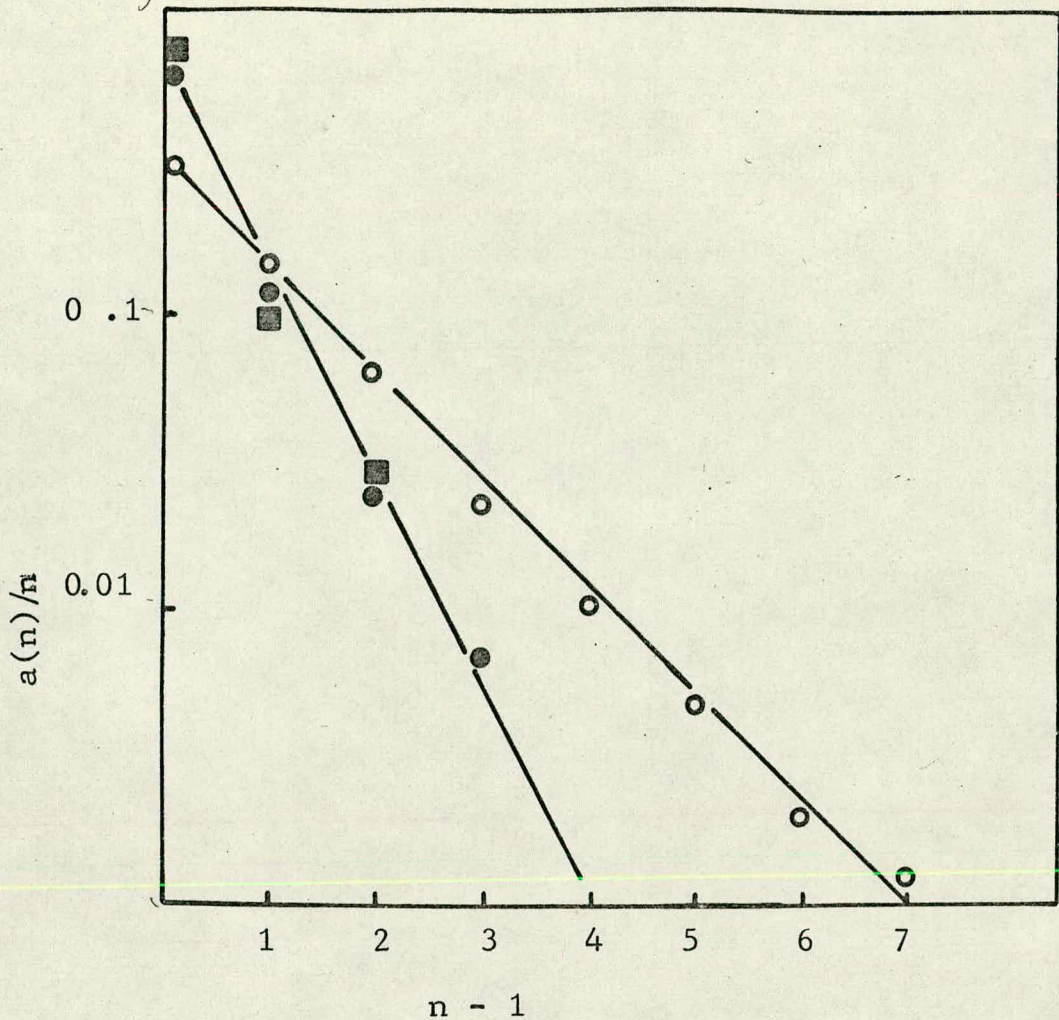


Fig. 13 Measurement of divergence in the restriction sites. For dimers and trimers arising from divergence in the restriction site, if the distribution of modified sites is random, the fraction  $(a)n$  of the DNA in a size class  $n$  is-

$$a(n) = nP^{(n-1)} (1-P)^2$$

where  $P$  is the fraction of modified sites  
 The above plot of  $\log a(n)/n$  would give a straight line slope  $\log P$  and intercepts  $2 \log(1-P)$ .

- Partial digest of *A. sylvaticus* sat.
- Complete digest of *A. sylvaticus* sat.
- Complete digest of *A. flavicollis* total DNA.



The divergence measured in this way is not comparable directly with the divergence, measured by thermal stability, of reassociated duplexes since the latter method gives an average value over all cross-reacting species of DNA molecules whereas restriction enzymes can only recognise short regions of a larger sequence.

Insufficient A. flavicollis satellite DNA was available pure to determine the proportions of the DNA in different size classes after restriction; total DNA was used and the peak areas measured from microdensitometer scans. This is less accurate than using the purified satellite because the underlying pattern of DNA fragments from other sequences makes estimation of the base line under the satellite peaks difficult. The results for partial and complete digests of A. sylvaticus satellite DNA and A. flavicollis total DNA are shown in figure 13. All points fall on straight lines indicating that divergence in the restriction sites is randomly distributed in the case of complete digests. As the linear relationship also holds for the partial digest these fragments must also be generated by a random process, i.e. the restriction enzyme does not act sequentially at adjacent sites by moving along the DNA since this would produce a distribution with an excess of monomers over the random expectation. The mechanism of the enzyme must involve steps of attachment to the DNA, cleavage and then release of the enzyme from the DNA.

#### The formation of $\frac{1}{2}$ mers of the basic repeat length.

One further important observation on the patterns given by electrophoresis of *Hin* digestion products of A. sylvaticus and A. flavicollis DNAs is the presence of intermediate bands in the partial digests. These are seen most clearly in partial digests of A. sylvaticus satellite (fig. 12b). These diffuse bands are centred around the positions expected for multiples of 1.5, 2.5, 3.5 etc. of the 370 base pair length. In the complete digest these bands





are reduced in quantity if not completely absent. There is no detectable amount of fragments 180 base pairs long as would be expected if these intermediate bands were composed entirely of multiples of 180 base pairs. On the other hand if these intermediate bands were an integral number of 370 base pair repeats plus one 180 base pair repeat tandemly linked then the proportion of each intermediate which is DNA with a 180 base pair repeat is 33% for the 1.5 mer, 20% for the 2.5 mer, 14% for the 3.5 mer etc. These peaks form only about 2% of the total peak area in partial digests of A. sylvaticus satellite DNA, the 180 base pair fragments released by complete digestion would thus be present in quantities too small to detect, especially if the peak were diffuse.

Interspersed 180 base pair repeats could arise by the introduction of a Hind III site into the 370 base pair sequence at its mid-point or by unequal crossing over between repeats as in figure 14. The two possibilities give rise to different predictions of the ratios of  $\frac{1}{2}$ mer to  $1\frac{1}{2}$ mer; divergence would produce two halfmers with  $1\frac{1}{2}$ mers arising when the adjacent site was diverged from the restriction sequence. The frequency of the  $1\frac{1}{2}$ mer would be the product of the frequency of finding a modified restriction site and the frequency of introduction of a restriction site. For the most favourable case where only one base change was required to introduce the restriction site this frequency is  $P \times C$  where  $C$  is the probability of introduction of a site. The ratio of the frequency of  $\frac{1}{2}$ mers to  $1\frac{1}{2}$ mers is then  $C/P \times C = 1/P$ , in this case 6:1. Unequal crossing over would form  $\frac{1}{2}$ mers and  $1\frac{1}{2}$ mers in equal proportions, this corresponds more closely with the amounts found.

In complete digests of A. sylvaticus satellite DNA less than 2% of the area of band I is found in the peak due to  $1\frac{1}{2}$ mers, since each out-of-register crossover would produce a  $1\frac{1}{2}$ mer and a  $\frac{1}{2}$ mer, one crossover per 75 monomers would give 2% by weight of the  $1\frac{1}{2}$ mers. This frequency is much lower than that found by Southern (1974) for the



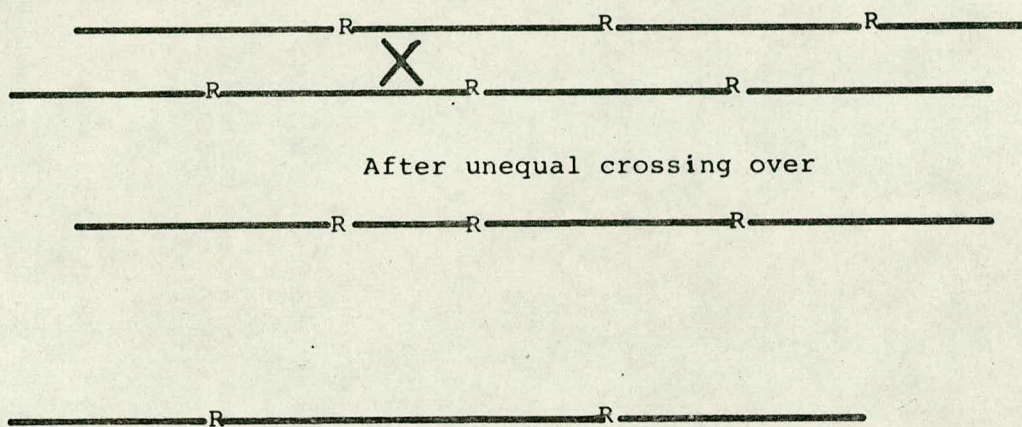


Fig. 14 The formation of  $\frac{1}{2}$ mers of the basic repeat length by unequal crossing over. R—R, positions of restriction sites



comparable situation in mouse satellite DNA and probably reflects the differences found in the reassociation registers of the two monomer bands. A. sylvaticus and A. flavicollis band I DNAs have a reassociation register of the same length as the restriction repeat whereas the mouse satellite band I DNA appears to have a reassociation register half the length of the restriction repeat. Thus it seems probable that the frequency of crossing over is proportional to the ability of the crossing over sequences to reassociate together, i.e. to their sequence homology.

#### Thermal denaturation of A. sylvaticus and A. flavicollis band I DNA

DNA in the 370 base pair restriction band was cut from the gel, recovered on hydroxylapatite and dialysed into 0.01 M NaCl, 0.001 M Tris/HCl, 0.0001 M EDTA pH 8.0. A similar volume of the buffer dialysed over night was used as a blank since it was found that dialysis reduced the 260/230 absorption ratio and hence the observed hyperchromicity. This is due to the leaching of light scattering material into the dialysed solution. When this precaution was taken the hypochromicity of the DNA was always greater than 30% after the initial thermal denaturation.

Figure 15a shows the thermal denaturation profiles of A. sylvaticus native DNA from the smallest restriction band (band I), figure 15b shows the corresponding profile for A. flavicollis. For ~~the~~ both species the DNAs have a melting temperature of  $63.5^{\circ}$ . Allan (1974) found that the  $T_m$  of the native isolated satellite DNAs of both species was  $67.5^{\circ}$  C, this  $4^{\circ}$  discrepancy is too large to be explained by experimental error, since over seven determinations of the  $T_m$  for the restriction band DNA, the variation of  $T_m$  was within  $\pm 0.5^{\circ}$  of  $63.5^{\circ}$ . Contamination of the satellite with sequences which do not have a repeat length of 370 base pairs is low and cannot explain this large a difference in  $T_m$ . Similarly the contamination of the isolated



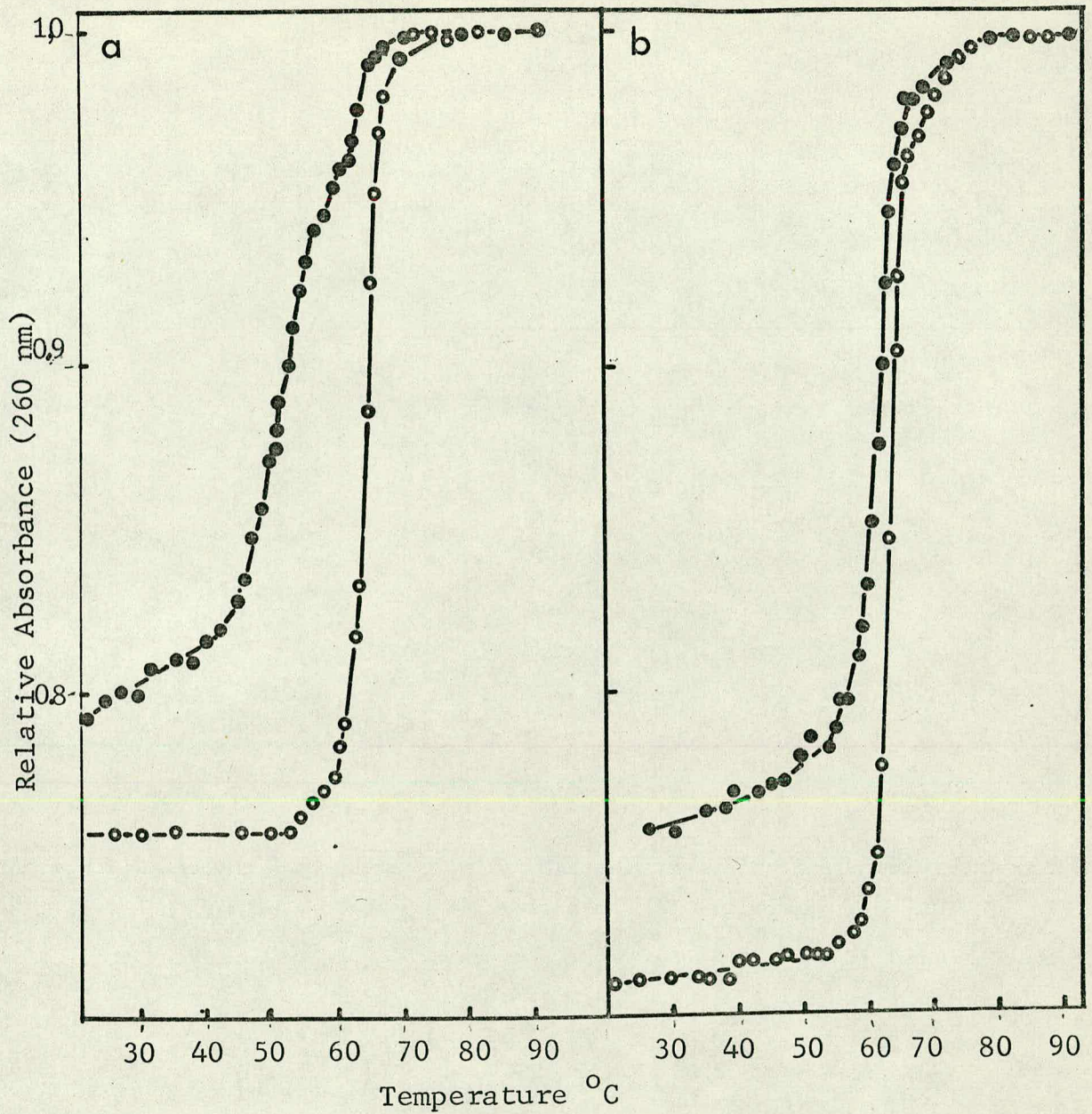


Fig. 15. Melting curves of 370 base pair band DNA from a) *A. sylvaticus* and b) *A. flavicollis*. DNA in 0.01 M NaCl, 0.001 M Tris/HCl, 0.0001 M EDTA was heated at 0.5°C per minute in sealed cuvettes.

—○— Native DNA  
 —●— DNA reassociated to  $Cot\ 10^{-1}$



band I DNA is too low for this to be a reasonable explanation.

The procedure for recovery of DNA from gels had no detectable effect on the  $T_m$  of *E. coli* DNA and thus the most feasible remaining explanation is that the difference in  $T_m$  reflects some organisation of sequences which differs between restriction fragments and sonicated DNA. This is more a rationalisation than an explanation.

Bands from both species give extremely sharp unimodal melts suggesting homogeneity of sequence and uniform base composition along the 370 base pair length of the isolated duplexes.

Marmur & Doty (1962) suggest, on the basis of determinations of the  $T_m$ s of bacterial DNAs, that a  $T_m$  of  $63.5^\circ$  in the salt concentration used here would correspond to a base composition of 38% G+C. Taking into account the problem of correcting the  $T_m$  for the different salt concentrations this value agrees well with the values of 34.7 and 33.7% G+C estimated from the buoyant densities of *A. sylvaticus* and *A. flavicollis* satellite DNAs (Allan 1974).

#### Renaturation of the restriction band DNA

Since Allan (1974) had shown that the satellite DNAs of *A. sylvaticus* and *A. flavicollis* reassociate too rapidly to follow in a normal spectrophotometer if the reaction is carried out in  $0.18 \text{ M Na}^+$ , the rates of reassociation of the band I DNAs were measured in  $0.04 \text{ M NaCl}$  and the rates corrected to give the rate expected in  $0.18 \text{ M Na}^+$  from the tables of Britten & Smith (1970).

The reassociation kinetics of the isolated Band I DNAs are shown in figure 16. Neither *A. sylvaticus* nor *A. flavicollis* band I material returns completely to the native optical density. A part of this residual hypochromicity will be due to mismatched bases but this can not account for all of the hypochromicity. Figure 9, traces of typical patterns of the total DNA after restriction and



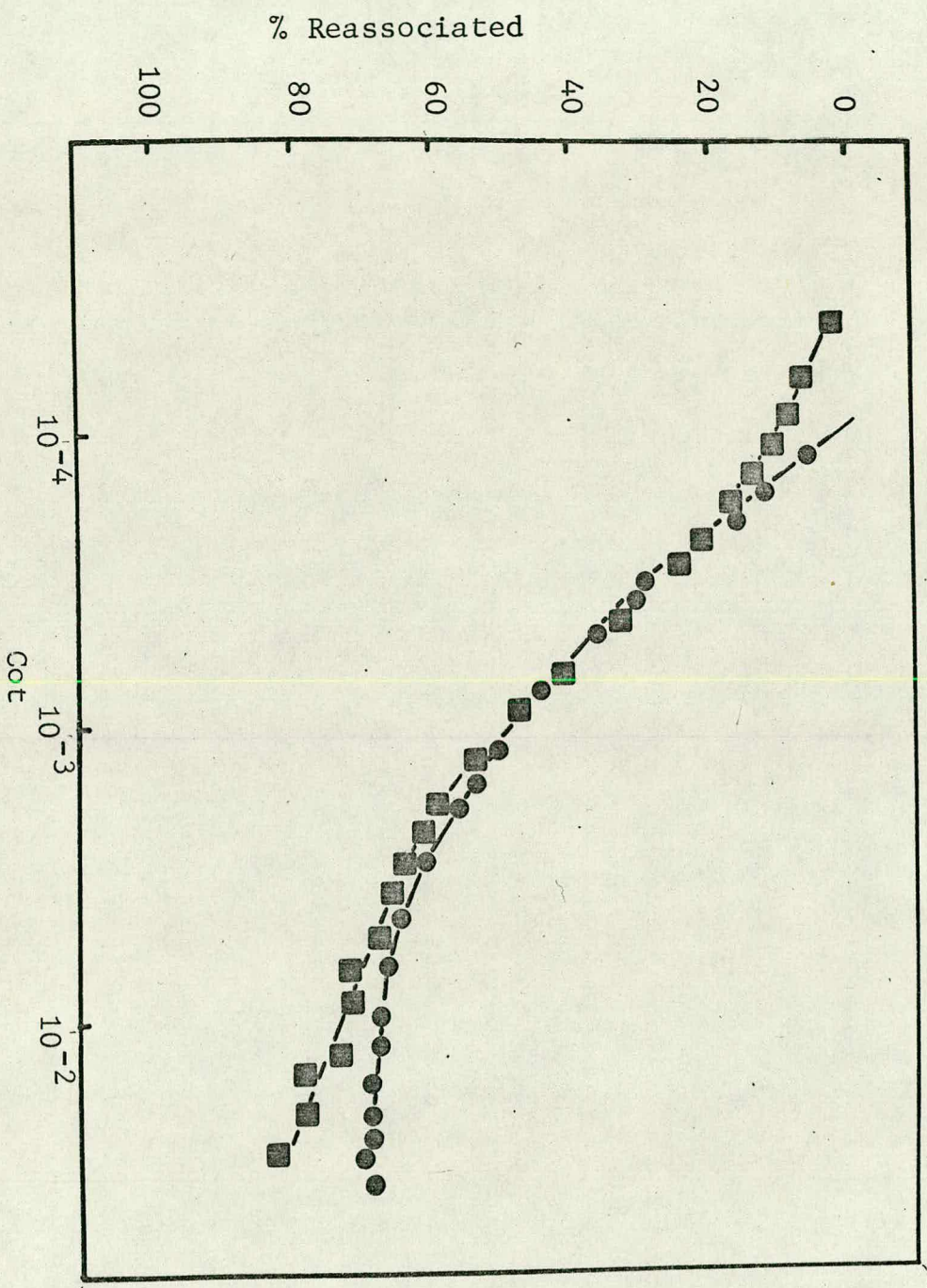


Fig. 16 Reassociation of 370 b.p. band DNA. DNA was reassociated in 0.04 M NaCl, 0.001 M Tris/HCl, 0.0001 M EDTA pH 7.9 at 50°C. Values of Cot are corrected to 0.18 M Na<sup>+</sup>.

*A. sylvaticus* —■— *A. flavicollis* —●—



analysis on agarose gels, shows that not all of the DNA in the slice cut from the gel is the satellite derived material; the proportion of non-satellite material is higher for A. flavicollis than for A. sylvaticus since there is a smaller fraction of the total DNA which is satellite in A. flavicollis.

The contaminating DNA is likely to be enriched in unique sequences since it is small (Botchan & McKenna, 1973; Southern & Roizes, 1973) and so will not reassociate significantly at the Cot values reached in these experiments.

The DNA which reassociates does so over a range of Cot values with a factor of 100 between the beginning and end of the reaction suggesting that a single species of molecule is reacting. Because of this, and because the extent of contamination of the band I DNA from the two species provides a reasonable explanation of the different extents of reaction observed, it is probable that the end point of the reaction represents complete reassociation of the band I DNA which is derived from satellite sequences.

On this basis the calculated values of  $Cot_{1/2}$  for A. sylvaticus and A. flavicollis respectively are  $4.8 \times 10^{-4}$  and  $2.8 \times 10^{-4}$  (corrected to  $0.18 \text{ M Na}^+$ ). If the end point of the reaction was a return to native optical density, both DNAs would have a  $Cot_{1/2}$  of  $6 \times 10^{-4}$ .

Sequence complexity, estimated according to Wetmur & Davidson (1968) on the basis of the  $Cot_{1/2}$  values are 430 base pairs for A. sylvaticus and 260 base pairs for A. flavicollis; if the  $Cot_{1/2}$  of the DNA from both species is taken to be  $6 \times 10^{-4}$  then the sequence complexity would be 540 base pairs.

These values for complexity correspond well to the repeat length of the fragments after restriction, this in turn suggests that the reassociation register is the same length as the restriction repeat with no shorter repeats detectable under these conditions of reassociation. A further implication is that the effect of mismatch on the rate of reassociation is small in this case.



### Thermal Stability of Reassociated DNA.

The reassociated duplex of restriction band I from A. sylvaticus melted with a distinctly biphasic profile (figure 15b); 80% of the DNA melted with a  $T_m$  of  $51.5^{\circ}\text{C}$  ( $\Delta T_m$   $12^{\circ}\text{C}$ ), the remaining 20% melted over  $5^{\circ}\text{C}$  with a  $T_m$  of  $61.5^{\circ}\text{C}$  ( $\Delta T_m$   $2^{\circ}\text{C}$ ). Allan (1974) did not observe a two step melt for reassociated, sonicated DNA; this could be due either to the lower extent of reassociation he observed (30% for satellite compared to 80% for the restriction band) giving low hypochromicities on melting the duplex or to differences in behaviour of randomly sonicated DNA compared to restriction fragments.

Randomly fragmented DNA about 500 nucleotides long, produced by sonication; would consist of a population of molecules which had on average 1.4 restriction repeats randomly positioned inside the molecules. On reassociation the possible spread of structures would be large and might not give a two step melt. One obvious difference between sonicated and restricted satellite DNAs from the Apodemus species is that reassociated sonicated satellites form hypersharp bands in CsCl gradients indicating the formation of large networks of DNA, restricted satellites do not behave in this way (see below).


A. flavicollis reassociated restriction band I does not show a two step melt, its profile is sharp with a  $T_m$  of  $61.5^{\circ}\text{C}$  (fig 15b). When the two species are compared it seems probable that the A. flavicollis reassociated material which has a  $\Delta T_m$  of  $2^{\circ}$  corresponds to the 20% of the A. sylvaticus band I DNA which has an identical  $T_m$ .

There is little difference between the reassociation rates of the band I DNAs from the two different species although there is a difference in  $\Delta T_m$  of  $10^{\circ}$ . This is in line with the data of McCarthy & Farquhar (1972) but does not confirm the predictions of Southern (1971) and Sutton & McCallum (1971). This may be explained simply if the constant sequences at the ends of the DNA fragments act as



Table 3 Properties of A. sylvaticus and A. flavicollis band I DNAs

|  | <u>A. sylvaticus</u>   | <u>A. flavicollis</u>  |
|--|------------------------|------------------------|
| Relative quantity                                    | 7%                     | 5%                     |
| Native T <sub>m</sub>                                | 63.5°C                 | 63.5°C                 |
| T <sub>m</sub>                                       | 12°C & 2°C             | 2°C                    |
| Cot <sub>1/2</sub> (moles sec litres <sup>-1</sup> ) | 4.8 x 10 <sup>-4</sup> | 2.8 x 10 <sup>-4</sup> |





stable nucleation sites and define the reassociation register.

The significance of differences between divergences of *A. sylvaticus* band I DNA measured by thermal stability and restriction.

Although the thermal stabilities of the 370 base pair band material after reassociation are very different in the case of *A. sylvaticus* and *A. flavicollis* the divergence measured by restriction is the same in both cases and close to the value from the  $\Delta T_m$  of *A. flavicollis*. One possible explanation could be that the restriction sites in *A. sylvaticus* satellite DNA are more conserved than the rest of the repeat unit. This would have important consequences since the implication is that these regions have some functional significance. If there were conserved regions in the *A. sylvaticus* satellite then the two step melt of the reassociated material could be explained on the basis that the DNA with a low  $\Delta T_m$  was the conserved region.

The relationship of the low melting material to the high melting material in *A. sylvaticus* reassociated duplex can be approached in a number of ways; there are two possible different structures, the high and low melts could reflect different extents of mismatch in different parts of the same molecules or two populations of molecules which have different extents of mismatch (figure 17). These two possibilities can be distinguished by heating the DNA to the inflection in the melting profile and cooling rapidly. If the biphasic melt is due to two different populations of molecules then at the temperature of the step one population will be denatured and one native; cooling will cause only the small reduction in absorbance due to base stacking in the single strands. On the other hand if the two parts of the melting curve reflect differently matched parts of the same molecule then at the temperature of the step the molecules will be partially double stranded.



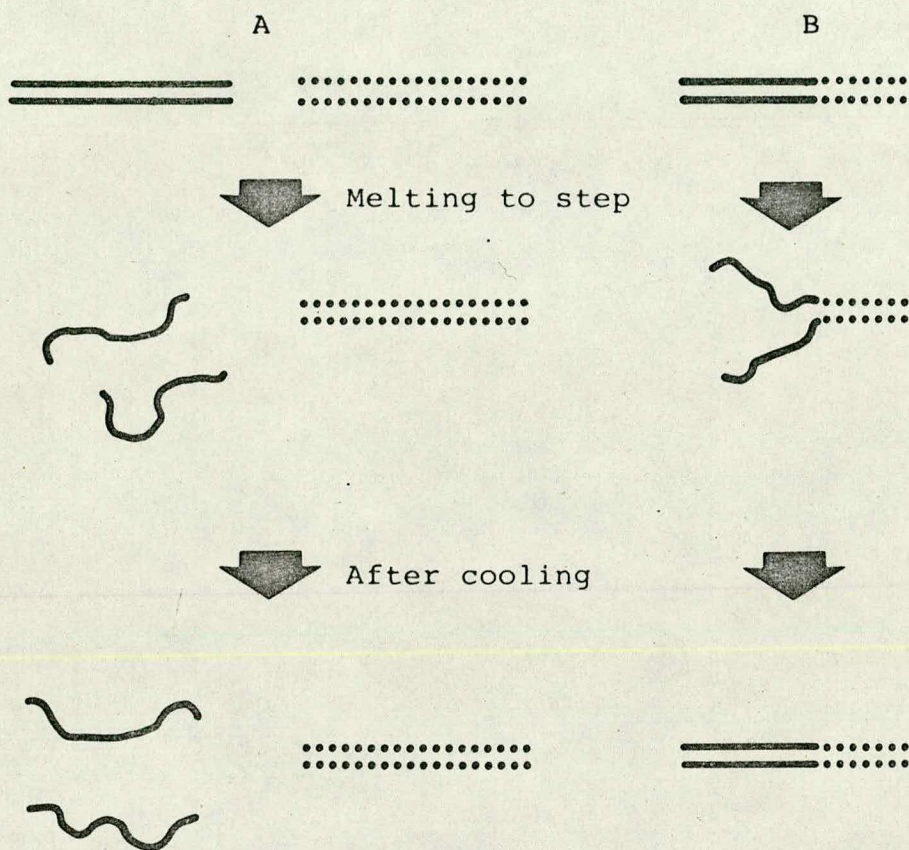


Fig. 17 Possible alternative situations giving a two step melt and differences between them. A, two populations of molecules with different extents of mismatch, B different extents of mismatch in different parts of the same molecule.



This region will keep the strands in register and the re-association rate of the single strands will be that of the fast zippering step of the reaction mechanism (Wetmur & Davidson 1968). In this case rapid cooling will return the optical density to the starting value instantly.

The results of such an experiment are shown in figure 18. When the temperature of the cuvette was brought to 20°C from temperatures below the step at 59°C the optical density returned to almost that of the reassociated material. At temperatures of 59°C or above returning the cuvette to a low temperature caused no return to the optical density of the reassociated DNA. The reassociated material must thus consist of two populations of molecules with different extents of mismatch and hence the biphasic melt is not due to the presence of a more conserved region.

A situation such as this could occur either with the two populations of molecules cross-reacting in reassociation or with two sets of molecules which do not reassociate with each other. The latter case would be very surprising here since it would suggest that there are two unrelated or very distantly related repeated sequences in A. sylvaticus satellite DNA which had identical repeat lengths and would point to a multiplication mechanism which itself determines the repeat length of the repeated sequence it creates.

A more reasonable explanation of the melting behaviour would be that a number of families were present in the band I DNA which had been derived from a common repeated sequence but from different diverged repeats. A small amount of divergence would occur after multiplication within each family. Denaturation of such a set of molecules would produce single strands which could cross-reassociate but the thermal stability of the resulting duplex would be lower if the strands were from different families. The restriction repeat would be present in all families (with some divergence) since they are derived from related



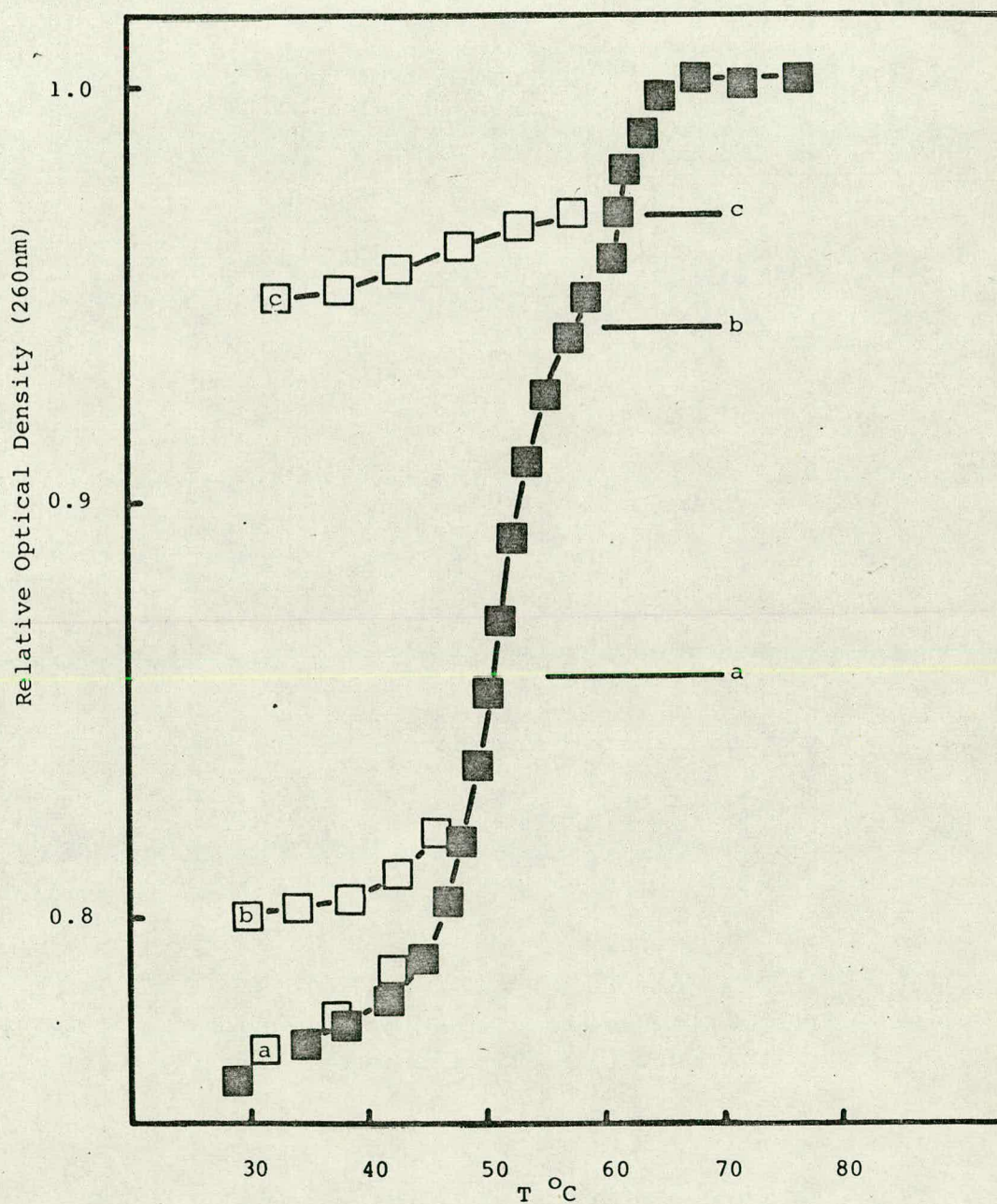


Fig. 18 Hypochromicity after melting and rapid cooling at different temperatures a) 50°C b) 57°C c) 60°C .



sequences and the divergence measured from restriction patterns would be the same in both cases since the divergence within each family would be the same. Between the families the 6% divergence would not be detected since only those families which were multiplied from repeats containing the restriction site would be found in the band DNA.

Cross-reassociation between families within the satellite of *A. sylvaticus*.

Cross reassociation between the related families postulated above can be shown by melting the reassociated DNA to the step in the melt and allowing the denatured DNA to reassociate. If cross-reassociation occurs then a proportion of the duplexes formed during the second reassociation should melt at the higher temperature and result in an increase in the proportion of the duplexes melting above the step. This increase should be 20% of the 80% which originally melted below the step, i.e. an increase from 20 to 34% of the DNA melting with a  $T_m$  of 61.5 on complete melting after the second reassociation.

Non cross reacting populations will show no change in the proportion of the DNA melting above the step after the second reassociation since the duplexes formed on the second reassociation will have the same degree of mismatch as the original duplexes.

Figure 19 shows the melt of half a sample of band I DNA from *A. sylvaticus* reassociated to a  $Cot$  of  $10^{-1}$ , also shown is the melt of the remaining half of this sample of reassociated DNA after melting to  $59^{\circ}$ , reassociating to a  $Cot$  of  $10^{-1}$  (based on the single strand concentration after melting to the step) in the reassociation buffer and dialysing back into the melting buffer. Hyperchromicity is expressed as a percentage of total hyperchromicity in each case for ease of comparison, the total hypochromicity was reduced in the melt after the second reassociation due



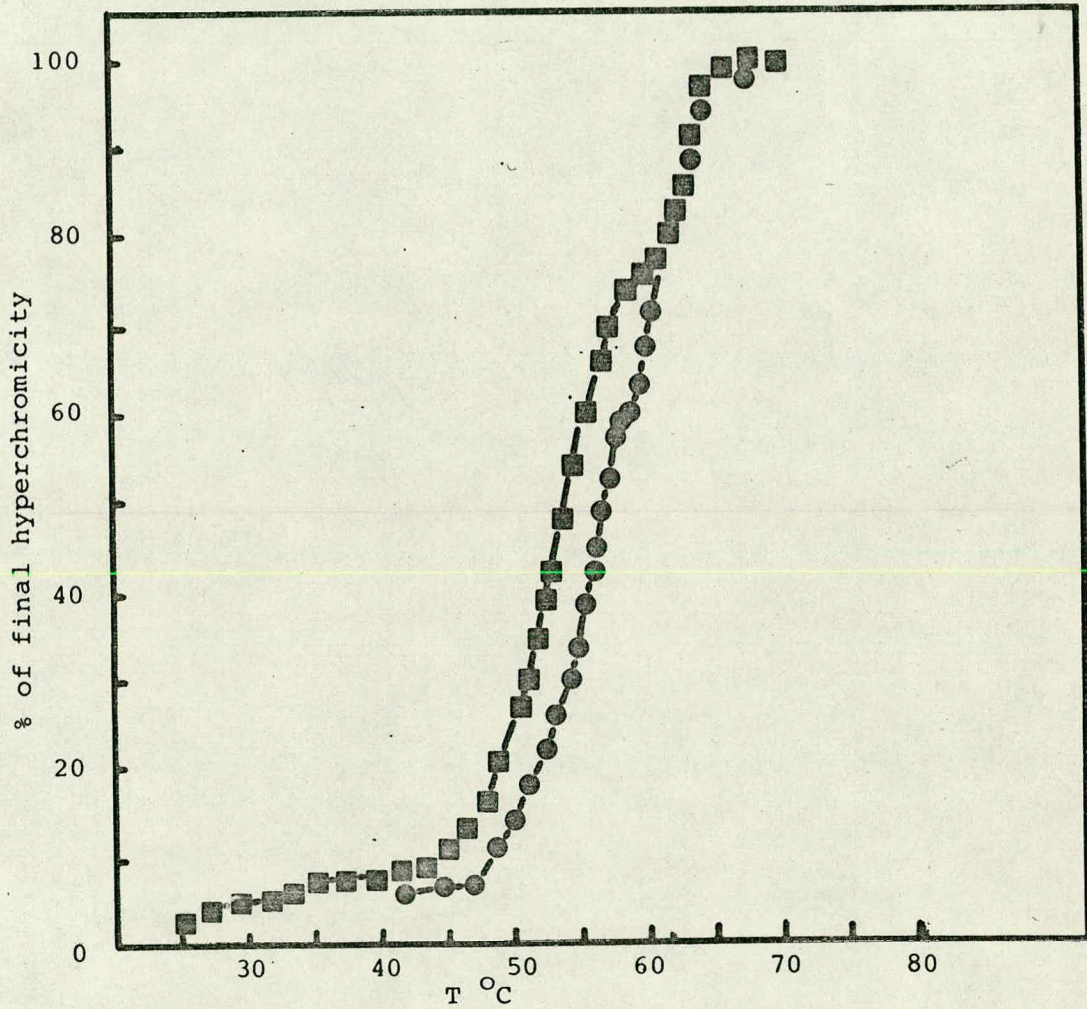


Fig. 19 Thermal denaturation of reassociated Band I DNA a) after melting to 80°C reassociating and melting to 80°C a second time ■—■ b) after melting to 80°C, reassociating, melting to 59°C and reassociating ●—●



to contamination caused by a total of three dialyses of the solutions. A slight inflexion in the melting profile is apparent at  $59^{\circ}$  but, more convincingly, there is 38% of the total hypochromicity above this temperature. This agrees well with the predicted increase to 34% confirming the predictions of the model shown in figure 20, that there are multiple cross-reassociating families of repeated sequences in the band I DNA and hence in A. sylvaticus DNA. If all the families were present in equal quantities the proportion of the band I DNA which melts above the step in the melting curve would indicate that there are five families present.

When the DNA is reassociated after melting to the step in the melt then melted again there is a small decrease in the  $\Delta T_m$  of the part of the curve below the step. If all related families could reassociate equally well with each other this would not occur. This change in  $\Delta T_m$  implies that there are fewer cross reassociating families after the second reassociation and hence that some of the families preferentially form the homoduplexes. The magnitude of the effect is small, the change in the  $\Delta T_m$  is about  $1^{\circ}\text{C}$ .

#### Melts on individuals

The situation in which there are a number of related families of repeated sequences in a restriction band can arise if each individual contributes a single family or if all families are represented in an individual. Whilst the former case is unlikely, as massive mutation rates would be required, it is a possibility which it was necessary to consider because experiments on A. sylvaticus used DNA from five to ten mice pooled but those on A. flavicollis were carried out on DNA from the sole survivor of the laboratory colony of this species.

To rule out the possibility that a single individual contributed one family of the repeated sequence, DNA from



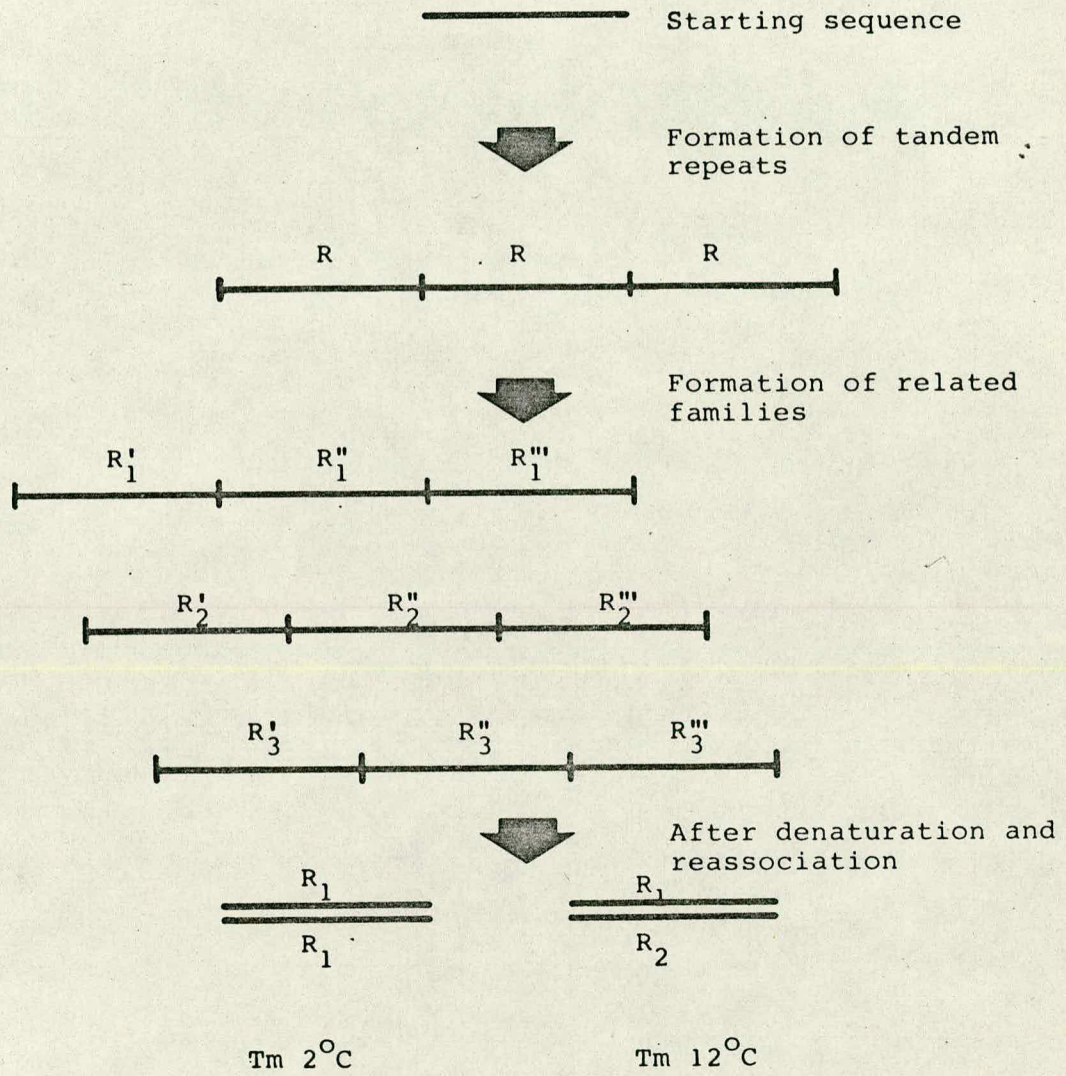


Fig. 20 Multiple related families



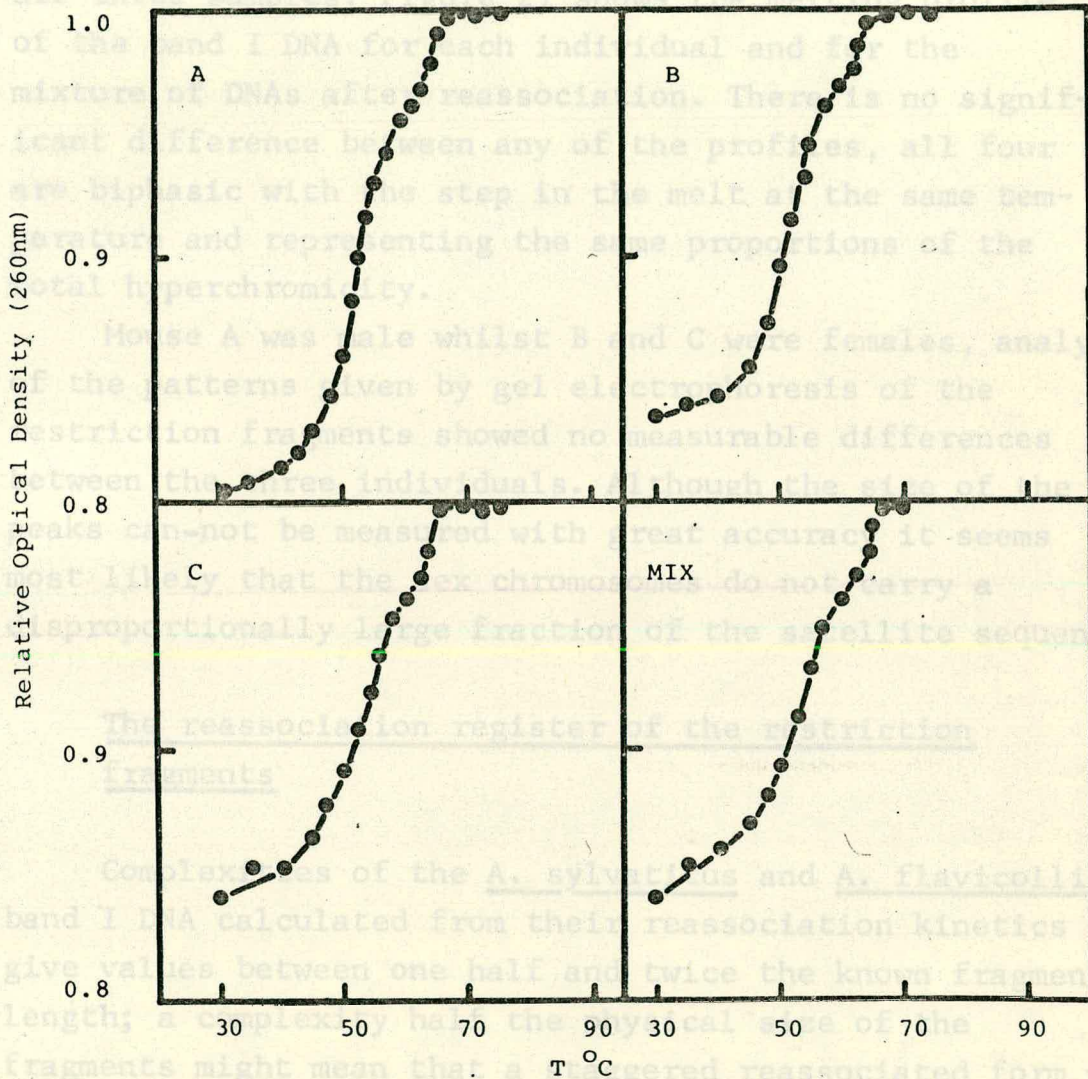


Fig. 21 Thermal denaturation of reassociated band I DNA from individual mice and of a mixture of three band I DNAs reassociated together. A, B and C individual DNAs, MIX mixture of all three DNAs.





Fig. 22 Size of reassociated band 1 DNA fragments. A, upper trace total *A. sylvaticus* DNA after restriction with *Hin* restriction endonuclease and electrophoresis: lower trace, reassociated band 1 DNA after electrophoresis. B, upper trace *A. flavicollis* total DNA: lower trace reassociated band 1 DNA.



DNA, in addition the band has moved ahead of the position of the native DNA. A small quantity of the DNA formed a band at the position of the dimer in this reassociated DNA but there is no trace of the  $\frac{1}{2}$ mer. In contrast the reassociated band I DNA from A. flavicollis ran as a sharp band at the same position as the native DNA with no trace of dimer or  $\frac{1}{2}$ mer. In both cases there was no high molecular weight DNA which proves that the reassociated duplexes have insufficient single stranded regions to permit concatemer formation under these conditions.

The other major difference between the band I DNA in the two species is the extent of mismatch, estimated at 2% for A. flavicollis and 2% and 12% for the duplexes found in A. sylvaticus. This difference could explain the mobility differences if mismatching reduced the segment length of a reassociated duplex and hence increased the mobility of such a molecule relative to native DNA of the same molecular weight. A. sylvaticus would give a broad peak moving faster than native because of the two populations of duplexes with different extents of mismatch, A. flavicollis would have more nearly native behaviour since it has one population with little mismatch.

Electron microscopy confirms that the reassociation is predominantly in the 370 base pair register, the frequency/length histogram (figure 23) demonstrates the size spread of the reassociated molecules. This is based on a latex sphere size standard and so does not give a reliable estimate of length.

No circular molecules were found suggesting that concatemer formation was not blocked by circularisation.

Measurement of the divergence between total A. sylvaticus DNA and A. flavicollis and A. microps DNAs.

Measurement of the divergence between total A. sylvaticus and total A. flavicollis DNAs provides a measure of the times at which the multiplication processes which



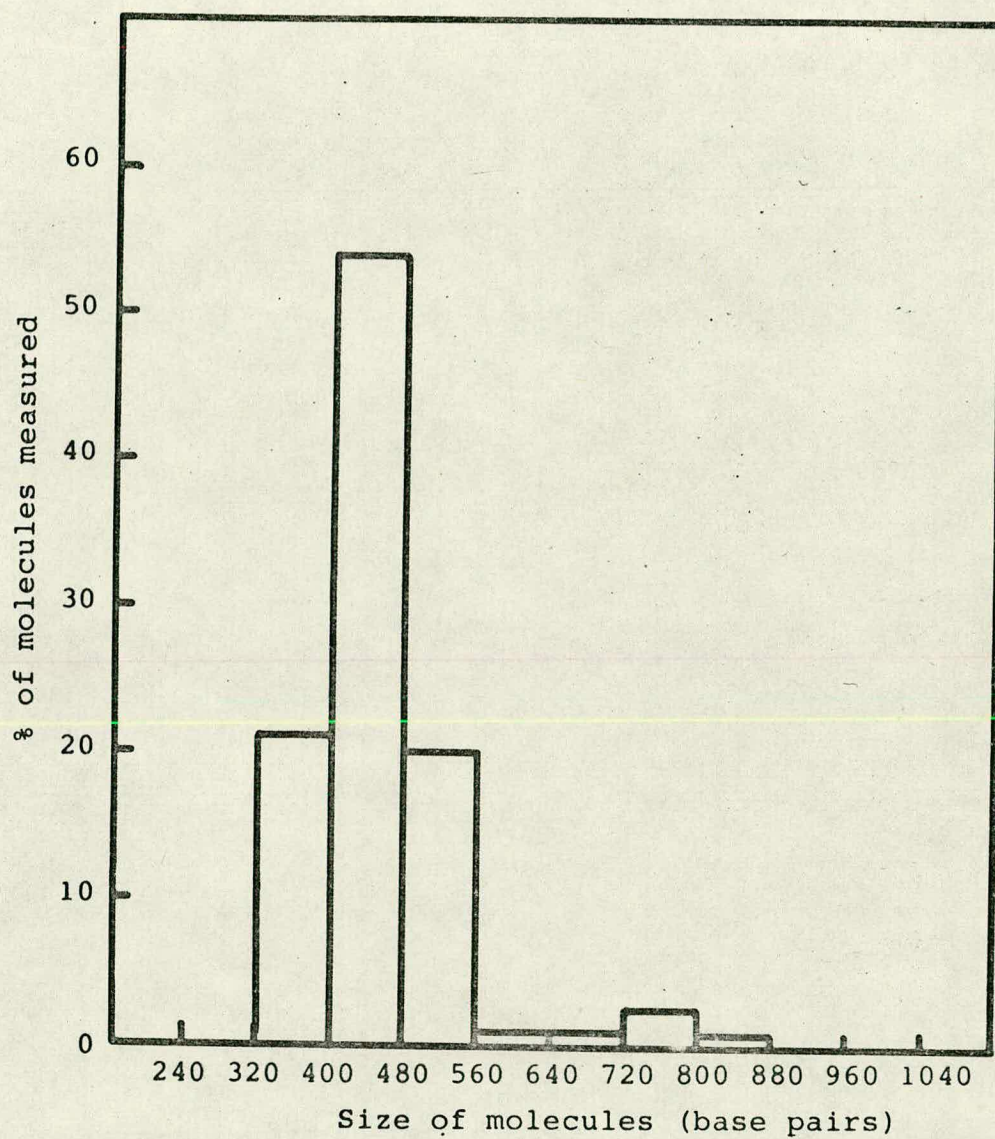


Fig. 23 Frequency/size histogram of *A. sylvaticus* band I DNA molecules. 100 molecules measured after aqueous spreading, latex spheres were used as a size standard.



formed the satellite, occurred compared to divergence of the species. Figure 24 shows the melting profile of DNA from A. sylvaticus sheared to about 500 nucleotides and renatured to a Cot of 7,000; the optical profile is that of the homologous reaction. A. flavicollis total  $^{125}\text{I}$  labelled DNA was present with a ratio of 1:1000 A. flavicollis : A. sylvaticus. The difference between the  $T_m$ s of the two profiles is  $5^\circ$ . Figure 25 shows a similar experiment in which the trace DNA was A. microps; here the difference between the  $T_m$ s was  $11^\circ$ .

From this data, the  $\Delta T_m$ s of the restriction fragments and the phylogenetic tree derived from the patterns of restriction bands (see below) a scheme can be drawn as in figure 26. Implicit in this is the assumption that mutation rates for all classes of DNA are constant in one genus. This may be so, but in measuring thermal stabilities one is measuring fixed mutations; if there is little selection for sequence in satellite DNA fixation of mutations will be more frequent than in protein coding or other sequences which have functional constraints. A  $\Delta T_m$  of  $12^\circ$  in satellite could represent a period of time which would produce only a smaller  $\Delta T_m$  in unique DNA. With respect to figure 26 this means that the points at which multiplications occurred can be moved relative to the branch points between the species.

#### Derivation of a phylogenetic tree from restriction patterns.

Plate 1 shows gel electrophoresis patterns of all the Apodemus total DNAs studied here after digestion with five different restriction enzymes using a partial digest of mouse satellite with Eco RII as a marker. There are a number of similarities between the patterns given by the different species. When digested with Eco RI A. sylvaticus, A. flavicollis and A. microps give a common pattern of four bands (see table 2 for size



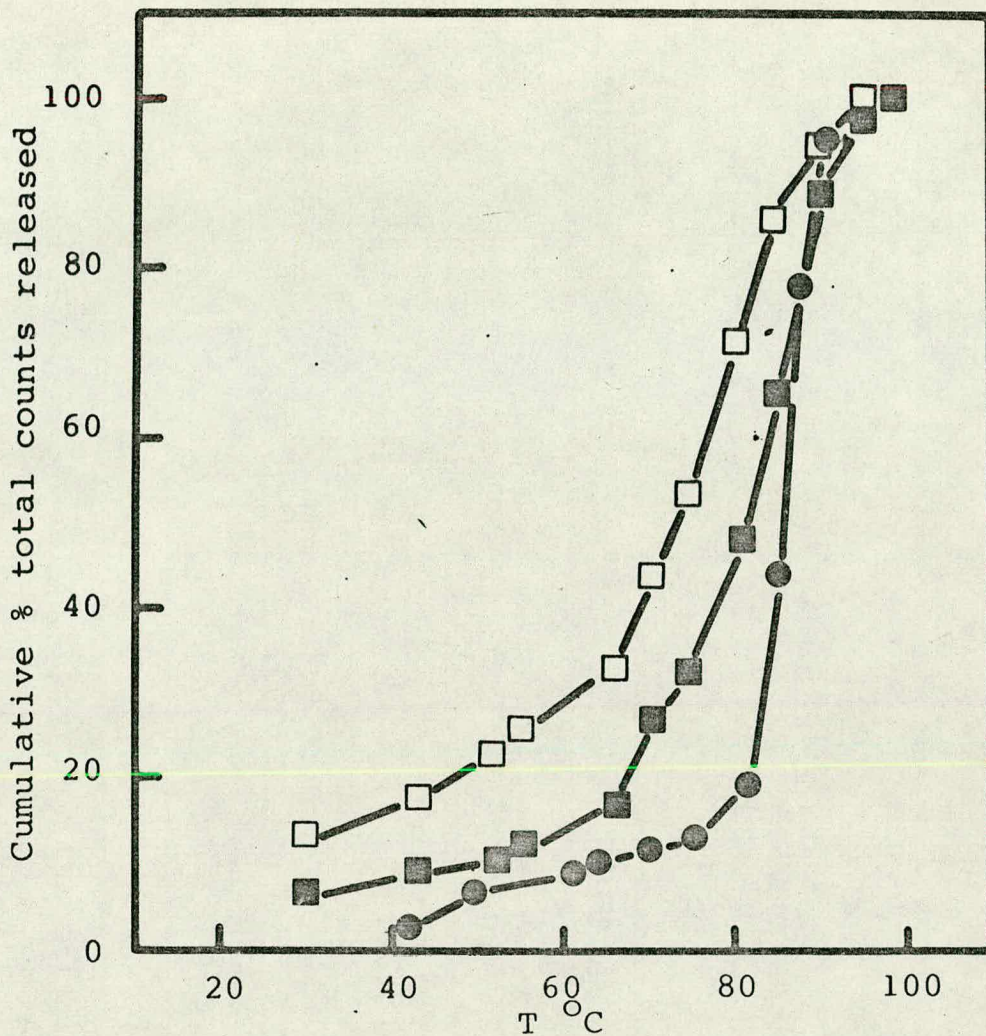


Fig. 24 Thermal stabilities of *A. sylvaticus* homoduplex and *A. sylvaticus/A. flavicollis* heteroduplex. 1 mg of *A. sylvaticus* total DNA and 1 microgramme of *A. flavicollis* I<sup>125</sup> total DNA was reassociated to a Cot of 7000 in 0.12M PB and a sample of the reaction mixture melted off HAP. —□— I<sup>125</sup> DNA released, —■— optical density (260nm), —●— native melt of *A. sylvaticus*. The extent of driver reaction was 94%. Self reassociation of the trace DNA at the same Cot was 5.7%



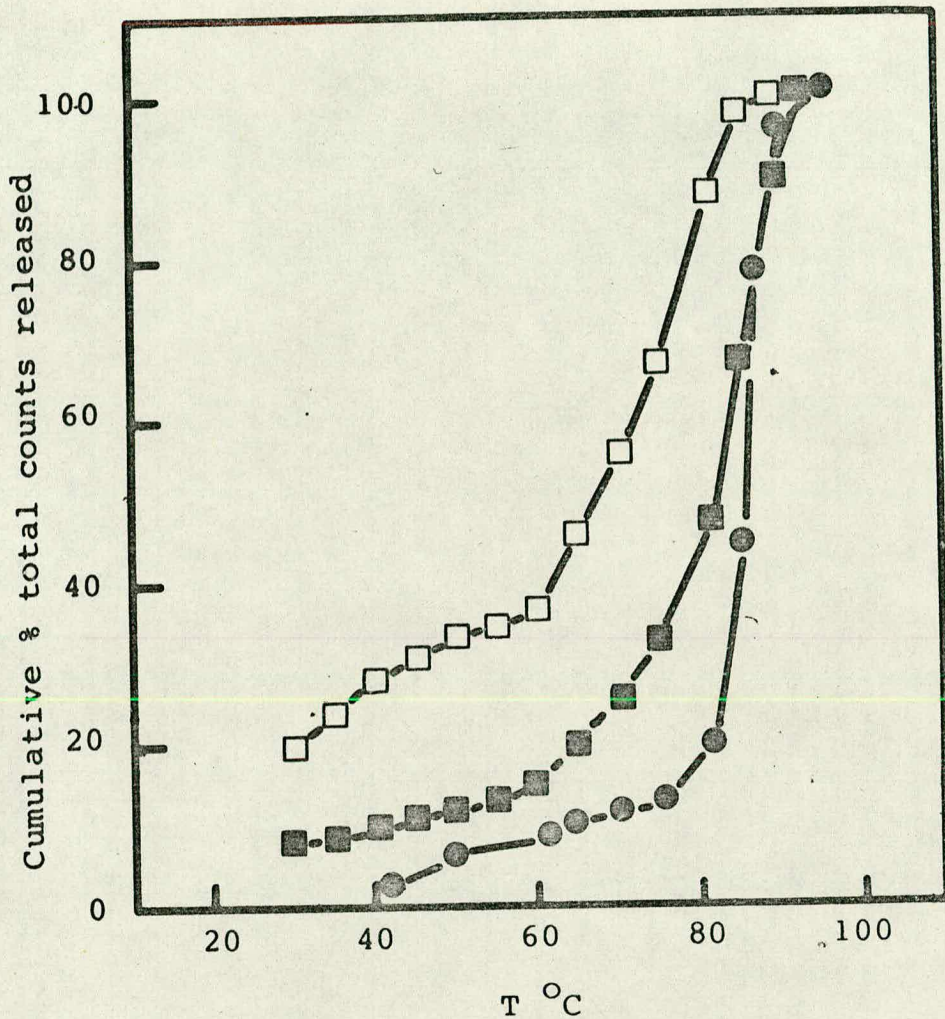


Fig. 25 Thermal stabilities of *A. sylvaticus* homoduplex and *A. sylvaticus/A. microps* heteroduplex. Legend as for fig 27 except  $\square-\square$  *A. microps* I<sup>125</sup> DNA. Driver DNA reassociation was 88%. Self reassociation of the trace DNA at the same Cot was 7.1%.



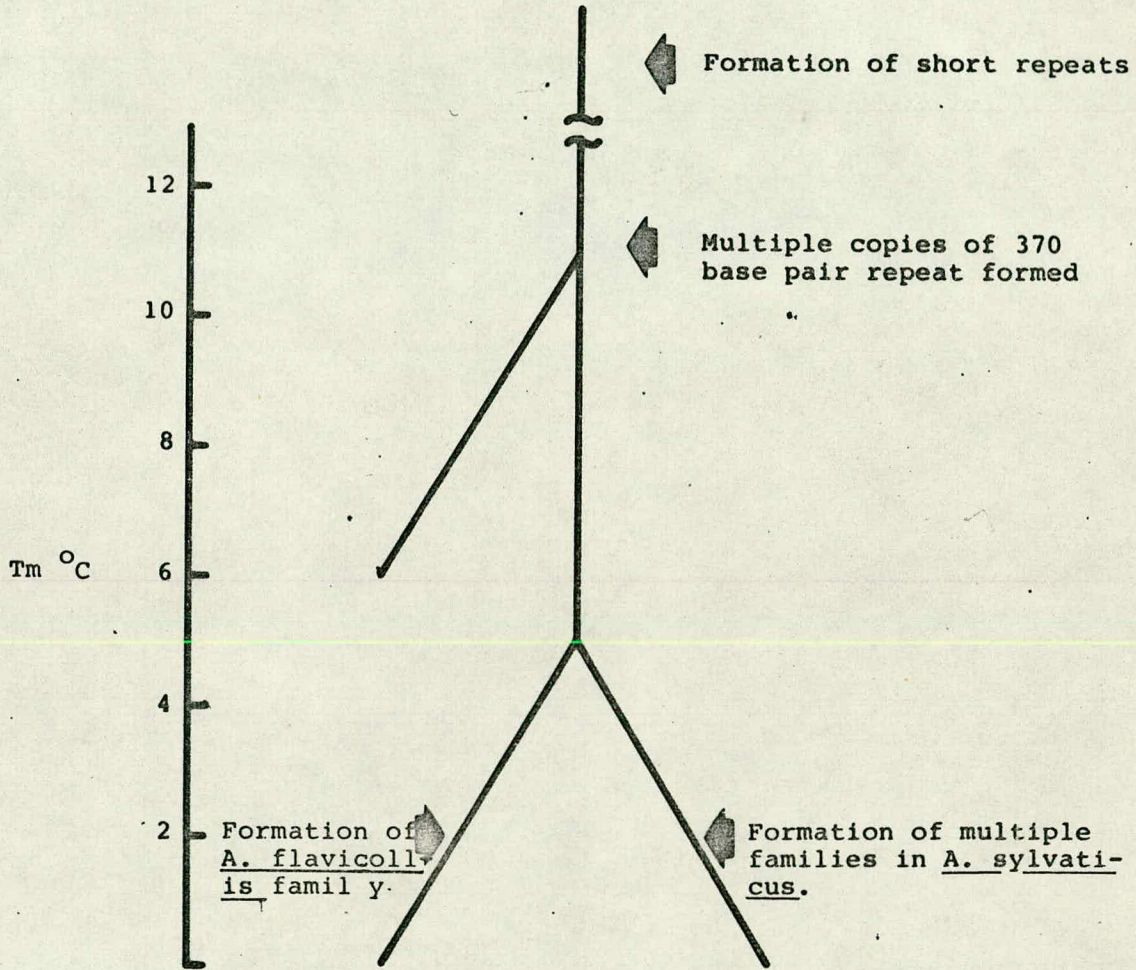


Fig. 26 Stages in the evolution of satellite DNA in the genus Apodemus



estimates), in addition A. agrarius and A. mystacinus clearly show the largest and smallest of these bands and A. mystacinus shows small quantities of the second smallest band.

The estimated sizes of these bands suggest that they represent four different repeated sequences since the 1850 and 1440 base pair bands are not simple multiples of the 500 and 430 base pair smaller bands. Partial digests gave no indication of dimers of these four bands implying that these repeated sequences are not tandemly arranged. They are not present in equimolar quantities (two are absent in A. agrarius) and hence are not parts of a longer repeated sequence.

Disregarding the mechanism of formation of these repeated sequences a family tree can be constructed for these members of the genus on the basis of similarities between the patterns of bands. From the Eco RI bands A. sylvaticus, A. flavicollis, and A. microps are more similar to each other than to A. mystacinus or A. agrarius. This is supported by the band patterns produced by Hin endonuclease which gives identical patterns for A. flavicollis and A. sylvaticus with a related pattern for A. microps. The positions of A. agrarius and A. microps in this tree are reversed in comparison to that which might be derived from CsCl density gradient centrifugation (figure 27).

This phylogenetic tree assumes that A. microps diverged before the formation of the 370 base pair repeat, but the divergence could also have occurred after the formation of the 370 base pair repeat. The presence of a 180 base pair repeat suggests two possible schemes; either the sequence was derived by the introduction of a restriction site into the 370 base pair repeat at a site close to the centre of the sequence or, this 180 base pair repeat was present in the common ancestor of the three species and multiplication took the basic 180 base pair repeat in A. microps but, in the case of



A. sylvaticus/flavicollis common ancestor multiplied a repeat or repeats which contained mutations at the restriction site giving the 370 base pair unit.

If the latter explanation is correct then there will be similarities between the two halves of the 370 base pair repeat which might give staggered duplex structures or concatenates when the 370 base pair length is reassociated. Similarities between the two halves of the 370 base pair unit might also be reflected in a kinetic complexity significantly less than this length.

Neither of these predictions is confirmed (see above) but the finding of  $1\frac{1}{2}$ mers in partial digests suggests that there may be a low level of similarity between the halves of the 370 base pair repeat.



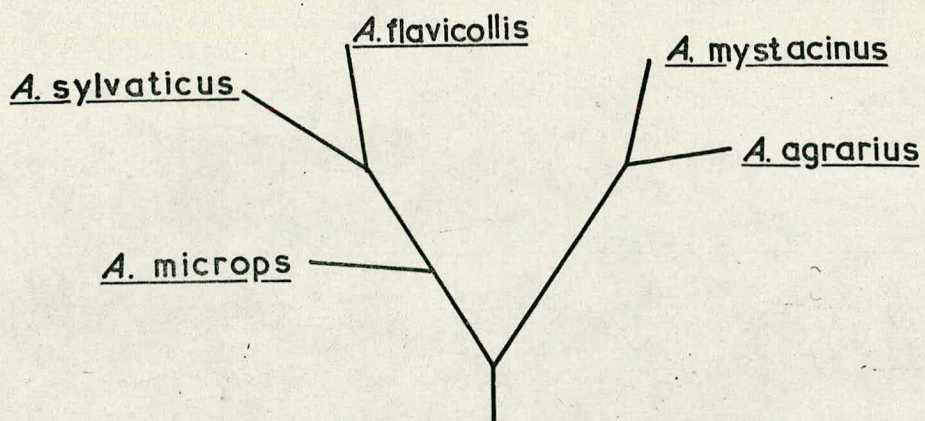


Fig.27 Phylogenetic tree for five members of the genus Apodemus based on similarities in the restriction patterns of the total DNAs.



CONCLUSION

A number of cross-reassociating families of sequences exist in the satellite DNA of A. sylvaticus which all have the same repeat length of around 370 base pairs. This 370 base pair periodicity is also found in two other members of the same genus. In one, A. flavicollis, only one family of sequences is found in the satellite DNA, in A. microps an earlier stage in the evolution of the sequences is observed in which sequences half the length of the 370 base pair periodicity are present in greater quantities than the 370 base pair repeat itself.

Multiple related satellite families have been found in Drosophila (Gall & Atherton 1974) and in a number of Mus species (Sutton & McCallum, 1972; Rice & Straus, 1973) but the use of restriction endonucleases in the case of A. sylvaticus, A. flavicollis and A. microps shows that within a single satellite there can be related families which maintain the restriction repeat length, not only between families but also between species.

Rice et al (1973) concluded that, since the divergence within the non repeated DNA between M. musculus and M. caroli or M. cervicolor is comparable to that between repeated DNAs in these species, the satellites originated at the time of speciation. This is not the case for the Apodemus species where the satellite sequences were present when A. sylvaticus and A. flavicollis diverged but were further multiplied in both A. sylvaticus and A. flavicollis after speciation.

Allan (1974) shows that the partial sequence analysis of A. sylvaticus and flavicollis satellite DNAs suggests a short ancestral sequence overlaid by a considerable amount of divergence. Slippage replication (Wells et al 1967) could lead to the formation of multiple, tandemly linked copies of this short sequence. The formation of a longer repeat could occur by a replication mechanism which took a large number of copies of the short



repeat and formed multiple copies of the sequence taken, a suitable candidate might be a rolling circle mechanism (see introduction). Alternatively divergence within a group of short sequences might make these sequences less homologous than a number of the sequences as a unit; unequal crossing over could then increase the similarities between the longer segments until this was the only recognisably repeating unit.

Smith (1973) has calculated relationships between the rate of crossing over and the rates of fixation of sequences towards a homogeneous set such as is found for the ribosomal spacer in Xenopus. He finds a linear relationship between the number of copies and the number of crossovers necessary for fixation. As the crossover frequency within the sets of short repeats is not known and also the number of repeats present at this stage is similarly unknown it is not possible to decide on the feasibility of this mechanism.

According to the model put forward here the 370 base pair repeat has survived two speciations and at least as many independent events forming different related families. Two types of mechanism can explain the stability of this repeat length. The first possibility is that an enzymic multiplication process, perhaps of the rolling circle type, takes a large number of copies of the 370 base pair repeat and by multiplication produces a family of the sequences within which divergence occurs. According to this hypothesis related families arise by taking different diverged segments of the original sequence for multiplication.

The second possibility is that the formation of a number of related families is not a multiplication event but is due to the processes of unequal crossing over. Smith (1973) points out that where there are a large number of copies of a sequence with either too high a mutation rate or too low a crossing over rate to maintain all the copies as a homogeneous set then unequal crossing over will form a number of families of repeated



sequences which are related to each other. An alternative form of this explanation is that in A. sylvaticus the families are prevented from crossing over in some way perhaps by being on different chromosomes.

Southern (1974) has argued that too many crossovers would be required to maintain the sequence homology found within the Mus satellite. It has been argued earlier that the frequency of crossing over is proportional to sequence homology, the A. sylvaticus band I DNA seems to be involved in crossing over to form  $\frac{1}{2}$ mers at a frequency of one crossover per 75 copies of the monomer even when there is little detectable sequence homology between the two halves of the molecule; crossing over with a 370 base pair periodicity would, perhaps, be much more frequent since this is the reassociation register of the restriction fragments.

The presence of the 370 base pair repeat length in three species does argue that there may be some functional reason for its retention; although it is not found in A. agrarius or mystacinus by restriction, Polito (personal communication) has found that there is a 370 base pair repeat in A. agrarius satellite. Sonicated A. agrarius satellite DNA forms circles which fall into size classes with this periodicity. There are two ways in which the periodicity might have been conserved without the restriction site; a) the A. agrarius satellite might have diverged from the ancestral sequence which contained the restriction site, b) the sequence multiplied up to form A. agrarius satellite, or formed by unequal crossing over, might not have contained the restriction site.

The presence of the repeat length but not the restriction sequence in A. agrarius implies that the important feature is perhaps the repetitiveness rather than the DNA sequence. This could be explained if tandem repeats were in some way responsible for a DNA structure with a functional significance.



If the importance of repeated DNA does lie in the occurrence of large numbers of the same repeat length then sequence conservation between copies of the same repeat length in related species may be irrelevant. Their function might not be as a recognition site for any universal process such as transcription or initiation of replication but rather for functions which are specific to the related genomes. Functions which fit this criteria might be maintenance of chromosomal structure, centromeric recognition or recognition between homologous chromosomes during pairing. In general these are functions which might involve repeat to repeat pairing and recognition between different DNA strands.



## REFERENCES

- Allan J. (1974) Doctoral thesis, University of Edinburgh.
- Arrighi, F. E. & Hsu, T.C. (1970) *Chromosoma* 32 224
- Bird, A. Rochaix, J. Bakken, A. *Molecular Cytogenetics*,  
Ed. Hamakalo, B. & Papaconstantinou, J.  
Plenum Press New York. (1973)
- Birnstiel, M.L., Spiers, J., Purdom, I., Jones, K. &  
Loening, U. (1968) *Nature* 219 454
- Biro, A. Carr-Brown, A., Southern E.M. & Walker P.M.B.,  
(1974) Manuscript submitted for publication.
- Blumenfeld, M. (1973) *Cold Spring Harbour Symp. Quant.  
Biol.* 38 423
- Bostock, C.J. (1972) *Adv. Cell Biol. Vol 2* 185 Ed.  
Prescott DM., Goldstein, L. & McConkey, E.  
New York, Appleton Century Crofts.
- Botchan, M., Kram, R., Schmid, C.W. & Hearst, J. (1971)  
*Proc. Natl. Acad. Sci. USA* 68 1125
- Botchan, M. & McKenna, G. (1973) *Cold Spring Harbour  
Symp. Quant. Biol.* 38 383
- Britten R.J., (1971) *Carnegie Inst. Wash. Year Book* 70  
369
- Britten, R.J. & Kohne, D.E. (1967) *Carnegie Inst. Wash.  
Year Book* 65 78
- Britten, R.J. & Kohne, D.E. (1968) *Science* 161 529
- Britten, R.J. & Davidson (1969) *Science* 165 349
- Britten, R.J. & Kohne, D.E. (1969) *Handbook of Molecular  
Cytology*. Ed. Lima de Faria A. North Holland.
- Britten, R.J. & Smith, J. (1970) *Carnegie Inst. Wash.  
Year Book* 68 378



- Brown, D.D. & Weber, C.S. (1968) *J. Molec. Biol.* 34 661
- Brown, D.D. & Dawid, I. (1968) *Science* 160 272
- Brunk, C.F. & Leick, V. (1968) *Biochem. Biophys. Acta* 179  
136
- Burton, K. & Peterson, G.B. (1960) *Biochem. J.* 75 17
- Commerford, S.L. (1971) *Biochemistry* 10 1993
- Danna, K.J., Sack, G.H. & Nathans, D. (1973) *J. Molec.*  
*Biol.* 78 363
- David, I., Brown, D.D. & Reeder, R.H. (1970) *J. Molec. Biol.*  
51 341
- Davidson, E.H. & Hough, B.R. (1968) *J. Molec. Biol.* 56 491
- Doerman, F. (1973) *Ann. Rev. Genetics.* 7 325
- Flamm, W.G., Bond, H.E. & Burr, H.E. (1966) *Biochem.*  
*Biophys. Acta* 129 310
- Flamm, W.G., Walker, P.M.B. & McCallum, M. *J. Molec. Biol*  
40 423
- Gall, J. & Atherton, D.D. (1974) *J. Molec. Biol.* 85 633
- Georgiev, G.P. (1969) *J. Theoretical Biol.* 25 473
- Gilbert, W. & Dresler, D. (1968) *Cold Spring Harbour*  
*Symp. Quant. Biol.* 33 473
- Hearst, J.E., Cech, T.R., Marx, K.A., Rosenfeld, A. &  
Allen, J.R. (1973) *Cold Spring Harbour Symp.*  
*Quant. Biol.* 38 329
- Hennig, W. (1970) *Chromosoma* 32 31
- Hennig, W. & Walker, PMBW (1970) *Nature* 225 915
- Hennig, W. (1972) *J. Molec. Biol.* 71 407
- Holliday (1964) *Genetics Research* 5 282



- Holliday (1968) *Reproduction and Recombination of Genetic Material*. Ed Peacock & Brunch  
C.A.A.S. 157
- Hourcade, D., Dressler, D. & Wolfson, J. (1973) *Cold Spring Harbour Symp. Quant. Biol.* 38 537
- Hutton & Wetmur, J.G. (1973) *Biochemistry* 12 3
- Jones, K.W. (1970) *Nature* 225 912
- Kimura, M. (1968) *Nature* 217 624
- King, J.L. & Jukes, T.H. (1969) *Science* 164 788
- Laird, C.D., McConnaughty, B.L. & McCarthy, B.J. (1970)  
*Nature* 224 149
- Laird, C.D. (1971) *Chromosoma* 32 378
- Lee, A.S. & Sinsheimer, R.L. (1974) *Proc Natl Acad Sci USA* 71 2882
- Lindsley & Novitski (1958) *Genetics* 43 790
- Loening, U.E. (1966) *Biochem. J.* 102 251
- Maio, J.J. & Schildkraut, C.L. (1969) *J. Molec. Biol* 40 203
- Maniatis, T. & Ptashne, M. (1973) *Nature* 246 133
- Marin & Prescott (1964) *J. Cell Biol.* 21 159
- Marmur, J. & Doty, P. (1962) *J. Molec. Biol* 5 109
- Martin, M.A. & Hoya, B.H. (1967) *J. Molec. Biol* 27 113
- Massie, H.R. & Zimm, B. (1965) *Proc. Natl. Acad. Sci. USA* 54 166
- Maurer, R., Maniatis, T. & Ptashne, M. (1974) *Nature* 249 221
- McCarthy, B.J. & Farquhar, M.N. (1972) *Evolution of Genetic Systems, Brookhaven Symp.* 23 1



- Mirsky, A.F. & Ris, H. (1951) *J. Gen. Physiol* 34 451
- Miyazawa, Y. & Thomas, C.A. Jr. (1965) *J. Molec. Biol.*  
11 223
- Old, R., Murray, K. & Roizes, G.P. (1974) manuscript  
submitted for publication.
- Pardue, M.L. & Gall, J.G. (1970) *Science* 168 1356
- Pardue, M.L. (1973) should be Weinberg, E.S., Birnstiel,  
M.L., Purdom, I.F. & Williamson, R. (1972)  
*Nature* 240 225
- Peacock, W.J. (1965) *Genetics* 51 573
- Rice, N. (1972) *Evolution in Genetic Systems*, Brookhaven  
*Symp.* 23 44
- Rice, N.R. & Straus, N.A. (1973) *Proc. Natl. Acad. Sci.*  
70 3546
- Rothfels, K.H. (1966) *Chromosoma* 20 54
- Sandler & Braver (1954) *Genetics* 39 365
- Schachat, F.H. & Hogness, D.S. (1973) *Cold Spring Harbour*  
*Symp. Quant. Biol.* 38 371
- Sharp, P.A., Sugden, B. & Sambrook, J. (1973) *Biochemistry*  
12 3055
- Slack, J.M.W. (1974) *Biopolymers* (in press)
- Smith, G.P. (1973) *Cold Spring Harbour Symp. Quant. Biol.*  
38 507
- Smith, H.O. & Wilcox, K.W. (1970) *J. Molec. Biol.* 51 379
- Sobell, H.M. (1973) *Advances in Genetics* 17 411
- Southern, E.M. (1970) *Nature* 227 794
- Southern, E. M. (1971) *Nature New Biol.* 232 82
- Southern, E.M. (1972) *Modern Aspects of Cytogenetics;*  
*Constitutive Heterochromatin in Man* 19



Schattauer Verlag, Stuttgart.

Southern, E.M. (1974) J. Molec Biol. in press

Southern, E.M. & Roizes, G.P. Cold Spring Harbour Symp.

Quant. Biol. 38 429

Sutton, WD & McCallum (1971) Nature New Biol. 232 83

Sutton, W.D. & McCallum (1972) J. Molec. Biol. 71 633

Sutton, W.D. (1972b) Nature New Biol. 237 70

Lee, C.S. & Thomas, C.A.Jr. (1973) J. Molec. Biol. 77 25

Pyeritz, R.E. & Thomas (1973) J. Molec. Biol. 77 57

Ullerich (1966) Chromosoma 18 316

Walen (1964) Genetics 49 905

Walker P.M.B. (1971) Prog. Biophys. Mol. Biol. 23 145

Walker P.M.B. (1971b) Nature 229 306

Walker P.M.B. & McClaren (1965) J. Molec. Biol. 12 394

Wang, A.C. & Kallenbach, N.R. (1971) J. Molec. Biol. 62

591

Waring, M. & Britten, R.J. (1966) Science 154 791

Wells, R.D., Ohtaku, E. & Khorana, H.G. (1965) J. Molec.

Biol. 14 221

Wetmur, J.G. & Davidson, E. (1968) J. Molec. Biol. 31 399

Wu, J.R., Hurn. A. & Bonner, J. (1972) J. Molec. Biol.

64 211

Yasmineh, W.G. & Yunis, J.J. (1970) Biochem. Biophys.

Res. Comm. 35 779

Yoshimuri, R.N. (1971) Doctoral Thesis, University of

California, San Fransisco.

Zehavi-Willner & Comb (1966) J. Molec. Biol. 16 250



This thesis has been composed by myself from my own work.