GENOME ORGANIZATION IN PLANTS WITH SPECIAL REFERENCE TO FOUR CUCURBITACEAE SPECIES

A THESIS

SUBMITTED TO THE

UNIVERSITY OF POONA

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN CHEMISTRY

BY

MRS. MRINAL RAVINDRA BHAVE
M. Sc.

DIVISION OF BIOCHEMISTRY
NATIONAL CHEMICAL LABORATORY
PUNE – 411 008 (INDIA)

1984

DEDICATED TO

MY PARENTS

DECLARATION

Certified that the work incorporated in the thesis "Genome organization in plants with special reference to four Cucurbitaceae species" submitted by Mrs. Mrinal Bhave was carried out by the candidate under my supervision. Such material as has been obtained from other sources has been duly acknowledged in this thesis.

P K Ranjekar

Dr. P.K. Ranjekar
Research Guide

CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

ABSTRACT

In the present work we have characterized nuclear genomes of four species of family Cucurbitaceae, namely, Luffa cylindria (sponge gourd), Luffa acutangula (ridge gourd), Benincasa hispida (ash gourd) and Coccinia indica (ivy gourd) with the following objectives:

i) To determine the content of repetitive DNA.

ii) To assess the nature of repetitive DNA sequences and to study their copy number, kinetic complexity and sequence divergence.

iii) To analyze the patterns of sequence organization and to evaluate the rates of turnover of these genomes.

iv) To determine the extent of homology of repeated DNA sequences.

Information about genome organization using several molecular approaches is extremely useful as it can help to answer basic questions such as mechanism of gene regulation, basis of chromosome evolution and phenotypic expression and functional role of different types of repetitive DNA sequences. It is also expected that such information will be useful in providing a molecular basis for the characteristic features of family Cucurbitaceae such as strong premating and incompatibility barriers leading to almost total species isolation.

The molecular studies on the four Cucurbitaceae genomes have been described in five main chapters and the results are summarized as follows:

1. High resolution thermal denaturation studies

The fine structure of DNA melting can be related to its

biological function and is useful in calculating its genetic and evolutionary origin. Derivative melting profiles have revealed species and family specific differences in plants. In the present investigation, the first derivative melting profiles of the nuclear genomes of the four Cucurbitaceae species have been obtained using a Gilford 250 spectrophotometer equipped with thermoprogrammer. The derivative melting profiles of ivy gourd and ash gourd are symmetrical, while those of ridge gourd and sponge gourd are skewed on the G + C rich side. Six melting components are common to all the four species. Fourteen out of the seventeen melting components are common in the two Luffa species, showing a great homology between them. The melting profiles of eukaryotic DNAs with many melting maxima are supposed to arise mainly because of the presence of high amounts of repetitive DNA in their genomes. Thus this approach gives an idea about the base composition and distribution of repetitive DNA families in Cucurbitaceae. Attempts are being made to draw a genetic distance map by comparing the areas under the derivative melting profiles.

2. Comparison of DNA reassociation kinetics

The reassociation kinetics studies of short DNA fragments (about 550 base pairs) were carried out in order to determine the number of DNA components reassociating with second order kinetics, the proportion of repeated

and single copy DNA sequences, and the frequency of repetition and kinetic complexity of different classes of DNA. The reassociation was monitored by optical method and hydroxyapatite column chromatography. The repeated DNA sequences account for 25% in ivy gourd, 48% in ash gourd, 51% in sponge gourd and 59% in ridge gourd. In ivy gourd, the repetitive DNA content of 25% is the lowest reported so far in all higher plant species. Out of the four species studied, only sponge gourd shows the presence of a fast reassociating component whose kinetic complexity and frequency of repetition are suggestive of a satellite fraction in its genome. The slow DNA component, with a reiteration frequency of a few hundred , comprises a major proportion of repetitive DNA in all the four species. Thus all the four species show about 2.5 fold variation in proportions of repetitive and single copy DNA sequences. Large differences are also noticed in proportions, copy number and kinetic complexity of different repetitive fractions. For example, though sponge gourd and ridge gourd belong to the same genus Luffa, the fast component with a reiteration frequency of 52,900 is absent in ridge gourd, while the kinetic complexity of single copy DNA of ridge gourd is twice as much as that of sponge gourd. It is suggested that these molecular variations may contribute to species isolation in family Cucurbitaceae, which otherwise is extremely conservative in its morphological and cytological features.

3. Reassociation kinetics studies at different criteria of stringency

Generally the reassociation kinetics of DNA are studied at a standard criterion of reassociation, i.e. Tm - 25 °C. At this temperature, only those sequences which have at least 25% homology can reassociate. However, it is likely that the genome has repetitive sequences which have more than 25% divergence (fossil repeats). Such sequences can be studied by relaxing the stringency of reassociation, e.g., reassociating the DNAs at a lower temperature (Tm - 30 °C to Tm - 35 °C). On the other hand, a higher incubation temperature can be employed to study the nature of the very well matched, conserved and perhaps functional repeats. In the present work, reassociation kinetics of repetitive DNAs have been studied optically at three different temperatures, ranging from about Tm - 30 °C to Tm - 5 °C, in addition to standard criterion (Tm - 25 °C). In all the species the repetitive DNA fraction reduces at a higher temperature of incubation. The sequences reassociating at high stringency conditions are shorter, have a higher copy number and a higher thermal stability than those reassociating at more relaxed conditions of reassociation. The effect of incubation temperature on the rate of reassociation indicates that the repetitive DNAs consist essentially of homogeneous

families. This view is also supported by the derivative melting profiles of repetitive DNAs isolated at different temperatures. Ivy gourd, ridge gourd and ash gourd contain more 'true' single copy DNA and less 'fossil repeats' than sponge gourd. Thus the genome of sponge gourd exhibits higher rate of turnover of repeated and single copy sequences than that of ridge gourd, ivy gourd and ash gourd.

4. Organization of repetitive and single copy DNA sequences

Repetitive sequences interspersed with single copy DNA sequences might function as regulatory elements for coordinating the gene activity. Hence it is essential to study the sequence organization in plant genomes to understand the role of repetitive sequences, as well as the amplification and translocation events responsible for origin and evolution of repetitive DNA. Several experimental approaches have been used to determine the interspersion patterns in these four species. The single strand specific enzyme S1 nucleasehas been used to isolate pure repetitive duplexes. The S1 nuclease resistant duplexes comprise 20%, 38%, 39% and 28% of the genome in ivy gourd, sponge gourd, ridge gourd and ash gourd respectively. The Agarose A50 gel filtration profiles of these pure duplexes show two size classes; a minor one, about 2000-4000 nucleotide pairs in length comprising

15-20% of the repetitive DNA, and a major one (80-85% of repetitive DNA) with an average size of 300 nucleotide pairs. Extensive interspersion of highly repetitive (very fast and fast reassociating) sequences with intermediately repetitive (slow reassociating) sequences is observed in sponge gourd and ridge gourd, while it is less apparent in ivy gourd and ash gourd. The length and proportion of interspersed single copy sequences have been determined by studying the fraction of repetitive DNA at different fragment lengths. The interspersed single copy sequences are about 2000 nucleotide pairs long in all four species. Most of the single copy DNA of sponge gourd and ridge gourd is involved in interspersion with repeated sequences, while a larger proportion of single copy fraction in ivy gourd and ash gourd is not interspersed with repetitive sequences even at a length of 7400 base pairs. The short interspersed repeat sequences show higher base sequence divergence than the long repetitive sequences.

5. DNA-DNA hybridization

The technique of DNA-DNA hybridization was used to determine the extent of sequence similarity in the repetitive DNAs. Sponge gourd total DNA and total repetitive DNA was labelled by nick-translation and hybridized with ridge gourd, ash gourd and ivy gourd DNAs to the respective

repetitive Cot value. Sponge gourd and ridge gourd repetitive DNAs show a great extent of sequence homology (85-95%) while repetitive sequences of ash gourd and ivy gourd show limited homologies (25-50%). The sequence divergence between sponge gourd and ridge gourd repetitive DNAs is also limited (0.65 - 1.27%) while that for ash gourd and ivy gourd repetitive DNA is extensive (3.6 - 8.00%). These results suggest that sponge gourd and ridge gourd have separated from each other relatively recently. On the other hand, ash gourd and ivy gourd DNAs seem to have undergone extensive divergence from the genus Luffa (sponge gourd and ridge gourd) since their time of origin.

CHAPTER I

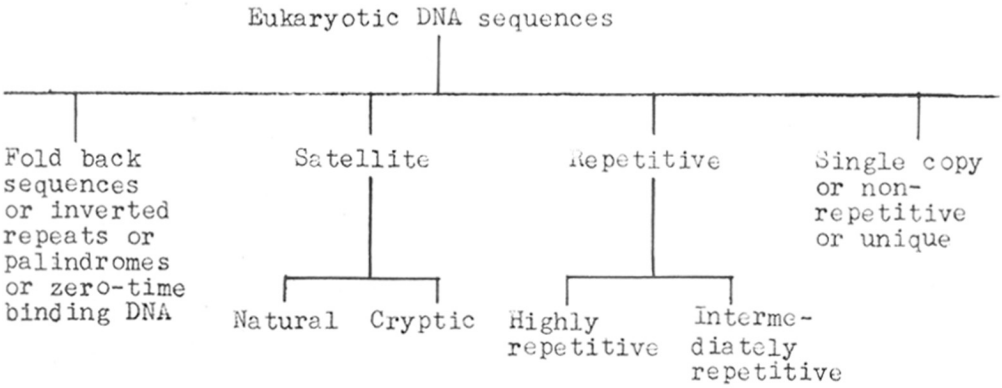
GENERAL INTRODUCTION

CURRENT CONCEPTS OF GENOME ORGANIZATION IN EUKARYOTES

DIFFERENT TYPES OF DNA SEQUENCES

Most of our current concepts of structure and organization of eukaryotic DNAs are based on DNA reassociation kinetics experiments and equilibrium centrifugation of DNAs in CsCl and Cs_2SO_4 gradients. Based on these studies, the nuclear genomes of eukaryotes have been shown to consist of mainly four types of DNA sequences as shown in Figure 1.1.

Figure 1.1



Inverted repeated sequences: These sequences generally represent 1-5% of the total genome (1-19, Table 1.1) and are usually dispersed with other DNA sequences. They can be observed as hairpin structures or bubbles under electron microscope (3,20,21). Ribosomal RNA as well as heterogeneous

Table 1.1 Reassociation kinetics studies of higher plants

DNA components of the plants	Fraction of genome	Copies per 1C genome	Kinetic complexity (base pairs)	Reference
Dicotyledons				
Arachis hypogea				
Highly repeated	0.12	38,000	2.9×10^5	1
Intermediately repeated	0.15	6,700	9.6×10^6	
Rare repeats	0.37	200	2.0×10^9	
Single copy	0.36	1	-	
Brassica nigra				
Very fast	0.03	-	-	2
Fast	0.25	22,640	5.1×10^3	
Slow	0.29	587	2.2×10^5	
Single copy	0.47	1	2.2×10^8	
Glycine max				
Very fast	0.05	-	-	
Fast	0.28	3,850	9.4×10^4	3
Slow	0.24	153	2.0×10^6	
Single copy	0.40	1	5.6×10^8	
Gossypium hirsutum				
Highly repeated	0.08	-	3.8×10^6	4
Repeated	0.27	134	9.7×10^8	
Single copy	0.61	1	2.9×10^{11}	

Table 1.1 (contd.)

Linum usitatissimum				
Zero time binding	0.03	-	-	5
Satellite	0.11	-	-	
Fast repeats	0.29	-	-	
Intermediate repeats	0.15	-	-	
Single copy	0.35	1	3.5×10^8	
Matthiola incana				
Very fast	0.07	-	-	2
Highly repetitive	0.20	973,000	2.8×10^2	
Slow	0.42	192	3.0×10^6	
Single copy	0.31	1	4.8×10^8	
Nicotiana tabacum				
Very fast	0.02	-	-	6
Fast	0.06	12,400	4.6×10^3	
Intermediate	0.59	252	2.5×10^6	
Single copy	0.22	1	2.8×10^8	
Petroselinum sativum				
Very fast	0.05	-	-	7
Highly repetitive	0.13	136,000	1.8×10^3	
Fast	0.48	3,000	3.0×10^5	
Slow	0.16	42	7.4×10^6	
Single copy	0.12	1	2.2×10^8	

Table 1.1 (contd.)

Phaseolus mungo				
Rapidly reassociating	0.06	-	-	8
Intermediately reassociating	0.29	833	2.8×10^5	
Slow	0.66	1	5.3×10^8	
Phaseolus aconitifolius				
Rapidly reassociating	0.05	-	-	
Intermediately reassociating	0.36	857	1.0×10^6	8
Slow	0.59	1	1.4×10^9	
Pisum sativum				
Very fast	0.04	-	-	9
Fast	0.46	10,000	2.1×10^5	
Slow	0.33	3,000	5.0×10^6	
Single copy	0.21	1	4.5×10^9	
Vigna radiata				
Very fast	0.02	-	-	10
Fast	0.05	10,000	2.8×10^3	
Slow	0.28	175	9.0×10^5	
Single copy	0.54	1	4.7×10^8	

Table 1.1 (contd.)

Monocotyledons				
Allium cepa				
Very fast	0.07	-	-	11
Fast	0.41	21,600	2.9×10^5	
Slow	0.36	225	2.4×10^7	
Single copy	0.06	1	9.3×10^8	
Echinochloa frumentacea				
Very fast	0.05	-	-	12
Fast	0.12	205,000	8.8×10^2	
Slow	0.24	288	1.2×10^6	
Single copy	0.47	1	6.8×10^8	
Eleusine coracana				
Very rapidly reassociating	0.17	-	-	13,14
Rapidly reassociating	0.07	1,500	1.7×10^4	
Intermediately reassociating	0.25	140	7.0×10^5	
Single copy	0.42	1	1.6×10^8	
Hordeum vulgare				
Very rapidly reassociating	0.08	-	-	15
Rapidly reassociating	0.34	-	-	
Slow repeats	0.30	-	-	
Single copy	0.28	-	-	

Table 1.1 (contd.)

Oryza sativa				
Very rapidly reassociating	0.09	200,000	1.0×10^3	13,14
Rapidly reassociating	0.19	12,000	3.9×10^4	
Intermediately reassociating	0.25	190	3.2×10^6	
Single copy	0.48	1	1.2×10^9	
Panicum miliare				
Very fast	0.18	-	-	12
Slow	0.08	240	2.8×10^4	
Single copy	0.46	1	3.8×10^7	
Pennisetum americanum				
Rapidly renaturing	0.23	-	-	16
Middle repetitive	0.47	298	3.1×10^5	
Single copy	0.17	1	3.3×10^7	
Secale cereale				
Highly repeated (I)	0.10	6×10^5 to 6×10^6	1.7×10^2 to 1.7×10^3	17
Highly repeated (II)	0.33	37,500	2.7×10^4	
Intermediately repeated	0.31	850	1.7×10^6	
Single copy	0.27	1	1.2×10^9	

Table 1.1 (contd.)

Setaria italica				
Very fast	0.18	-	-	12
Slow	0.08	240	2.8×10^4	
Single copy	0.46	1	3.8×10^7	
Triticum aestivum				
Rapidly reannealing	0.04-0.10	-	-	18
Intermediately reannealing	0.83	4,300	-	
Slow reannealing	0.12	1	-	
Zea mays				
Highly repeated	0.42	79,333	3.6×10^4	19
Middle repeated	0.17	1,133	2.5×10^6	
Non-repeated	0.30	1	2.8×10^9	

nuclear RNA from mammals have also been shown to contain a low percentage of foldback sequences (22-24). Palindromic sequences flank most of the transposable elements and are probably involved in transposition of genetic material (25). They are supposed to function as binding sites for regulatory proteins (26,27) and cleavage sites for restriction enzymes (28). These sequences are also assumed to be involved in the mechanism of gene expression or in the processing of primary transcription products (29,30).

Satellite DNA sequences: Many of the very highly repeated DNA sequences are present in long tandom arrays and appear as 'satellites' after equilibrium density gradient centrifugation of DNA in neutral CsCl (patent satellites, 31) or in Cs_2SO_4 containing heavy metals such as Ag^{+} or Hg^{++} (cryptic satellites). Both natural and cryptic satellites have been shown to occur in many plant species (Table 1.2, 32-55).

In plants, the satellites are particularly most abundant in the species of Cucurbitaceae and Rutaceae. The largest amount of satellite DNA is present in the genome of Cucumber (35). The plant satellite DNAs are often heterogeneous in base composition and can be separated into subfractions differing in their melting properties, reassociation rates and degrees of methylation e.g. musk melon (56-58), Brassica nigra (59), wheat

and barley (51) and plants of Citrinae (60). The cryptic satellite of Sinapis alba is especially peculiar in lacking palindromes or very short repeats and methylation and thus is not similar to any of the known satellite sequences (53). The basic repeating unit of satellite DNA is longer in plants as compared to that in animals (47,58,61,62).

In situ localization of satellite DNAs in plant chromosomes has shown that these sequences are generally distributed in the heterochromatic regions of chromosomes, chromocentres or nucleolar organizer regions (48,50,63). In rye, the highly repetitive DNA (Cot 0.02 M x sec/1) was isolated and shown to be localized mainly within C-bands near the distal ends of most chromosomal arms (52). The elucidation of the relationship between satellite DNA and heterochromatin permits some conjecture concerning the origin and function of this significant portion of the genome of higher organisms (64-67). A role has been suggested for satellite sequences in chromosome condensation, homologous chromosome recognition and pairing, in providing specific attachment sites for spindle fibres and nuclear membrane and in maintaining proper association between particular regions of non-homologous chromosomes during interphase. The additional functions such as genetic regulators, spacers and

Table 1.2 Satellite DNA sequences in higher plants

Species	Satellite DNA % of total genome	Reference
NATURAL SATELLITES		
DICOTYLEDONS		
Ammi visnaga	-	33
Aquelagia alpina	-	34
Brassica rapa	21	35
Brassica pekinensis	17	35
Brassica 9 sp.	37	36
Bryonia dioica	5	35
Calendula officinalis	-	35
Citrullus vulgaris	3	35
Citrus (4 sp.)	19-23	35
Cucumis melo	30	35,37
Cucumis sativus	20-44	35,38,39
Cucurbita pepo	16-18	35
Daucas carota	-	33,40
Drimys piperata	7	35
Hammamelis mollis	6	35
Fortunella sp.	24	35
Gossypium sp.	-	41
Lagenaria vulgaris	9	35

Table 1.2 (contd.)

Linum usitatissimum	15	35,5
Lobularia maritima	32	35
Lupinus angustifolius	40	42
Luffa cylindrica	6	35
Lycopersicon esculantum	6	43
Oenothera fruticosa	-	34
Phaseolus aconitifolius	-	44
Phaseolus aureus	5	35
Phaseolus coccineus	24	35
Phaseolus mungo	-	44
Phaseolus vulgaris	19	35,45,46
Raphanus sativus	10	38
Sinapis alba	-	47
Solanum crispum	6	35
Solanum tuberosum	4	35
Tropaeolum majus	-	48
MONOCOTYLEDONS		
Cymbidium hybrid	15	49
CRYPTIC SATELLITES		
DICOTYLEDONS		
Citrus paradasi	-	32
Cucumis melo	-	32
Cucumis sativus	-	32

Table 1.2 (contd.)

Cucurbita pepo	-	32
Linum usitatissimum	-	32
Vicia faba	-	32,50
MONOCOTYLEDONS		
Avena sativa	-	32
Hordeum vulgare	8	32,51
Scilla campanulata	-	32
Scilla sibirica	19	32,50
Secale cereale	-	52
Sinapis alba	12	53
Triticum aestivum	12	51,54,55

mutation - accumulating loci have also been attributed to satellite DNA (65).

In many instances, differential replication of satellite DNA sequences has been observed under certain physiological conditions, and thus satellite sequences could be involved in response to some physiological variations (68-70). However, most of these studies are correlative, and closely related species often differ with respect to the content of satellite or highly repeated DNA sequences (Table 1.2). Satellite DNA sequences are considered to be very sensitive indicators of evolutionary divergence. It is suggested that they may be involved in establishing fertility barriers during speciation (64,66).

Repeated DNA sequences: Most of the higher eukaryotic genomes are composed of different classes of repeated DNA sequences differing in proportion, kinetic complexity and frequency of reiteration (Tables 1.1, 1.3) (1-19,71-93).

The highly repeated DNA sequences occur generally as long tandom arrays, are repeated 10^3-10^6 times, and have a kinetic complexity of 10^3-10^5 base pairs. The moderately repeated DNA sequences occur about 10^1-10^3 times in the genomes and have kinetic complexities in the range of 10^5-10^6 base pairs. These sequences are

generally dispersed throughout the genome and appear to be more conserved than the highly repeated DNA sequences. Thermal denaturation studies have indicated that wide variations exist in the degree of sequence divergence in repetitive DNAs of different plant species. In rye, wheat, barley, tobacco, finger millet and maize, for example, upto 15% base mismatch is observed in repetitive DNA duplexes (6,13,15,17-19). In rice, on the other hand, the base mismatching is only 1% and is the lowest reported so far (13). Phaseolus species show an intermediate degree of base mismatch (8,46).

Many functional roles have been assigned to repeated DNA sequences. Most of the early suggestions were based on cytological observations. A clear evidence for the functional involvement of repeated DNA sequences in meiosis comes from the studies on crossing over in Lilium microsporocytes, wherein it is observed that a fraction of moderately repeated DNA sequences is synthesized during repair synthesis in pachytene (94,95). In addition to the roles in chromosome mechanics, certain fractions of repeats have been assigned some other functions which will be described in the following sections. However, significance of the much larger fraction of repetitive DNA remains, more or less, unknown or speculative.

Table 1.3 Repeated DNA sequences in higher plants

Species	Repetitive DNA % of total genome	Reference
DICOTYLEDONS		
Anacyclus sp.	70.0 - 89.0	71
Anemone sp.	53.0 - 67.0	72
Anthemis sp.	63.0 - 67.0	71
Arachis hypogea	64.0	1
Beans	40-70	73-75
Brassica pekinensis	57.0	2
Cinnamomum camphora	37.3	76
Cucumis sativus	55.0	77
Decaisnea fargesi	48.5	76
Glycine max	60.0	78,3
Gossypium hirsutum	40.0 - 50.0	4,79
Helianthus annus	69.0	80
Helianthus tuberosus	45.0	81
Lathyrus sp.	56.0 - 72.0	82
Linum usitatissimum	56.0	5
Liriodendron tulipifera	47.5	76
Magnolia soulargiana	39.5	76
Matthiola incana	69.0	2
Nicotiana tabaccum	70.0	6
Petroselinum sativum	50-70	7

Table 1.3 (contd.)

Phaseolus 4 sp.	35-47	8,46
Pisum sativum	65.0 - 75.0	9,83-85
Raphanus sativus	45.0	77
Tropaeolum majus	70.0	48
Vicia 3 sp.	70-90	86
Vigna radiata	35.0	10
MONOCOTYLEDONS		
Allium 7 sp.	44.0 - 60.0	87
Allium 3 sp.	83.0 - 86.0	86
Daffodil lily	50.0	88
Eleusine coracana	49.0	13
Hedera helix	44.0 - 56.0	89
Lilium henryi	60.0	90
Orchids 4 sp.	45.0	63
Oryza sativa	52.0	13
Pennisetum americanum	69.0	16
	54.0	13
Poa annua	87.0	91
Poa trivialis	82.0	91
Scilla sibirica	40.0	92
Wheat, oat rye, barley	70.0 - 83.0	15,17,18,93
Zea mays	78.0	91
	60.0	19

Dispersed repeat families. Discovery of restriction endonucleases and development of molecular cloning and DNA sequencing techniques have revolutionized the study of structure of complex eukaryotic genomes. The repeated DNA sequences and their genomic organization are found to be much more complex than previously suspected. Digestion of DNA with a specific restriction endonuclease divides it into a set of fragments whose size depends upon the spacing between the recognition sites of the enzyme. The mixture of sizes thus produced can then be separated by agarose (or polyacrylamide) gel electrophoresis. This technique has led to the identification of two different classes of highly repeated DNA sequences in mammals. Dispersed families with unit lengths under 500 base pairs are found in hundreds of thousands of copies, and are termed SINES (short interspersed repeated sequences). The other class, composed of repeat units several kilo base pairs in length, occurring about 10^4 (or fewer) times in the genome, is termed LINES (long interspersed repeated sequences) (96-115, Table 1.4).

Because of the almost ubiquitous occurrence of LINES and/or SINES in the mammalian genomes, they have been suggested to have various functions. For example, the 4.5 S and 7S small nuclear RNAs of mouse and human cultured cells are considered as Alu member transcripts (96).

Table 1.4 Dispersed repeat families in Mammals

Type of family	Organism	Copy number	Length of sequences (base pairs)	References
Alu I	Human	100,000-500,000	300	96-98
Alu-equivalent	Galago	-	-	98
	Lemur	-	-	98
	New world monkey	-	300	99
	Chinese hamster	-	140,96	96,98
	Mouse	40,000-80,000	130,190	96
	Rat	-	-	96
C family	Rabbit	-	-	100
	Chicken	-	-	96
R family	Mouse	100,000	475	101,102
Hinf	Human	50-100	319	103
Kpn I	Primates	$6x10^4 - 10^5$	1200,1500	104-106
			1800,1900	107
EcoRI	Human	-	340	108
MIF	Mouse	50,000	1300	109-111
	Apodemus	-	-	109-111
	Rat	-		109-111
	Chinese hamster	-		109-111

Table 1.4 (contd.)

EcoRII	Mouse	1500, 1700	112
Xba I	Human	850, 1300, 1900	105
	Gibbon	-	105
	Great ape	-	105
	Old world monkey	1300, 850	105
Hind III	Human	1900	113, 114
Bam HI	Mouse	5600	115

Kpn I sequences are also transcribed in HeLa cells by RNA polymerase III into abundant heterogeneous species of RNA (106). Alu I and Kpn I members are often interspersed with structural genes and could be involved in recombination and gene conversion events (104,11
Some of the other suggested functions of these sequences are replication origins (117), involvement in genome reorganization, DNA replication and pre-mRNA processing (118). These sequences are also assumed to have role in definition of discrete chromosomal and nuclear domains that are distinct from those occupied by the predominant centromeric repeats (113).

Insertion sequences or mobile genetic elements: Many genetic systems under the control of mobile sequence elements are established todate, both in pro- and eukaryotes. The plant mobile genetic elements are listed in Table 1.5 (119-131). The most well-studied transposons in higher plants are those in maize, e.g. (i) the Ds-Ac insertion elements affecting the 'shrunken' gene that encodes the enzyme sucrose synthetase and thus leading to mutants with collapsed endosperm (119-122), (ii) the Ds-Ac system active at 'waxy' loci that determines amylose content of pollen and endosperm tissue (123-125), and (iii) the En-I (126) and Ds-Mp (127) systems causing instability at the loci which are important in anthocyanin

synthesis. The mutant waxy loci have been cloned to isolate Ac and Ds elements (125). Ds element is found to be 4.1 and 2.0 kilo base pairs in length, while Ac element is 4.3 kilo base pairs long. The 4.1 kilo base pairs long Ds element is highly homologous to Ac element, while the 2.0 kilo base long Ds is homologous to the ends of Ac. Sequences homologous to ends of Ac are present in many copies in the maize genome, while those homologous to centre of Ac are very few. The Ac element at waxy locus is structurally different from other genomic Ac-like sequences, as shown by restriction enzyme digestions.

In majority of instances, the mobile sequences lead to inactivation of genes. However, these sequences are also known to mobilize genes and activate silent genes lacking normal promoters (132-134). Moreover, they may also be involved in certain developmental pathways (135). The intervening sequences in eukaryotic split genes are generally flanked by short imperfect direct repeats (136). It is suggested that the fluidity of DNA rearrangement resulting from the frequent transposition events can offer considerable scope for adaptations (135,137). Mobile elements can cross barriers which normally separate cells, organisms or species. Phage Mu, retroviruses and

Table 1.5 Mobile DNA sequences in higher plants

System	Plant	Gene affected	Reference
Ds-Ac	Maize	Shrunken gene coding for sucrose synthetase, waxy locus determining amylose content of endosperm.	119-125
En-I	Maize	Loci involved in anthocyanin systesis	126
Ds-Mp	Maize	Loci involved in anthocyanin systesis	127
U_q	Wheat	Strains infected with Wheat Streak Mosoic Virus	128
Cin 1	Maize	-	129
3.4 Kb unit	Soybean	L1 lectin gene	130,131

T-DNA of Agrobacterium tumefaciens are interspecific and even intergeneric vectors or agents for genome reorganization (138). In addition a single mobile element in Drosophila (*P* element) can behave differently in different cell types (138). Flavell et al. (139) suggest that the movement of short pieces of DNA to new sites is a relatively common phenomenon in plant chromosome evolution.

Multigene families: Although most structural genes are single copy sequences, a variety of multigene families are known in which the gene sequences are present in enough copies to be included with the moderately repetitive DNA. The examples of plant multigene families are given in Table 1.6 (140-155). Many multigene families such as ribosomal RNA genes are organized in arrays of tandemly repeating units which include a large amount of spacer DNA and sometimes noncoding intervening sequences within the genes themselves. The function of spacer and insertion sequences in such units remains unknown. It is suggested that they play a role in the operation of correction mechanisms (such as unequal crossing over) required for the parallel evolution of tandem genes (156).

In addition to tandemly repeated genes, recent evidence suggests that copies of some repeated genes can exist at many different loci e.g. the genes for the

Table 1.6 Multigene families in plants

Multigene family	Species	Number of genes	Reference
Ribosomal DNA	Rice	850	140
"	Varieties of *Triticum durum*	2308-6457	141
"	Microseridineae	1100-3200	142
"	*Allium cepa*	5520-11870	143
Zein	Maize	100	144-146
B1 hordein	Barley	-	147
Conglycinin	Soybean	5-20	148
Storage protein	Cotton	-	149 150
Phaseolin	*Phaseolus vulgaris*	14	151
Actin	Soybean	10	152
Leghaemoglobin	Soybean	40	153
Ribulose bisphosphate carboxylase	Wheat	-	154
"	Soybean	-	155

maize storage protein zein are distributed on different regions of chromosome 4 and 7 (157).

Single copy DNA: The most slowly reassociating DNA sequences in eukaryotic genomes are usually assumed to be single copy sequences. There are some plant species where truly unique DNA sequences are not present, e.g. the genome of Vicia faba contains 60 identical copies of the so-called unique DNA sequences whereas V.sativa has four identical copies (75). In plants, the proportion of single copy DNA varies in the range of 10-65% (Table 1.1).

The unique or single copy DNA sequences are considered to include the sites of classical genetic activity, i.e., the structural gene loci, such as fibroin (158), globin (159,160) and ovalbumin (161). However, usually only a very small fraction of unique DNA (1-10%) is found to be transcribed into mRNA (162,163). Nuclear RNA typically contains 4-10 times excess transcripts, but most of them never leave the nucleus and, therefore, do not code for proteins (164,165). Thus the amount of unique DNA does not correlate with the biological complexity of the organism, any more than the total DNA content does.

GENOME ORGANIZATION IN EUKARYOTES : Satellite DNA and very highly repetitive DNA generally occur as tandemly linked simple sequences. On the other hand, members of the intermediately repetitive class are located in all the regions of the genome. Britten and Kohne (166) were the first to demonstrate the interspersion of intermediately repetitive and unique DNA sequences. Two general types of interspersion patterns are discerned. The Xenopus pattern or 'short period' interspersion pattern is characterized by repeat units about 300 base pairs in length, interspersed with single copy sequences about 800 base pairs long (167). The Drosophila pattern or 'long period' interspersion pattern is characterized by families of repeat units several kilo base pairs in length separated by tens of kilo base pairs of single copy sequences (168). The studies in Xenopus (167), sea urchin (169), an insect (170), fishes (171), mammals (172-174) and a fungus (175) demonstrated a widespread occurrence of short period interspersion pattern and suggested that this arrangement could have a functional role. The model of gene regulation proposed by Britten and Davidson (176) and since updated (177,178) is an attempt to integrate the information available on DNA sequence organization and gene regulation in eukaryotes; here repetitive sequences have been suggested

to represent regulatory elements for the expression of adjacent structural genes.

Subsequent studies, however, have revealed the presence of long period interspersion pattern in a wide variety of unrelated organisms such as birds (179,180), nematode (181) and fungi (182-185). Studies on DNA sequence organization patterns in plants have also revealed a great diversity (Table 1.7). From the table it is clear that while plants like wheat, rye, maize and tobacco exhibit mainly short period interspersion pattern, the rest of the plants have either mixed or long period interspersion patterns. Rice and flax are unique in having no interspersion of repeat and single copy DNA sequences at the specific fragment length studied.

Another interesting point that has emerged from the analysis of sequence organization in plants is the dependence of pattern of interspersion on the genome size. Plants with genome size (1C) smaller than 2×10^9 base pairs show predominance of long period pattern, while genomes larger than 1×10^9 base pairs show the predominance of short period pattern.

It is suggested that the great diversity in genome organization patterns may not reflect differences in gene regulation mechanisms. Instead, the pattern of sequence organization appears to reflect the nature of processes

Table 1.7 DNA sequence organization patterns in plants

Systematic Name	Haploid DNA content (pg)	Length of the interspersed repetitive sequences (np)	Length of interspersed single copy sequences (np)	Pattern of inter-spersion	Reference
Vigna radiata	0.50	300-120, 1200-6700	750, 3950	Interme-diate	10
Oryza sativa	0.60	50-100, 6000-6400	-	Absent	186
Linum usitatissimum	0.60	> 10,000	-	Long	5
Gossypium hirsutum	0.80	1250	1800, 4000	Interme-diate	4
Matthiola incana	1.50	350-400	1080	Short	2
Eleusine coracana	1.60	150-200, 4000-4200	1900	Mixed	187
Phaseolus aureus	1.80	1900	1300	Interme-diate	188
Phaseolus vulgaris	1.85	1545	1300	Interme-diate	188
Petroselinum sativum	2.0	300-400	1000	Mixed	7
Pennisetum americanum	2.5	4300-4500	1900	Long	16
Glycine max	3.25	300-400, 1500	1235	Short	3,78
Nicotiana tabaccum	3.90	300, 1500	1400	Short	6
Pisum sativum	4.95	300- 400	300-1000	Short	9
Zea mays	7.70	500-1000, 7500	2100	Short	19
Secale cereale	9.0	300-800	1800	Short	189
Allium cepa	16.75	300	1000	Short	11
Triticum aestivum	18.1	300-400	1500	Short	190

like amplification and translocation responsible for evolution of repeated DNA sequences. Frequent amplification would lead to highly repetitive genomes, and when coupled with relatively frequent transposition events, genomes with extensive short period interspersion would result (191). Thus the pattern of sequence arrangement is determined by the rate of 'turnover' of the genomes, the rate being different even among organisms with similar genome sizes.

MODELS OF REPETITIVE DNA IN EVOLUTION: Various models have been presented to account for the accumulation and persistence of repetitive DNA in eukaryotic genomes. These models can be classified according to the level at which natural selection has operated during production of repeats, and whether the action of natural selection is sequence dependent or independent.

The function-dependent models have been suggested particularly by Britten and Davidson (176) and Davidson and Britten (177,178). These models suggest a sequence-dependent positive selection for the repeats, because of their effects on the developmental programs of the organism. The proposed roles of these repeats include (i) regulation of transcription, (ii) production of evolutionary novelty, (iii) encoding of regions within

large RNAs which allows them to function as regulatory activators and (iv) integration of chromosome structure and recognition.

The more recent sequence-dependent models include that of 'selfish DNA' proposed by Orgel and Crick (192) and Doolittle and Sapienza (193). However, this model suggests that only those sequences accumulate which can make use of the cells' replication machinery. Phenotypic selection acts only to check the accumulation when it begins to affect severely the working of the genome.

The highly sequence-independent 'ignorant' model of Dover (194) also suggests phenotypic selection only as a check. The model postulates no direct selection on repeats at all, and suggests that they are rather occasional randomly generated side-products of the genomic machinery for DNA metabolism.

The 'nucleotypic' model of Cavalier-Smith (195,196) is also sequence-independent, but suggests a link between selection based on phenotype and accumulation of moderate repeats. Selection acts only on overall DNA content. This determines the physical volume of the nucleus, and thus indirectly the cell size, cell cycle time, nucleus-cytoplasm transport and other aspects of nuclear and cell size.

According to Bouchard (197), the genomes are 'predisposed' to produce repeats, and the 'tolerance' of the genome is the rate at which it can accept such events. According to the 'ignorant' model, the production of periodic saltation is an unavoidable byproduct of the complex machinery of DNA metabolism of the cell. This may be true to some degree, since Darlington (198) has argued that characteristics which have an overall selective advantage are adopted during evolution, even if they carry small disadvantages. Once such mechanisms are incorporated, they are quite difficult to change or replace. Jain (199) suggests that there may have been positive selection for saltatory replication as a potential source of evolutionary novelty not provided by point mutations. The tendency towards saltation may vary from one genome to another.

Finally, under the scheme suggested by Bouchard (197), the potential for accumulation of moderate repeats within a given genome will be governed by the extent of predisposition, receipt of certain sequences from external sources, and tolerance exhibited by the ancestral lineages. The amount and characteristics of repeats observed at present will be a function of the passive rate of accumulation over time (balanced to an unknown

degree by probably passive deletion). In addition, some portion of the accumulated repeats may have become fixed due to having possessed or acquired a selectively advantageous function or because of autonomous tendencies (e.g. transposons).

All the models of evolution discussed above are not mutually exclusive. Different genomes may be subjected to various processes included in these models to different degrees.

SCOPE OF THESIS

The family Cucurbitaceae contains a number of common vegetable and fruit crops such as gourds and melons, distributed in the warmer parts of the world. All these plants are climbing herbs or shrubs, and are very similar in their morphological features. The family in general shows the following salient features:

1. Absence of polyploidy during evolution of cultivated species.
2. Presence of monoecious or dioecious forms. Occurrence of a heteromorphic pair of chromosomes in some instances.
3. Adoption of both premating and incompatibility barriers, leading to almost total species isolation.
4. Occurrence of small chromosomes, with chromosome number generally 2n = 22, 24 or 26. Occurrence of loss of chromatin during evolution.
5. Small genome size (1C), varying in a narrow range of 1.75 to 2.25 pg.
6. Presence of satellite DNAs in many species.
7. Presence of mitochondrial genome which is large and variable in size.

The molecular biological data available so far in the family Cucurbitaceae is limited to mainly satellite DNA

sequences, ribosomal cistrons and mitochondrial genomes. Since no information was available about the structure and organization of nuclear genomes in Cucurbitaceae plant species, studies were undertaken to analyse the nuclear DNA of a few cucurbits. It is expected that such studies would be extremely useful in understanding the role of repeated DNA sequences in speciation, genome repatterning during evolution, and molecular basis for the strong species isolation.

The following four species belonging to three different genera have been chosen for study.

1. Luffa cylindrica (sponge gourd)
2. Luffa acutangula (ridge gourd)
3. Benincasa hispida (ash gourd)
4. Coccinia indica (ivy gourd).

Chapter II includes isolation and physico-chemical characterization of the nuclear DNAs including high resolution thermal denaturation studies. Chapter III deals with determination of amount of repetitive DNA in these genomes by optical reassociation and hydroxyapatite column chromatography. The Cot curves are analyzed using nonlinear least squares method to reveal the components reassociating with second order kinetics. The reassociation data are also used to estimate the frequency of repetition, kinetic complexity and proportions of different repetitive fractions.

Further analysis of repetitive DNA sequences has been summarized in chapter IV. Reassociation kinetics have been carried out at different criteria of stringency to determine the proportion of very highly diverged and of less well-matched repetitive sequences. Thermal denaturation and optical reassociation of different Cot fractions have been studied to determine their kinetic heterogeneity and base mismatch. In chapter V, arrangements of repeated and single copy DNA sequences using different complementary approaches are described. In chapter VI, results obtained by DNA - DNA hybridization between repeated DNAs of sponge gourd and of ridge gourd, ash gourd and ivy gourd are given. Finally, in chapter VII, our conclusions on the Cucurbitaceae nuclear genomes are described.

1. Luffa: A cosmopolitan genus - Out of 9 species 7 occur in India, and two are cultivated.

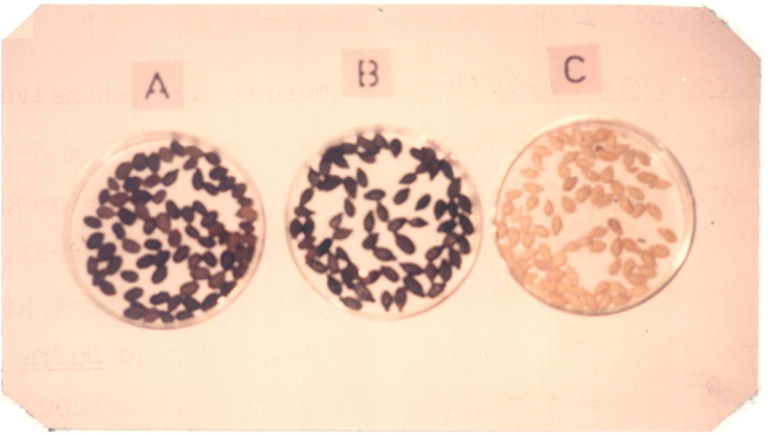
a) Luffa cylindrica (Sponge gourd): An extensive climber with bright yellow flowers. Fruits are 15-50 x 6-10 cms, fusiform, strongly fibrous inside. Seeds are smooth, usually black, with slightly winged margins. Tender fruits are used as vegetable. The ripe fruits produce the best kind of bath sponge. This species has great affinity to Luffa acutangula with respect to many floral characteristics (200).

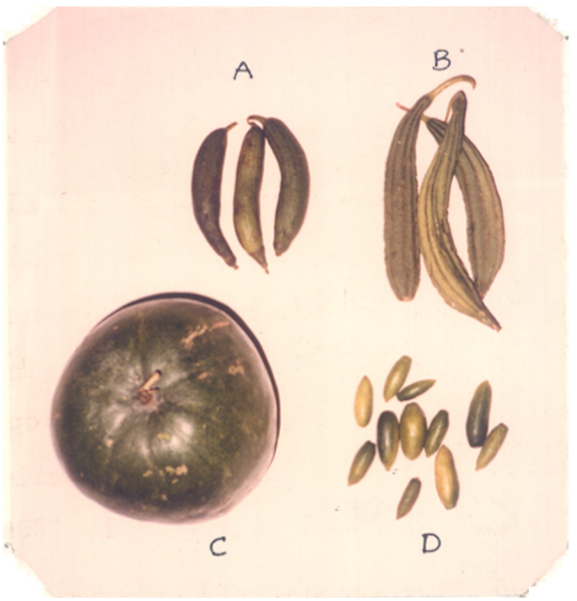
b) Luffa acutangula (Ridge gourd): Extensive climber with pale yellow flowers. Fruits 15-30 x 6-10 cms, angular, acutely 10-angled. Seeds are black, thick, compressed. Ridged gourd is a popular vegetable and eaten when tender (200).

2. Benincasa hispida (Ash gourd): An annual climber. The plant is softly hairy. Flowers are large, yellow, solitary. Fruit is thick, hispid, fleshy, hairy when young and waxy when mature. Seeds are yellowish white, marginate and compressed. Cultivated in tropical and subtropical regions of India. No where it is recorded to occur wild. Both young and ripe fruits are used as vegetable (200).

3. Coccinia indica (Ivy gourd): 13 species, mostly confined to tropical Africa. Only one occurs in India. A dioecious herb with white flowers. Fruit 5.0 x 2.5 cms, with red and juicy pulp. Fruit is a common vegetable (200).

Figure 1.2 : Seeds and Fruits of Cucurbitaceae





A Sponge gourd

B Ridge gourd

C Ash gourd

D Ivy gourd

REFERENCES

1. Dhillon, S.S., Rake, A.V. and Miksche, J.P. (1980) Plant Physiol. 65, 1121-1127.

2. Wenzel, W. and Hemleben, V. (1979) Plant Syst.Evol.Suppl. 2, 29-40.

3. Gurley, W.B., Hepburn, A.G. and Key, J.L. (1979) Biochim.Biophys.Acta 561, 167-183.

4. Walbot, V. and Dure, L.S. (1976) J.Mol.Biol. 101, 503-536.

5. Cullis, C.A. (1981) Biochim.Biophys.Acta 652, 1-15.

6. Zimmerman, J.L. and Goldberg, R.B. (1977) Chromosoma (Berl.) 59, 227-252.

7. Kiper, M. and Herzfeld, F. (1978) Chromosoma (Berl.) 65, 335-351.

8. Seshadri, M. and Ranjekar, P.K. (1979) Indian J.Biochim.Biophys. 16, 1-15.

9. Murray, M.G., Cuellar, R.E. and Thompson, W.F. (1978) Biochemistry 17, 5781-5790.

10. Murray, M.G., Palmer, J.D., Cuellar, R.E. and Thompson, W.F. (1979), Biochemistry 18, 5259-5266.

11. Stack, S.M. and Comings, D.E. (1979) Chromosoma (Berl.) 70, 161-181.

12. SivaRaman, L. and Ranjekar, P.K. (1984) Indian J.Biochem.Biophys. (In press).

13. Deshpande, V.G. and Ranjekar, P.K. (1980) Hoppe-Seyler's Z.Physiol.Chem. 361, 1223-1233.

14. Gupta, V.S. (1981) Structure and sequence organization of DNA in eukaryotes with special reference to three Gramineae plant species, Ph.D. Thesis, University of Poona, Poona, India.

15. Ranjekar, P.K., Pallotta, D., Lafontaine, J.G. (1976) Biochim.Biophys.Acta. 425, 30-40.

16. Wimpee, C.F. and Rawson, J.R.Y. (1979) Biochem.Biophys.Acta. 562, 192-206.

17. Ranjekar, P.K., Lafontaine, J.G. and Pallotta, D. (1974) Chromosoma (Berl.) 48, 427-440.

18. Smith, D.B. and Flavell, R.B. (1975) Chromosoma (Berl.) 50, 223-242.

19. Hake, S. and Walbot, V. (1980) Chromosoma (Berl.) 79, 251-270.

20. Walbot, V. and Dure, L.S. (1976) J.Mol.Biol. 101, 503-536.

21. Bazetoux, S., Jouanin, L., and Huguet, T. (1978) Nucleic Acids Res. 5, 750-768.

22. Wellauer, P.K. and Dawid, I.B. (1974) J.Mol.Biol. 89, 379-395.

23. Jelinek, W., Molloy, G., Solditt, M., Wall, R., Sheiness, D. and Darnell, J.E. (Jr) (1973) Cold Spring Harbor Symp. Quant.Biol. 38, 891-898.

24. Ryskov, A.P., Saunders, G.F., Farashyan, V.R. and Georgier, G.P. (1973) Biochim.Biophys.Acta. 312, 152-164.

25. Calos, M.P. and Miller, J.H. (1980) Cell. 20, 579-595.

26. Dickson, R.C., Ahelson, J., Barnes, W.M. and Reznikoff, W.S. (1975) Science 187, 27-35.

27. Darlix, J.L. and Horaist, M. (1975) Nature (Lond.) 256, 288-292.

28. Kelly, T.J. and Smith, H.O. (1970) J.Mol.Biol. 51, 393-409.

29. Dunn, J.J. and Studier, F.W. (1973) Proc.Natl. Acad.Sci. USA, 70, 3296-3300.

30. Vogeli, G., Ohkubo, H., Sobel, M.E., Yamada, Y., Pastan, I. and de Crombrugghe, B. (1981) Proc.Natl.Acad.Sci. USA. 78, 5334-5338.

31. Skinner, D.M. (1977) Bioscience 27, 790-796.

32. Deumling, B., Sinclair, J., Timmis, J.N. and Ingle, J. (1976) Cytobiologie 13, 224-232.

33. Ohyama, K., Gamborg, O.L. and Miller, R.A. (1972) Plant Physiol. 50, 319-321.

34. Ingle, J., Timmis, J.N. and Sinclair, J. (1975) Plant Physiol. 55, 496-501.

35. Ingle, J., Pearson, G.G. and Sinclair, J. (1973) Nature (New Biol). 242, 193-197.

36. Beridze, T.G. (1975) Biochim.Biophys.Acta. 395, 274-279.

37. Bendich, A.J. and Anderson, R.S. (1974) Proc.Natl.Acad.Sci. USA 71, 1511-1515.

38. Ranjekar, P.K., Pallotta, D. and Lafontaine, J.G. (1978) Can.J.Biochem. 56, 808-815.

39. Kadouri, A., Atsmon, D. and Edelman, M. (1975) Proc.Natl.Acad.Sci. USA 72, 2260-2264.

40. Duhrssen, E. and Neumann, K-H. (1980) Z.Pflanzenphysiol. 100, 447-474.

41. Aripdzhanov, Sh A., Bespal'Ko N.V. and Ibraginsov, A.P. (1976) Uzbek Biol Zh. 1, 9-10.

42. Strubbe, K., van Oostveldt, P., Brockaert, D. and van Parijs, R. (1982) Planta 155, 238-343.

43. Chilton, M.D. (1975) Genetics 81, 469-483.

44. Seshadri, M. and Ranjekar, P.K. (1979) Indian J.Biochem.Biophys. 16, 1-15.

45. Beridze, T.G. and Bragvadze, G.P. (1976) Molekul.Biol. 10, 1279-1289.

46. Seshadri, M. and Ranjekar, P.K. (1980) Hoppe-Seyler's Z.Physiol.Chem. 361, 1041-1048.

47. Capesius, I. (1979) Plant Syst.Evol. 133, 1-14.

48. Deumling, B. and Nagl, W. (1978) Cytobiologie 16, 412-420.

49. Capesius, I., Bierweiler, B., Bachmann, K., Rücker W. and Nagl, W. (1975) Biochim.Biophys.Acta. 395, 67-73.

50. Timmis, J.N., Deumling, B. and Ingle, J. (1975) Nature (Lond.) 257, 152-155.

51. Ranjekar, P.K., Pallotta, D. and Lafontaine, J.G. (1976) Biochim.Biophys.Acta 520, 103-110.

52. Appels, R., Driscoll, C. and Peacock, W.J. (1978) Chromosoma (Berl.) 70, 67-89.

53. Capesius, I. (1983) Biochim.Biophys.Acta 739, 276-280.

54. Huguet, T. and Jouanin, L. (1972) Biochim.Biophys. Res. Commun. 46, 1169-1174.

55. Peacock, W.J., Lohe, A.R., Gerlach, W.L., Dunsmuir, P., Dennis, E.S. and Appels, R. (1977) Cold Spring Harbor Symp. Quant.Biol. 42, 1121-1135.

56. Bendich, A.J. and Taylor, W.C. (1977) Plant Physiol. 59, 604-609.

57. Reis, R.S., Timmis, J.N. and Ingle, J. (1981) Biochem.J. 195, 723-734.

58. Hemleben, V., Leweke, B., Roth, A. and Stadler, J. (1982) Nucleic Acids Res. 10, 631-644.

59. Beridze, T.G. (1979) Biochem. (USSR) 44, 1546-1552.

60. Bragvadze, G.R. (1983) Biochem. (USSR) 48, 48-55.

61. Reisner, A.H., Colless, V.M. and Bucholtz, C.A. (1980) Aust.J.Biol.Sci. 33, 623-632.

62. Lee, C.S. (1981) Chromosoma (Berl.) 83, 367-379.

63. Nagl, W. and Capesius, I. (1977) Chromosomes Today 6, 141-152.

64. Yunis, J.J. and Yasmineh, W.G. (1971) Science 174, 1200-1209.

65. Sharma, A.K. (1978) Nucleus 21, 113-116.

66. Fry, K. and Salser, W. (1977) Cell 12, 1069-1084.

67. Walker, P.M.B. (1971) Nature 229, 306-308.

68. Grisvard, J. and Tuffet-Anghileri, A. (1980) Nucleic Acids Res. 8, 2843-2858.

69. Durante, M., Geri, C., Grisvard, J., Guille, E., Parenti, R. and Buiatti, M. (1983) Protoplasma 114, 114-118.

70. Nagl, W. and Rücker, W. (1976) Nucleic Acids Res. 3, 2033-2039.

71. Fuhrmann, B. and Nagl, W. (1979) Plant Syst.Evol. Suppl. 2, 235-245.

72. Cullis, C.A. and Schweizer (1974) Chromosoma (Berl.) 44, 417-421.

73. Chooi, W.Y. (1971) Genetics 68, 213-230.

74. Straus, N.A. (1972) Carnegie Inst.Wash. Year Book 71, 257-259.

75. Gnucheva, E.,Yu, Kir'yanov, G.I., Polyaka, V. Yu and Chenstov Yu S. (1977) Biokhimiya 42, 2186-2191.

76. Schaan, M.E. and Nagl, W. (1979) Plant Syst.Evol.Suppl. 2, 67-71.

77. Ranjekar, P.K., Pallotta, D. and Lafontaine, J.G. (1978) Can.J.Biochem. 402, 171-178.

78. Goldberg, R.B. (1978) Biochem.Genet. 16, 45-68.

79. Wilson, J.T., Katterman, F.R.H. and Endrizzi, J.E. (1976) Biochem.Genet. 14, 1071-1075.

80. Nagl, W. and Capesius, I. (1976) Plant Syst.Evol. 126, 221-237.

81. Nze-Ekekang, L., Patillon, M., Schafer, A. and Kovoor, A. (1974) J.Expt.Bot. 25, 320-329.

82. Narayan, R.K.J. and Rees, H. (1976) Chromosoma (Berl.) 63, 101-107.

83. Sivolap, Y.M. and Bonner, J. (1971) Proc.Natl.Acad.Sci. USA. 68, 387-389.

84. Pearson, W.R., Smith, S.L., Wu, J.R. and Bonner, J. (1978) Plant Physiol. 62, 112-115.

85. Pearson, W.R., Wu, J.R. and Bonner, J. (1978) Biochemistry 17, 51-59.

86. Gaddipatti, J.P.B. (1976) Characterization of chromosomal DNA in higher plants with varying genome size. Ph.D. Thesis, Banaras Hindu University, Varanasi, India.

87. Ranjekar, P.K., Pallotta, D., and Lafontaine, J.G. (1978) Biochem.Genet. 16, 957-970.

88. Bendich, A.J. and Anderson, R.S. (1977) Biochemistry 16, 4655-4663.

89. Schaffner, K.H. and Nagl, W. (1979) Plant Syst. Evol.Suppl. 2, 105-110.

90. Smyth, D.R. and Stern, H. (1973) Nature (New Biol.) 245, 94-96.

91. Flavell, R.B. and Smith, D.B. (1975) Stadler Symp. 7, 47-70.

92. Stanziano, P. and Maggini, F. (1976) Atti.Ass.Genet. Ital. 21, 66-69.

93. Bendich, A.J. and McCarthy, B.J. (1970) Genetics 65, 545-565.

94. Stern, H. and Hotta, Y. (1977) Phil.Trans.R.Soc. London Ser.B. 277, 277-294.

95. Stern, H. and Hotta, Y. (1978) Annu.Rev.Plant Physiol. 29, 415-436.

96. Jelinek, W.R. and Haynes, S.R. (1983) Cold Spring Harbor Symp. Quant.Biol. 47, 1123-1130.

97. Soriano, P., Meunier-Rotival, M. and Bernardi, G. (1983) Proc.Natl.Acad.Sci. USA. 80, 1816-20.

98. Daniels, G.R. and Deininger, P.L. (1983) Nucleic Acids Res. 11, 7595-7610.

99. Daniels, G.R., Fox, G.M., Lowensteiner, D., Schmid, C.W. and Deininger, P.L. (1983) Nucleic Acids Res. 11, 7579-7593.

100. Fritsch, E.F., Shen, C.K.J., Lawn, R.M., Maniatis, T. (1981) Cold Spring Harbor Symp. Quant.Biol. 45, 761-775.

101. Gebhard, W., Meitinger, T., Hochtl, J. and Zachau, H.G. (1982) J.Mol.Biol. 157, 453-471.

102. Gebhard, W. and Zachau, H.G. (1983) J.Mol.Biol. 170, 255-270.

103. Shimizu, Y., Yoshida, K., Ren, C-S., Fujinaga, K., Rajagopalan, S. and Chinnadurai, G. (1983) Nature (Lond.) 302, 587-590.

104. Shafit-Zagaro, B., Brown, F.L., Maio, J.J. and Adams, J.W. (1982) Gene 20, 397-407.

105. Cillespie, D., Adams, J.W., Costanzie, C. and Jo Caranfa, M. (1982) Gene 20, 409-414.

106. Shafit-Zagardo, B., Brown, F.L., Zavodny, P.J. and Maio, J.J. (1983) Nature (Lond.) 304, 277-280.

107. Lerman, M.I., Thayer, R.E. and Singer, M.F. (1983) Proc.Natl.Acad.Sci. USA. 80, 3966-3970.

108. Wu, J.C. and Manuelidis, L. (1980) J.Mol.Biol. 142, 363-386.

109. Brown, S.D.M. and Dover, G. (1981) J.Mol.Biol. 150, 441-446.

110. Dubnick, M., Chou, J., Petes, T.D. and Farber, R.A. (1983) J.Mol.Evol. 19, 115-121.

111. Heller, R. and Arnheim, N. (1980) Nucleic Acids Res. 8, 5031-5042.

112. Manuelidis, L. (1980) Nucleic Acids Res. 8, 3247-3258.

113. Manuelidis, L. (1982) Nucleic Acids Res. 10, 3211-19.

114. Manuelidis, L. and Biro, P.A. (1982) Nucleic Acids Res. 10, 3221-3239.

115. Meunier-Rotival, M., Soriano, P., Cuny, G., Straus, F. and Bernardi, G. (1981) Proc.Natl.Acad.Sci. USA. 79, 355-359.

116. Fox, G.M., Hess, J.F., Shen, C.K.J. and Schmid, C.W. (1983) Cold Spring Harbor Symp. Quant.Biol. 47, 1131-34.

117. Taylor, J.H. (1984) Molec. and Cell Biol. 61, 99-109.

118. Krayer, A.S. Markusheva, T.V., Kramerov, D.A., Ryskov, A.P., Skryalin,K.G., Bayer, A.A. and Georgier, G.P. (1982) Nucleic Acids Res. 10, 7461-7475.

119. McClintock, B. (1950) Proc.Natl.Acad.Sci. USA. 36, 344-355.

120. McClintock, B. (1951) Cold Spring Harbor Symp. Quant. Biol. 16, 13-47.

121. Burr, B. and Burr, F.A. (1981) Genetics 98, 143-156.

122. Tebbe, V.C., Doring, H.P., Federoff, N. and Starlinger, P. (1983) Cell 34, 383-393.

123. Echt, C.S. and Schwartz, D. (1981) Genetics 99, 275-284.

124. Shure, M., Wessler, S. and Fedoroff, N. (1983) Cell 35, 225-233.

125. Federoff, N., Wessler, S. and Shure, M. (1983) Cell 35, 235-242.

126. Reddy, A.R. and Peterson, P.A. (1983) Molec.Gen.Genet. 192, 21-31.

127. Kermicle, J.L. (1980) Science 208, 1457-1459.

128. Friedemann, P. and Peterson, P.A. (1982) Molec.Gen.Gent. 187, 19-29.

129. Shepherd, N.S., Schwarz-Sommer, Z., Vel Spalve, J.B., Gupta, M., Weinand, U. and Saedler, H. (1984) Nature (Lond.) 307, 185-187.

130. Goldberg, R.B., Hoschek, G. and Vodkin, L.O. (1983) Cell 33, 465-475.

131. Vodkin, L.O., Rhodes, P.R. and Goldberg, R.B. (1983) Cell 34, 1023-1031.

132. Hinton, D.M. and Musso, R.E. (1983) J.Mol.Biol. 169, 53-81.

133. Jqurin, B. and Normark, S. (1983) Cell 32, 809-816.

134. Simons, R.W., Hoopes, B.C., Mc Clure, W.R. and Kleckner, N. (1983) Cell 34, 673-682.

135. Calos, M.P. and Miller, J.H. (1980) Cell 20, 579-595.

136. Crick, F. (1979) Science 204, 264-271.

137. Nerbert, D.W. (1979) Molec.Cell Biochem. 27, 27-46.

138. Shapiro, J.A. and Cordell, B. (1982) Biol.Cell 43, 31-54.

139. Flavell, R.B., O'Dell, M. and Hutchinson, J. (1981) Cold Spring Harbor Symp. Quant.Biol. 45, 501-508.

140. Ono, K. and Suguiura, M. (1980) Chromosoma (Berl.) 76, 85-89.

141. Romanani, N.M.G. (1972) J.Hum.Evol. I, 23-40.

142. Hemleben, V., Bachmann, K. and Price, H.C. (1978) Experientia 34, 1452-1453

143. Maggini, F. and Carmona, M. (1981) Protoplasma 108, 163-171.

144. Langridge, P. and Rix, G. (1983) Cell 34, 1015-1022.

145. Rubenstein, I. (1982) In: Maize for Biological Res. (Sheridan, W.F., ed.) Plant Molec.Biol.Asso., Charlottesville, VA 22905, pp.189-195.

146. Viotti, Sala, E., Marotta, R., Alberi, P., Balducci, C. and Soave, C. (1979) Eur.J.Biochem. 102, 211-222.

147. Hopp, H.E., Rasmussen, S.K. and Brandt, A. (1983) Carlsberg Res.Commun. 48, 201-206.

148. Schuler, M.A., Schmitt, E.S. and Beachy, R.N. (1982) Nucleic Acids Res. 10, 8225-8244.

149. Galau, G.A., Chlan, C.A. and Dure, L. III (1983) Plant Mol.Biol. 2, 189-198.

Pyle D.B. Baker J.C.

150. Dure, L.III, Galau, C.A. and Chlan, C.A. (1983) Plant Mol.Biol. 2, 199-206.

151.* Hall, T.C. (1983) Structure and Function of Plant Genomes pp. 123-142.

Shah D.M., Eckenrode V.K. and Meagher R.B.

152. Nagao, R.T (1981) DNA 1, 1-10.

153. Baulcombe, D. and Verma, D.P.S. (1978) Nucleic Acids Res. 5, 4141-4153.

154. Broglie, R., Coruzzi, G., Lamppa, G., Keith, B. and Chua, N. (1983) Bio/Technol. 1, 55-61.

155. Smith, S.M., Bedbrook, J. and Speirs, J. (1983) Nucleic Acids Res. 11, 8719-8723.

156. Smith, G.P. (1976) Science 191, 528-535.

157. Soave, C., Reggiani, R., DiFonza, N. and Salamini, F. (1982) Biochem.Genet. 20, 1027-1038.

158. Suzuki, Y., Cage, L.P. and Brown, D.D. (1972) J.Mol.Biol. 70, 637-649.

159. Bishop, J.O. and Rosbash, M. (1973) Nature (New Biol). 241, 204-207.

160. Harrison, P.R., Birnie, G.D., Hell, A., Humphries, S., Young, B.D. and Paul, J. (1979) J.Mol.Biol. 84, 539-554.

161. Harris, S.E., Means, A.R., Mitchell, W.M. and O'Malley, B.W. (1973) Proc.Natl.Acad.Sci.USA. 70, 3776-3780.

162. Kleiman, L., Birnie, G.D., Young, B.D. and Paul, J. (1977) Biochemistry 16, 1213-1233.

163. Goldberg, R.B., Hoschek, G., Kamalay, J.C. and Timberlake, W.E. (1978) Cell 14, 123-131.

164. Galau, G.A., Britten, R.J. and Davidson, E.H. (1974) Cell 2, 9-20.

165. Hough, B.R., Smith, M.J., Britten, R.J. and Davidson, E.H. (1975) Cell 5, 291-299.

166. Britten, R.J. and Kohne, D.E. (1968) Science 161, 529-540.

167. Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973) J.Mol.Biol. 77, 1-23.

168. Manning, J.E., Schmid, C.W. and Davidson, N. (1975) Cell 4, 141-155.

169. Graham, D.E., Neufeld, B.R., Davidson, E.H. and Britten, R.J. (1974) Cell 1, 127-137.

170. Efstratiadis, A., Crain, W.R., Britten, R.J., Davidson, E.H. and Kafatos, F.C. (1976) Proc. Natl.Acad Sci. USA. 73, 2289-2293.

171. Olmo, E., Stingo, V., Odierna, G. and Capriglione, T. (1982) Experientia 38, 339-340.

172. Mayfield, J.E., Mc Kenna, J.F. and Lessa, B.S. (1980) Chromosoma (Berl.) 76, 271-294.

173. Avvedimento, V.E., Acquaviva, A.M. and Varrone, S. (1976) Nucleic Acids Res. 3, 2491-2505.

174. Schmid, C.W. and Deininger, P.L. (1975) Cell 6, 345-358.

175. Firtel, R.A. and Kindle, K. (1975) Cell 5, 401-411.

176. Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357.

177. Davidson, E.H. and Britten, R.J. (1975) Quart. Rev.Biol. 48, 565-613.

178. Davidson, E.H. and Britten, R.J. (1979) Science, 204, 1052-1059.

179. Epplen, J.T., Leipoldt, M., Engel, W. and Schmidtke, J. (1978) Chromosoma (Berl.) 69, 307-321.

180. Wagenmann, M., Epplen, J.T., Bachmann, K., Engel, W. and Schmidtke, J. (1981) Experientia 37, 1274-1276.

181. Schachat, F., O'Conner, D.J. and Epstein, H.F. (1978) Biochim.Biophys.Acta 520, 683-692.

182. Ullrich, R.C., Kohorn, B.D. and Specht, C.A. (1980) Chromosoma (Berl.) 81, 371-378.

183. Harshey, R.M., Jayaram, M. and Chamberlin, M.E. (1979) Chromosoma (Berl.) 73, 143-151.

184. Hudspeth, M.E.S., Timberlake, W.E. and Goldberg, R.B. (1977) Proc.Natl.Acad.Sci.USA, 74, 4332-4336.

185. Krumlauf, R. and Marluf, G.A. (1979) Biochem. 18, 3705-3713.

186. Gupta, V.S., Gadre, S.R. and Ranjekar, P.K. (1981) Biochim.Biophys.Acta 656, 147-154.

187. Gupta, V.S. and Ranjekar, P.K. (1981) J.Biosci. 3, 417-430.

188. Seshadri, M. and Ranjekar, P.K. (1980) Biochim.Biophys.Acta 610, 211-220.

189. Smith, D.B. and Flavell, R.B. (1977) Biochim.Biophys. Acta 474, 32-97.

190. Flavell, R.B. and Smith, D.B. (1976) Heredity 37, 231-252.

191. Thompson, W.F. and Murray, M.G. (1980) In: The Plant Genome (Davies, D.R. and Hopwood, D.A., eds.) pp. 31.45.

192. Orgel, L.E. and Crick, F.H.C. (1980) Nature (Lond.) 284, 604-607.

193. Doolittle, W.F. and Sapienza, C. (1980) Nature (Lond.) 284, 601-603.

194. Dover, G. (1980) Nature (Lond.) 285, 618-620.

195. Cavalier-Smith, T. (1978) J.Cell.Sci. 34, 247-278.

196. Cavalier-Smith, T. (1980) Nature (Lond.) 285, 617-618.

197. Bouchard, R.A. (1982) In: Int.Rev.Cytol. (Bourne, G.H. and Danielli, J.F., eds.) Vol.32, Academic Press, New York, pp. 113-193.

198. Darlington, C.D. (1971) Nature (Lond.) 234, 521-525.

199. Jain, H.K. (1980) Nature (Lond.) 288, 647-648.

200. Chakravorty, H.L. (1983) In: Fascicles of Flora of India. (Thothathri, K., ed.) Fascicle II Cucurbitaceae, Botanical Survey of India, Howrah, India.

151* Hall T.C., Slighton J.L., Ersland D.R., Murray M.G., Hoffman L.M., Adang M.J., Brown J.W.S., Ma Y., Matthew J.A., Cramer J.H., Barker R.F., Sutton D.W. and Kemp J.D. (1983) In: Structure and Function of Plant Genomes, Ciferri O. and Dure III. L.(eds.). Plenum Publishing corpn., New York. pp. 123-142.

CHAPTER II

HIGH RESOLUTION THERMAL DENATURATION ANALYSIS OF DNAs

HIGH RESOLUTION THERMAL DENATURATION ANALYSIS OF DNAs

SUMMARY

Thermal denaturation studies of DNAs of four Cucurbitaceae plant species have been carried out to determine the base composition, hyperchromicity and transition width of melting. The melting temperatures of the total, unsheared DNAs fall in the range of 83.2 to 85.7°C. The G+C contents calculated from the melting temperatures range from 34 to 40%, the lowest being in ivy gourd and the highest being in ridge gourd. The transition widths of 6 to 8°C represent a moderate sequence divergence in the four species. The high resolution thermal denaturation profiles have shown species-specific differences. About 12 out of 17 to 22 melting components are common to all the species, when a range of 1.5% G+C is considered for comparison. The two Luffa species show very similar first derivative melting profiles, and hence may have a significant DNA sequence homology. The profiles of Luffa species are, however, very distinct from those of ash gourd and ivy gourd.

INTRODUCTION

Melting of DNA indicates the separation of the two component strands which are initially held together by hydrogen bonds and hydrophobic interactions. The DNA melting

profile in general is a two dimensional display obtained by plotting a change in some measurable property of DNA like alterations in viscosity, optical rotation, buoyant density, light scattering and extinction coefficient, against an external variable responsible for melting of DNA, such as temperature, pH, ionic strength, organic solvents and concentration of ligand molecules. The DNA melting profile is primarily characterized by the following parameters: (i) The melting temperature (Tm), which is defined as the temperature corresponding to the midpoint of the melting curve (1); (ii) the sharpness or breadth of melting, ΔT_a^b or σ 2/3, defined as the difference between the temperatures of 'a' (17%) and 'b' (83%) melting (2,3); (iii) hyperchromicity which is expressed in terms of percent increase in absorbance after complete denaturation. Under constant conditions of ionic strength and pH, a linear relationship exists between the melting temperature and the base composition of DNA expressed in terms of G+C.

The most popular method for studying the melting of DNA is to monitor the change in ultra violet absorbance caused by elevating the temperature of DNA solution. This method is simple and reliable, and environmental conditions can be set with precision. The UV absorbance by nucleotide bases in the wavelength region around 260 nm is due to the π-π^* electronic transition in purine and pyrimidine bases.

The intensity of absorbance increases by about 40% when the hydrogen bonds between the bases are disrupted. A linear relationship exists between this hyperchromicity and the extent of disruption of base pairs.

If a given DNA molecule contains regions within it which are markedly different in G+C content from the average G+C content of the whole molecule, then such regions will have different Tms from the rest of the molecule resulting in a heterogeneous denaturation behaviour. Such heterogeneity is seldom visible in the simple denaturation profiles, but can be readily detected by plotting "derived" melting curves, where the ratio of hyperchromicity over a small rise in temperature (e.g. 0.1°C) to the total hyperchromicity is plotted against temperature. Such studies have revealed an ordered and non-random distribution of base sequences in most organisms. Three kinds of derivative melting profiles are generally recognized: (i) Polyphasic melting of short, homogeneous viral DNAs consisting of unique sequences (4-8); (ii) smooth, monophasic melting of bacterial and large viral DNAs (8,9) and (iii) broad, polyphasic melting of higher eukaryotes (8-12). The peaks which appear distinctly in the melting of DNAs of higher organisms are not the products of internal base sequence heterogeneity, but may come from some special families such as satellite DNA sequences or some other

regions specific to functions of higher organisms, and at least in part from repeated sequences (8-13). The fine melting technique can be applied to establish a way in which the DNA double helix is constructed and to relate it with its biological function to find the genetic and evolutionary origin of this nonrandom sequence arrangement and also in recombination and gene manipulation technology (14).

Since the melting profiles provide a histogram of the local G+C content arranged over several hundred base pairs, it provides a convenient method to examine the homology or to explore the taxonomic relationships among DNAs, like deletion mutants of lambda phage (15), kinetoplast DNAs of hemoflagellates (16), and a series of related bacteriophages of Bacillus subtilis (4). Species - and family-specific melting curves have been obtained in some higher plants (12). This technique has been shown to have advantages over the other techniques of genetic distance evaluation, like restriction mapping and DNA - DNA hybridization (4). In addition, fine melting has apparently a greater resolution compared to techniques like density gradient centrifugation and electron microscopy employed for studying distribution in base composition (9,14).

This chapter describes the studies carried out on denaturation characteristics and high resolution melting studies of DNAs of sponge gourd, ridge gourd, ash gourd

and ivy gourd, to determine the heterogeneity in base composition and also to assess the extent of similarities and differences among these four species of Cucurbitaceae.

MATERIALS AND METHODS

DNA sources and seed materials

Seeds of Luffa cylindrica (sponge gourd), Luffa acutangula (ridge gourd), Benincasa hispida (ash gourd) and Triticum aestivum (wheat) and fresh fruits of Coccinia indica (ivy gourd) were obtained locally. The source of seeds and the variety were kept constant in order to avoid qualitative differences, if any. The seeds were washed thoroughly with distilled water, and treated with 0.001% mercuric chloride for 10 min. to reduce the bacterial contamination (17). Seeds were grown on wet cotton in dark under controlled conditions of temperature and humidity. About 6-8" long shoots (grown for 8-10 days) were cut and stored at -20°C till further use. In case of ivy gourd, fresh fruits were thoroughly washed and used for DNA extraction.

E.coli B NCIM 2089 was used for DNA extraction. Calf thymus DNA was obtained commercially.

Chemicals

All the chemicals used throughout the work were of Analytical Reagent (AR) or Guaranteed Reagent (GR) grade.

They were obtained from British Drug House (BDH), Sarabhai Chemicals or E. Merck. Fine chemicals/reagents such as RNase A, Tris (Tris-hydroxy methyl aminomethane) and agarose were obtained from Sigma Chemical Co., MO, USA. Molecular weight markers (λ DNA digested with Hind III and ϕ X 174 RF DNA digested with Hae III), and restriction enzymes were obtained from Boehringer Mannheim, West Germany or Bethesda Research Laboratories, USA.

DNA isolation

DNAs were extracted by a combination of procedures of Marmur (18) and Ranjekar et al. (19). Frozen plant tissue or fresh fruits cut to smaller pieces were homogenized in sucrose buffer (0.5 M sucrose, 0.05 M Tris, 0.05 M maleic acid, 0.003 M $CaCl_2$, pH 6.0, containing 0.1% Triton X-100) in a Remi blender at maximum speed for 1 1/2 min. The homogenate was filtered through two layers of cheese cloth or muslin cloth and the filtrate obtained was centrifuged at 1000 x g for 30 min. The crude nuclear pellet was resuspended in sucrose buffer and centrifuged again at 1000 x g for 20 min. This procedure was repeated twice. The washed 'nuclear' pellet was found to be composed of nuclei as well as a lot of cell debris, and was free of chloroplasts. The pellet was resuspended in saline EDTA (0.15 M NaCl, 0.1 M sodium salt of EDTA, pH 8.0) and the nuclei were lysed by the addition of sodium dodecyl

sulphate to final concentration of 2% followed by incubation at 62°C for 30 min. with intermittent gentle shaking . The suspension was then cooled to 10°C and the proteins were denatured by adding sodium perchlorate to an effective concentration of 1M, followed by deproteinization with chloroform-isoamyl alcohol (24:1 v/v). The latter step was repeated till the interphase was totally absent. The aqueous layer was chilled and the nucleic acids were precipitated with 1.5 volumes of ethanol. The fibres were spooled out, dried, dissolved in 1 x SSC (0.15 M NaCl, 0.015 M trisodium citrate, pH 7.0) and then incubated with RNase A (50 µg/ml) at 37°C for 1 h to remove the contaminating RNA. Prior to use, RNase was heated at 80°C for 10 min. to make it free from any DNase activity. The solution was then deproteinized with chloroform - isoamyl alcohol (24:1 v/v) and the DNA was precipitated with 1.5 volumes of ethanol and dissolved in 1 x SSC. About 5-8 mg of DNA was obtained from 1 kg tissue. Wheat DNA was also isolated according to this procedure. E. coli DNA was isolated according to Marmur's procedure (18).

Criteria of purity of DNA preparations

(I) All the DNA preparations were analysed routinely for RNA, protein and polysaccharide contamination by orcinol test (20), Lowry's assay (21) and paper chromatography of sugars obtained by hydrolysis of polysaccharides (22)

respectively. The DNA preparations had less than 1% contamination of RNA and proteins.

Procedure for polysaccharide estimation: 10-15 µg of DNA precipitate was dissolved in 1 N HCl and hydrolysed in a boiling water bath for 30-35 min. All operations were carried out in a 1.5 ml capacity Eppendorf microfuge vial. After hydrolysis, a pinch of animal charcoal was added and the vial was shaken. The solution was then centrifuged and the supernatent was discarded. Only sugars get adsorbed on charcoal. The adsorbed sugars were eluted from charcoal by suspending the charcoal pellet in ethanol and collecting the supernatent by centrifugation. The ethanol containing the sugars was allowed to evaporate at room temperature to reduce the volume to 5-10 µl. It was then chromatographed on Whatman 1 (ascending) in N-propanol:acetic acid:water (7:2:1) system. The chamber and the paper were saturated before the run. The chromatograph was air-dried and developed in the following sequence: i) dipping in saturated solution of silver nitrate in acetone and drying, ii) dipping in 0.5% NaOH in 90% ethanol and drying, iii) finally dipping in 0.5% sodium thiosulphate solution.

Calf thymus DNA treated similarly, and the sugars like glucose, ribose and deoxyribose were used as standards. The standard sugars gave dark brown spots, while the DNA preparationsshowed only deoxyribose sugar (Figure 2.1).

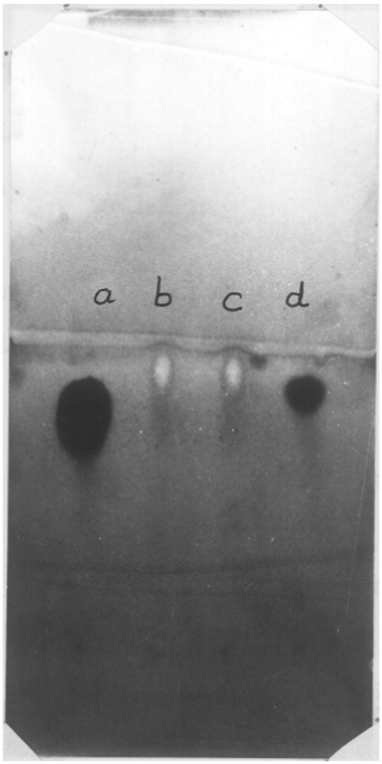


Fig. 2.1 : Determination of polysaccharide contamination of DNAs by paper chromatography.

Lane a : D-glucose

Lane b : DNA of sponge gourd

Lane c : DNA of ash gourd

Lane d : D-ribose

(II) The wavelength scanning of the DNAs was carried out in the range of 220 to 300 nm. All the DNA preparations exhibited an optical density ratio A_{280}/A_{260} = 0.55 and A_{230}/A_{260} = 0.45 and had absorbance less than 0.100 at 300 nm. Figure 2.2 represents a typical absorption spectrum for the DNA preparations.

(III) The DNAs showed approximately 25% hyperchromicity after denaturation. The presence of sharp melting profiles and absence of a 'foot' or shallow absorbance rise before the actual start of melting is indicative of absence of single-stranded DNA and RNA (Figure 2.4) (23).

DNA sizing

Average size of the DNA preparations was determined by agarose slab gel electrophoresis using 1% agarose (Sigma) gels in TAE buffer (40 mM Tris, 20 mM acetic acid, 2 mM EDTA, pH 8.1). Electrophoresis was carried out at 50 mV. The gels were stained in ethidium bromide (0.5 µg/ml) in dark for 10 min., visualized on ultraviolet transilluminater and photographed using a Pentax ES II camera with a 50 mm lens. Hind III digest of λ DNA and Hae III digest of ϕ X 174 RF DNA were used as molecular weight markers. The DNA preparations showed a molecular weight greater than 10 kilo base pairs (Figure 2.3).

Figure 2.2

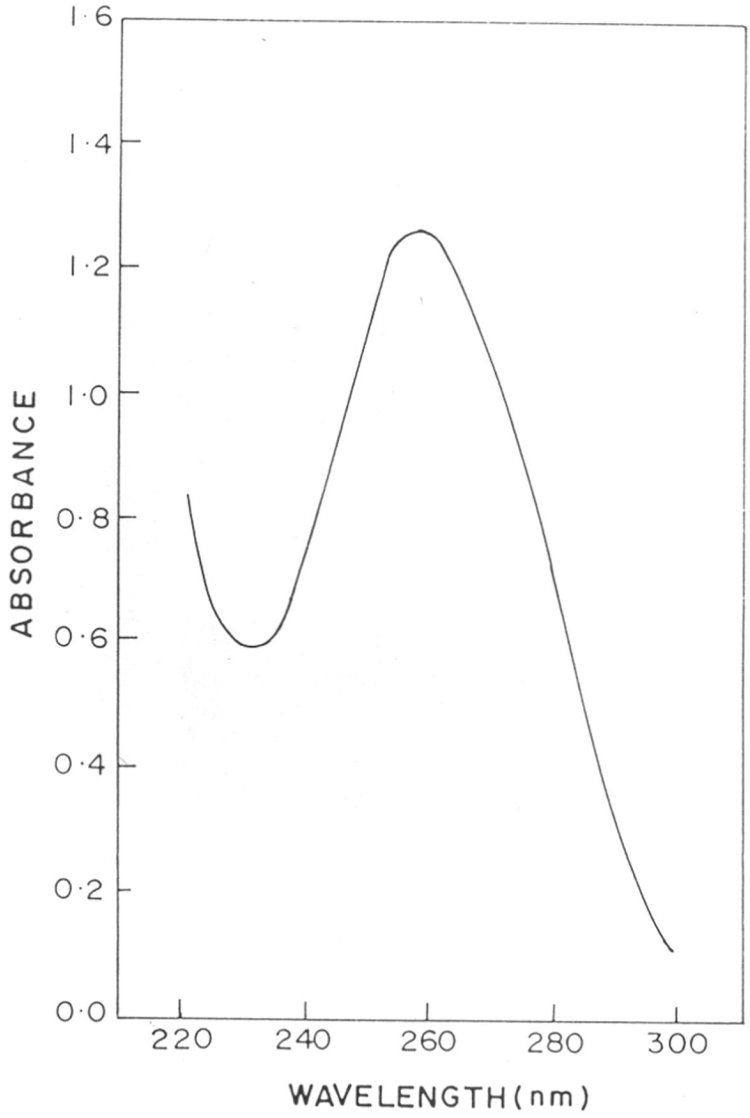


Figure **2.2** : Absorption spectrum of sponge gourd DNA. DNA was dissolved in 1 x SSC and scanned in the wavelength range of 220 to 300 nm.

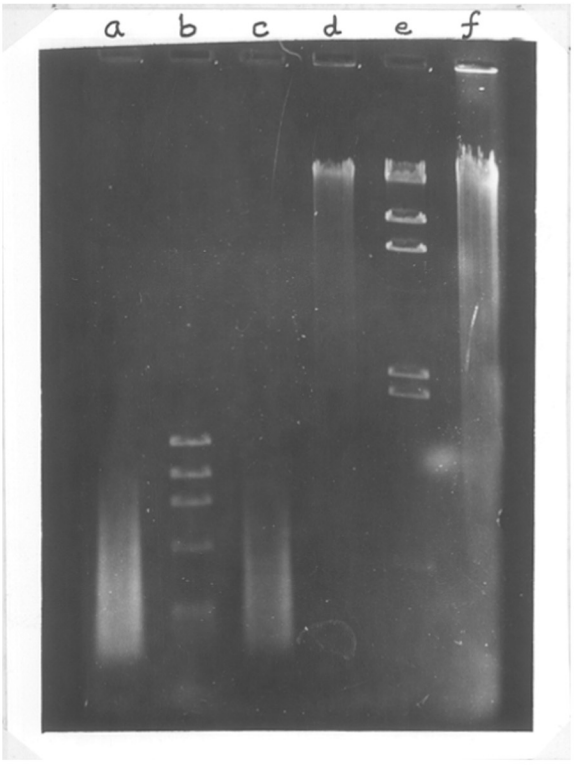


Fig. 2.3 : Agarose slab gel electrophoresis of sonicated ivy gourd (lane a), sponge gourd (lane c) DNAs and unsonicated ash gourd (lane d) and ridge gourd (lane f) DNAs. ϕ X 174 RF DNA digested with Hae III (lane b, 1342, 1078, 872, 606, 310 base pairs) and DNA digested with Hind III (lane e, 23300, 9500, 6400, 4200, 2200, 1800 base pairs) were used as molecular weight markers.

Thermal denaturation

Thermal denaturation was carried out using a Gilford 250 spectrophotometer equipped with thermoprogrammer (Model 2527), analog multiplexer (Model 6046) and automatic reference compensator (19). Thermal cuvettes were filled with 0.3 ml DNA solution (25-50 µg/ml) in 0.12 M sodium phosphate buffer, pH 6.8, and the absorbance was recorded at room temperature. Cuvettes were filled in such a way that the air-space was less than 10% of the solution volume and evaporation was prevented using teflon stoppers. The temperature was then raised to 98°C at the heat rate of 1°C/min. and the absorbance change occurring during the heating process from 62°C to 98°C was monitored, either with the recorder (Model 6051) or programmed printer. The hyperchromicity of the DNA samples was calculated using the formula

$$H = \frac{A_{260}\ (98^{\circ}C) - A_{260}\ (62^{\circ}C)}{A_{260}\ (98^{\circ}C)} \qquad (24,25)$$

where H is the hyperchromicity and A_{260} is the absorbance at 260 nm. The total hyperchromicity was normalized to 100% and the graph of percent hyperchromicity versus temperature was plotted. The melting temperature (Tm) of the DNAs was calculated as the temperature at which half of the total increase in absorbance is obtained. The G+C

content of DNAs was determined according to the formula

$$G+C = (Tm - 69.3)\ 2.44 \qquad (1,26).$$

The transition interval (ΔT) was calculated as

T_b °C (83% hyperchromicity) - T_a °C (17% hyperchromicity) (2,3).

DNAs from E.coli, calf thymus and wheat were used as standards for comparison in these studies.

High resolution thermal denaturation analysis

For such analysis, DNAs were denaturated at a heat rate of 0.25°C/min. and absorbance values were recorded at 0.1°C increment, from 62°C to 98°C. Thus the experimental curve consists of 300-400 experimental points depending on the melting range of DNA. Noise levels were reduced with a large slit width (1 mm) together with a high initial DNA concentration (50 µg/ml). Data smoothing and differentiation were performed by fitting ten contiguous data points at a time to a simple polynomial by the least squares method. A linear function was used as a fitting polynomial (14). The differential curves were obtained by plotting the temperature derivative of the absorbance of DNA against temperature (dA/dT vs T). To check the reproducibility of the procedure, at least five DNA preparations, each in duplicate, were melted and the resulting curves were analyzed separately. Fitting 10 points at a time, each indicating 0.1°C increment, represents a temperature rise of 1°C,

which is expected to resolve individual regions melting below and above the average Tm of the DNA. The limit of resolution of this analysis is 0.001 absorbance, and peaks separated by an interval as narrow as 0.4°C (approx. 1% G+C) could be detected reproducibly.

RESULTS

The melting data obtained by denaturing all the DNAs at 1.0°C increments are depicted in Figure 2.4 and summarized in Table 2.1. All the melting curves are smooth, sharp and monophasic. The Tm values range from 83.2 to 85.7°C. The G+C contents as calculated from the Tm values fall in the range of 34 to 40% as in most other higher plant species (12). The G+C contents of E. coli, calf thymus and wheat DNAs compare well with the reported values (3,8,9). The ΔT (transition interval), which is an indication of base composition heterogeneity in DNA, ranges from 6 to 8°C. This range is narrow to moderate as compared to that in some higher plants (3). The hyperchromicities fall in the range of 25 to 27.5%.

The first derivative melting profiles reveal several details about the heterogeneity in distribution of components differing in base composition (Figure 2.5). The derivative profiles of DNAs of E. coli, calf thymus and wheat were also obtained under identical conditions and found to compare well with the reported data (3,8,9). All the four Cucurbitaceae plant species show 17-22 melting components

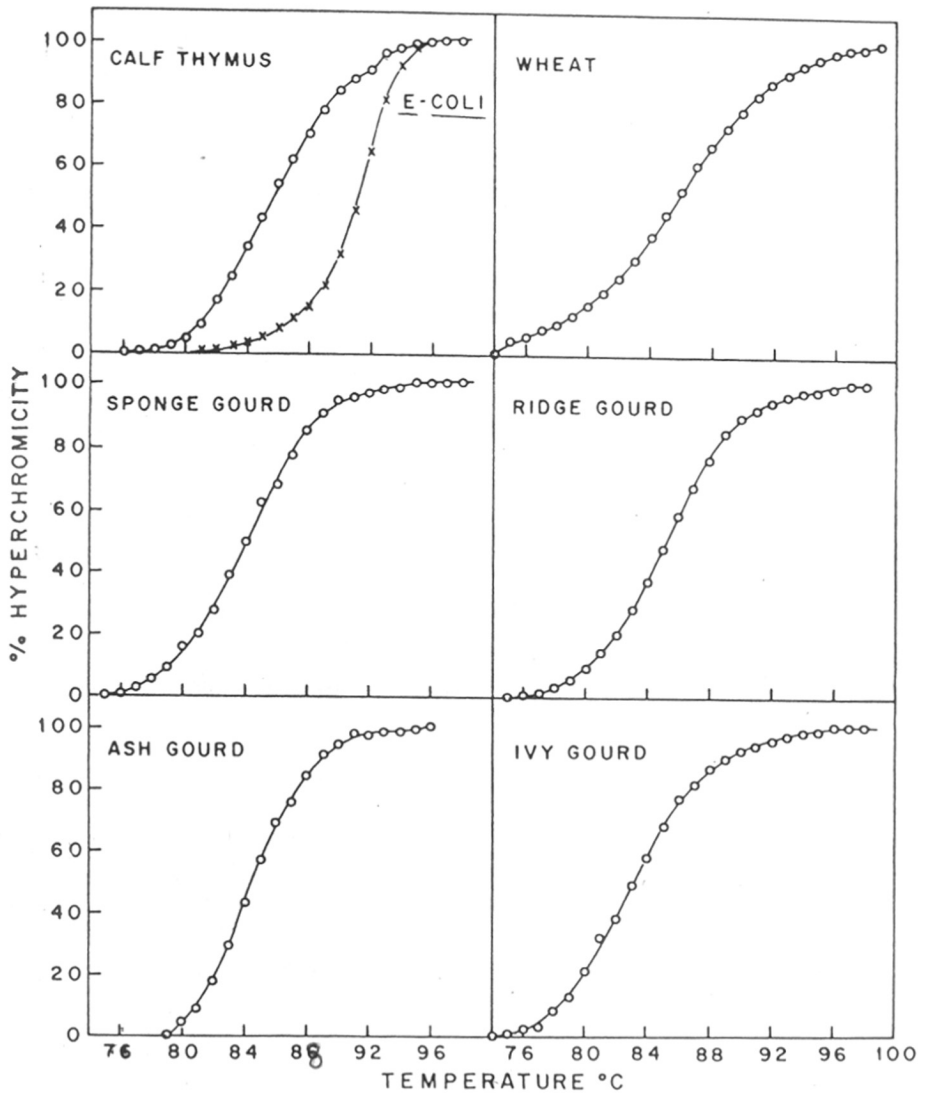


Figure 2.4 : Thermal denaturation profiles of Cucurbitaceae DNAs. High molecular weight DNAs ($>$10 kilo base pairs) in 0.12 M sodium phosphate buffer (pH 6.8) were melted in a Gilford 250 spectrophotometer at a heat rate of $1^{o}C$/min. and absorbance values were recorded at an interval of $1^{o}C$.

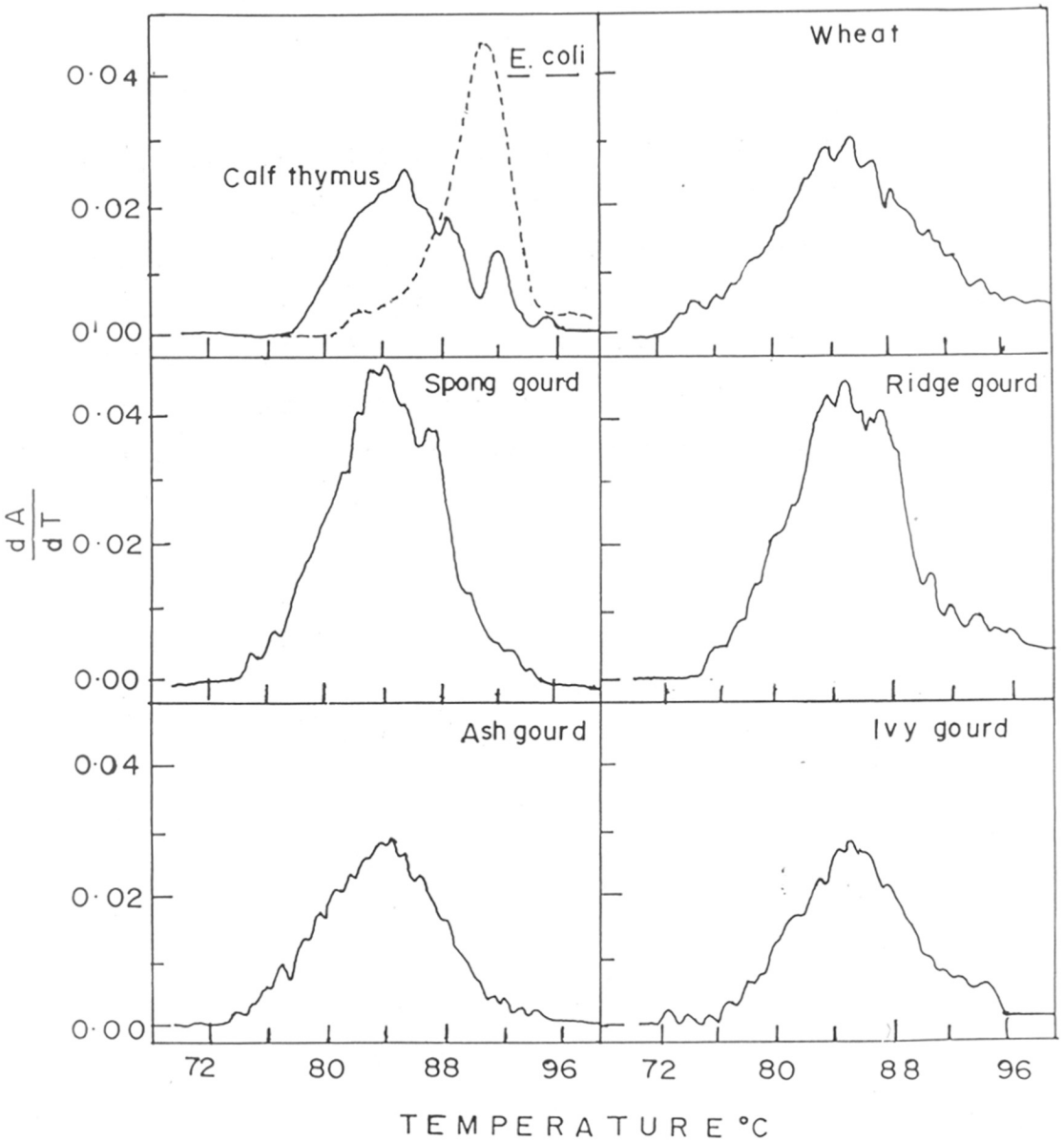
Table 2.1 Melting data of native unsheared DNAs

Species*	Tm (°C)	G+C content (%)	Hyper-chromicity (%)	ΔT °C (°C)
E. coli (6)	90.83±0.96	52.54±2.34	21.30±0.81	4.60±0.23
Calf thymus (5)	86.18±0.41	41.19±1.00	24.89±1.68	7.36±0.33
Wheat (3)	85.56±0.12	39.69±0.28	24.34±0.29	10.67±0.23
Sponge gourd (7)	84.39±0.23	36.83±0.56	28.82±0.78	7.16±0.60
Ridge gourd (6)	85.67±0.23	39.94±0.56	27.56±0.41	7.53±0.61
Ash gourd (7)	84.21±0.46	36.39±1.13	25.27±1.07	6.39±0.44
Ivy gourd (8)	83.24±0.46	34.01±1.12	26.57±1.02	7.96±0.39

* Figures in parentheses indicate the number of experiments carried out in each case.

Fig. 2.5 : First derivative melting profiles of unsheared DNAs of four Cucurbitaceae species and of E.coli, calf thymus and wheat. All the DNA melting experiments were carried out in 0.12M sodium phosphate buffer (pH 6.8). The DNA concentrations used were 50-60 µg/ml. All the DNAs were denatured at a heat rate of 0.25°C/min. and the absorbance values were recorded at 0.1°C increments.

Figure 2.5


E. coli
Calf thymus
Wheat
Spong gourd
Ridge gourd
Ash gourd
Ivy gourd
0·04
0·02
0·00
dA/dT
72
80
88
96
TEMPERATURE °C

which can be arbitrarily divided into three groups. The first group includes maxima peaks at the Tm and peaks that are in the range of ± 1°C on either side of the Tm. The second group comprises early melting, A+T rich components, upto the melting maxima, while the third group represents late melting, G+C rich components melting at temperatures higher than the melting maxima.

In all the species, two to three maxima peaks which coincide with the mean melting temperature are observed. While the melting starts from 74-76°C in sponge gourd, ridge gourd and ash gourd, in ivy gourd it starts much earlier (72°C), and two very early melting components (72.5°C, 73.0°C) are observed. All the species show 7 to 9 A+T rich components and 7 to 8 G+C rich components. Thermal denaturation is complete by 95°C in sponge gourd, ash gourd and ivy gourd but continues upto 97°C in ridge gourd.

The derivative melting profiles of sponge gourd and ridge gourd are very similar and are skewed on G+C rich side, with distinct peaks between 87.5 and 88.0°C. The melting profile of ridge gourd, however, can be distinguished by its high melting component at 96.6°C. The fine melting profiles of ash gourd and ivy gourd are symmetrical, with apparently equal distribution of A+T rich and G+C rich components on either side of the melting maxima. The profile of ivy gourd, however, is distinguished by the presence of the two early melting components.

In order to determine the number of components that are shared among the four species, the positions of all the components are indicated in Figure 2.6. A temperature range of 0 to 0.6°C was considered to estimate the number of shared components and these data are summarized in Table 2.2. In the temperature range of 0.6°C (1.5% G+C), 12 melting components are shared by all the four species, six each on the A+T rich and G+C rich sides, of the melting profiles. The number of common melting peaks is reduced to 11, 10, 7, 5 and 0 as the range for comparison is reduced to 0.5°C, 0.4°C, 0.3°C, 0.2°C and 0.1°C temperature difference (1.25%, 1.00%, 0.75%, 0.5% and 0.25% G+C, respectively). In the two Luffa species, fourteen out of seventeen components are common at a temperature difference of 0.6°C. Out of these, as many as eight are common even at 0.1°C difference and three appear at identical positions.

DISCUSSION

Since fine melting analysis is a very sensitive method, several precautions were taken to avoid artifacts. High molecular weight DNA preparations (molecular weight > 10 kilo base pairs), free from detectable levels of polysaccharides, RNA and single stranded DNA were used. Experimental conditions such as salt concentration (0.18M Na^+), and DNA concentrations (50-60 µg/ml) were

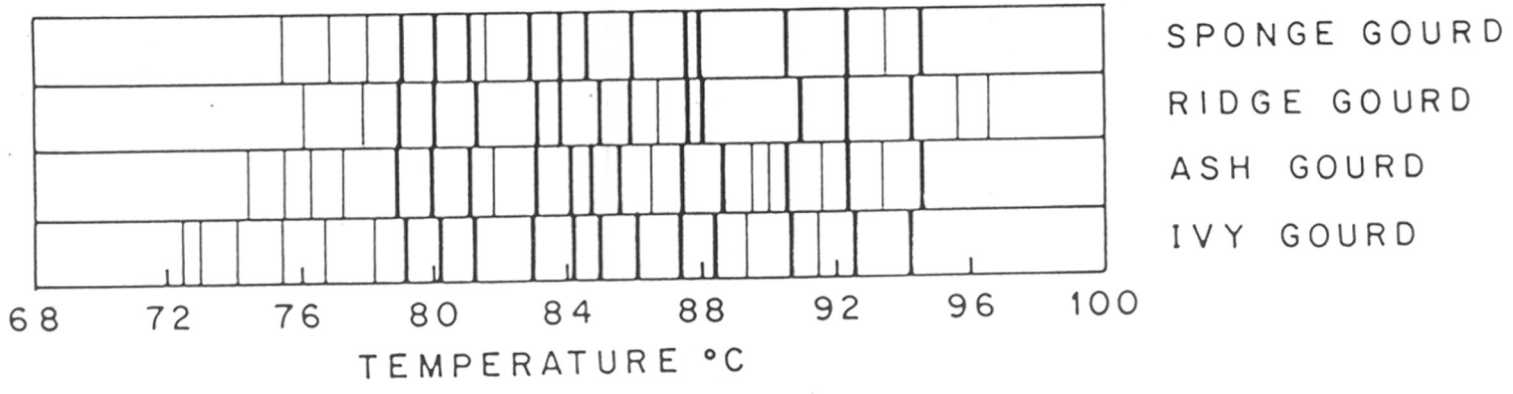


Figure 2.6 : The positions of melting components observed in the derivative profiles of four Cucurbitaceae species, shown as light lines. The bold lines indicate the components that are shared by all the four species within a temperature range of 0.6°C.

Table 2.2 The number of melting components shared by the four Cucurbitaceae species

Temperature range (°C)	0.6	0.5	0.4	0.3	0.2	0.1	0.0
G+C content range (%)	1.46	1.22	0.98	0.73	0.49	0.24	0.00
Number of components shared by all the four species	12	11	10	7	5	-	-
Number of components shared by two *Luffa* species	14	13	13	11	10	8	3

kept constant. The number and position of all the components in every DNA preparation were reproducible in the case of all the DNAs.

From Figure 2.5, it is clear that each species shows a polyphasic derivative melting profile, and distinct species-specific differences are observed. The appearance of polyphasic melting profiles in eukaryotic DNAs is considered to be the effect of satellite and repetitive families. The bacteriophage DNAs are made of unique sequences and show the presence of 'thermalities' melting in very narrow intervals (9). However, as the genome size increases, the overlap between the neighbouring sequences increases and this results in smooth, monophasic melting observed in the bacterial DNAs, which are also composed of only unique sequences (8). This should also be true for the higher plants and animals, if their genomes are to consist of only unique sequences. Thus the multiphasic melting profiles of higher organisms could be arising from the different repetitive sequences in their genomes. Each repetitive family in general should have a distinct base composition, and behave as a 'block' in the DNA melting (9). The technique thus allows identification of major blocks of repetitive DNA, even if they are interspersed with other DNA sequences, and therefore not detectable in cesium chloride gradients. The satellite sequences studied so far

in Cucurbitaceae are G+C rich (27,28). The present four species do not show any such distinct high melting component. However, our work on other species of Cucurbitaceae has shown that satellite sequences can be resolved very effectively (29). Considering that the present technique allows the qualitative determination of repetitive families (by their melting temperatures), it appears that the repeated sequences of the two Luffa species, sponge gourd and ridge gourd, are very similar. On the other hand, they are distinctly different from those of ash gourd and ivy gourd.

It is necessary to further complement this qualitative analysis by quantitative studies. Attempts were made to resolve these derivative profiles into a number of Gaussian components, to compare the areas under each, and to determine the proportion of different repetitive families in the genomes. However, the extremely high number of melting maxima, and the tremendous overlap of their areas has made it impossible to develop a suitable programme for such resolution. Personal correspondence with Dr. R.D. Blake (Department of Biochemistry, University of Maine, Orono), Dr. C.E. Cuellar (Carnegie Institution of Washington, Stanford, California) and Dr. A. Wada (Department of Physics, University of Tokyo, Japan) has indicated that it is possible to resolve derivative

80

profiles with 5-9 melting maxima fairly well; however, it is not possible to resolve the profiles of higher organisms characterized by numerous peaks satisfactorily. Attempts are being made at present to overlap the profiles and calculate the % area under the curves that is shared and not shared by the species, and thus derive the genetic distance between them.

REFERENCES

1. Marmur, J. and Doty P. (1962) J.Mol.Biol. 5, 109-118.

2. Mahler, H.R. and Dutton, G. (1964) J.Mol.Biol. 10, 157-175,

3. Huguet, T. and Jouanin, L. (1972) Biochim.Biophys.Acta 262, 431-440.

4. Ansevin, A.T., Vizard, D.L., Mandel, M.L. and Dean, D.H. (1983) Biochim.Biophys.Acta 740, 127-133.

5. Lyutchenko, Y.L., Vologadskii, A.V., Frank Kamenetskii, M.D. (1978) Nature (Lond.) 271, 28-31.

6. Tachibana, H., Wada, A., Gotoh, O. and Takanami, M. (1978) Biochim.Biophys.Acta 517, 319-328.

7. Vizard, D.L. and Ansevin, A.J. (1976) Biochemistry 15, 741-750.

8. Blake, R.D. and Lefoley, S.G. (1978) Biochim.Biophys. Acta 518, 233-246.

9. Mayfield, J.E. (1977) Biochim.Biophys.Acta 477, 97-101.

10. Szecsi, A. (1979) Acta Microbiol.Acad.Sci. Hung. 26, 337-347.

11. Szecsi, A. and Dobrovolszky, A. (1980) Nucleic Acids Res. 8, 2517-2525.

12. Pivek, L., Horska, K., Vitek, A. and Daskocil, J. (1974) Biochim.Biophys.Acta 340, 199-206.

13. Guttman, T., Vitek, A. and Pivec, L. (1977) Nucleic Acids Res. 4, 285-297.

14. Wada, A., Yabuki, S. and Husimi, Y. (1980) In: CRC Critical Review in Biochem. 9, 87-144.

15. Yabuki, S., Gotoh, O. and Wada, A. (1975) Biochim.Biophys.Acta 395, 258-273.

16. Steinert, M. and Van Ansel, S. (1974) Biochim. Biophys.Res.Commun. 61, 1249-1255.

CHAPTER III

NONLINEAR REGRESSION ANALYSIS OF DNA REASSOCIATION KINETICS

17. Mascarenhas, A.F., Hendre, R.R., Nadgir, A.L., Ghugale, D.D., Godbole, D.A., Prabhu, R.A. and Jagannathan, V. (1976) Ind.J.Exp.Biol. 14, 604-606.

18. Marmur, J. (1961) J.Mol.Biol. 3, 208-218.

19. Ranjekar, P.K., Pallotta, D. and Lafantaine, J.G. (1976) Biochim.Biophys.Acta 425, 30-40.

20. Volkin, E. and Cohn, W.E. (1954) In: Methods Biochem. Anal. (Glick, D., ed.) Vol.1, Interscience Publishers Inc., New York, pp. 287-303.

21. Lowry, O.H., Rosenbrough, N.J., Farr, A.L. and Randall, R.L. (1951) J.Biol.Chem. 193, 265-275.

22. Darby, G.K., Jones, A.S., Kennedy, J.F. and Walker, R.T. (1970) J.Bacteriol. 103, 159-165.

23. Walbot, V.and Goldberg, R.B. (1979) In: Nucleic Acids in Plants (Hall T.C. and Davies, J.W., eds), Vol.I, CRC Press, Florida, pp. 3-40.

24. Zimmerman, J.L. and Goldberg, R.B. (1977) Chromosoma (Berl.) 59, 227-252.

25. Wimpee, C.F. and Rawson, J.R.Y. (1979) Biochim. Biophys.Acta 562, 192-206.

26. Mandel, M. and Marmur, J. (1968) In: Methods in Enzymol. (Grossman, L. and Moldave, K., eds.) Vol.12B, Academic Press, New York, pp. 195-206.

27. Reis, R.S., Timmis, J.N. and Ingle, J. (1981) Biochem. J. 195, 723-734.

28. Ranjekar, P.K., Pallotta, D. and Lafontaine, J.G. (1978) Can.J.Biochem. 56, 808-815.

29. Bhave, M., Lagu, M., Gadre, S.R. and Ranjekar, P.K. (1984) Indian J.Biochem.Biophys. 21, 181-184.

NONLINEAR REGRESSION ANALYSIS OF DNA REASSOCIATION KINETICS

SUMMARY

Employing the nonlinear regression analysis method, reassociation kinetics of DNAs from four Cucurbitaceae plant species have been analysed to determine (i) the number of DNA components reassociating with second order kinetics, (ii) proportion of repeated and single copy sequences, and (iii) frequency of repetition and kinetic complexity of different classes of DNA. The repeated DNA sequences account for 25 to 59% of the total DNA. In ivy gourd, the repetitive DNA content of 25% is the lowest reported so far in all the higher plant species. Only sponge gourd shows the presence of a fast-reassociating component with a reiteration frequency of 5.29×10^4, and a kinetic complexity of only about 700 base pairs. The slow component exhibits a low reiteration frequency of 9×10^1 to 3.2×10^2 and the total length of DNA included in this fraction (i.e. kinetic complexity) is 1.7×10^5 to 1.0×10^6 base pairs. The kinetic complexity of single copy fraction ranges from 8.37×10^7 to 1.69×10^8 base pairs. Though sponge gourd and ridge gourd belong to the same genus Luffa, they differ in their total repetitive and single copy DNA proportions. These variations in proportions, frequencies of reiteration and

kinetic complexities of different classes of DNAs of the four species may contribute to their interspecies incompatibility.

INTRODUCTION

The occurrence of repeated DNA sequences is a typical feature of all eukaryotic genomes. This conclusion is primarily based on DNA reassociation kinetics measurements. Various techniques such as hydroxyapatite chromatography and optical and fluorometric measurements are employed to study reassociation kinetics. Hydroxyapatite chromatography is the most convenient and widely used methodology for the fractionation of DNA on the basis of the reassociation rates of its components. Since strand separation produces a hyperchromic shift which is reversed as base pairs reform, reassociation may also be followed optically by monitoring the decrease in optical density (hypochromic shift) as a function of time. A curve of percentage reassociation of denatured DNA corresponding to different Cot values which are dependent on the concentration of DNA and period of incubation can be obtained by both the methods. From this Cot curve, the percentages of repetitive and nonrepetitive DNA and their reassociation rates can be calculated. Repetitive DNA fraction can also be resolved

into different subfractions and estimates of their proportions, frequencies of repetition and kinetic complexities can be obtained.

The only information available so far on reassociation kinetics in Cucurbitaceae species is that on the satellite DNAs of cucumber (1) and muskmelon (2). The two satellites of cucumber were found to differ from each other in their reassociation behaviour. The satellite I was found to be composed of a major (75%) rapidly reassociating and a minor (25%) slowly reassociating fraction. In case of satellite II, 40% of the DNA reassociated fast, while the remaining 60% reassociated slowly, suggesting the presence of many components with different complexities. The muskmelon satellite was also found to be heterogeneous, with a fast reassocating component comprising 37% and a slower one comprising 44% of the satellite.

This chapter deals with the reassociation kinetics of the DNAs of the four cucurbits, namely, sponge gourd, ridge gourd, ash gourd and ivy gourd. We wish to find out 1) how many DNA components, each reassociating with second order kinetics, are present in the total DNA, 2) what is the proportion of total repetitive DNA, 3) what is the frequency of repetition and kinetic complexity of different repetitive DNA components.

MATERIALS AND METHODS

DNA isolation

DNAs were isolated from seedlings of sponge gourd, ridge gourd and ash gourd and fruits of ivy gourd as described in chapter II.

DNA shearing and sizing

Since the size of DNA fragments is an important parameter in the study of DNA reassociation kinetics (3), it was controlled by shearing the DNAs to small fragments. Native high molecular weight DNAs, dialysed against 0.12 M sodium phosphate buffer (pH 6.8) were fragmented to a modal length of 550 base pairs using a Sonic Oscillator (Bronwell Model, Biosonik III, 250 W, 20 Khz) fitted with a half-inch (green) probe. Prior to sonication, nitrogen gas was bubbled through the DNA solution for 5-10 min (4). Approximately 30 ml of the DNA solution was sonicated at a maximum intensity for 3 min at 1 min pulse and at an interval of 10-15 min. between successive pulses of sonication. The temperature of the DNA solution was strictly maintained below $4^{\circ}C$ by constant cooling in an ice bath to prevent denaturation due to heating. The average size of DNAs was determined before and after shearing as described in chapter II. The DNAs before sonication had a molecular weight of > 10 kilo base pairs,

while the sonicated DNAs gave a modal size of 550 base pairs, with sizes ranging between 400 to 600 base pairs.

Hydroxyapatite preparation

Hydroxyapatite was prepared essentially according to the procedure described by Tiselius et al. (5) and stored at 4°C with a few drops of chloroform. The extent of DNA recovery from each batch of hydroxyapatite was checked by loading known quantities of DNA. The recovery of DNAs was always more than 95%.

DNA reassociation kinetics

The major steps in a typical reassociation experiment are (a) strand separation by heat denaturation and (b) reassociation between complementary strands under appropriate conditions. Proper choice of incubation temperature (Tm -25°C) and salt concentration (0.12 M sodium phosphate buffer, pH 6.8; Na^+ 0.18 M) provides optimum conditions for reassociation. Two techniques were used to measure the progress of reassociation:

1) Optical reassociation: Since strand separation produces a hyperchromic shift (increase in absorbance) which is reversed when base pairs reform, reassociation may be followed by monitoring the decrease in optical density (hypochromic shift) as a function of time.

DNA samples, 550 base pairs long, in 0.12 M sodium phosphate buffer (pH 6.8) were denatured in a Gilford 250 spectrophotometer equipped with a thermo-programmer (Model 2527), analog multiplexer (Model 6046) and automatic reference compensator. Thermal cuvettes were filled with 0.3 ml DNA solution (concentration 25-50 µg/ml) and the temperature was raised to 98°C. The absorbance change occurring during the heating process from 62°C to 98°C was monitored at 260 nm. The hyperchromicity of DNA samples was calculated as given in chapter II. The DNA samples were maintained at 98°C for 5 min. to ensure complete strand separation and then quickly cooled to 62°C by circulating water. The drop in temperature from 98°C to 62°C took about 60-90 sec. The change in absorbance of DNA solutions was continuosly monitored with time. The midpoint between 98°C and 62°C (i.e. 80°C) was calculated according to Britten et al. (6) and was considered as the zero time. The increase in absorbance from 62°C to 98°C was taken as 100% hyperchromicity.

The Cot value at different time intervals is defined as the half product of absorbance of dissociated DNA x h(s) of incubation. ($Cot = \frac{A}{2} \times t\ (h)$) (7,8). Thus a Cot value of 1 is equivalent to incubation of a DNA solution of concentration 83 µg/ml (approx. 2 O.D)

for one hour (7). The percent reassociation at different Cot values was calculated according to the formula

$$\% R = 100 \times \frac{A_{260} \text{ of denatured DNA} - A_{260} \text{ at a given time}}{A_{260} \text{ of denatured DNA} - A_{260} \text{ of native DNA at } 62^{\circ}C}$$

Reassociation was monitored optically in the Cot range of 1.0×10^{-3} to 5.0×10^{0} M.sec/l.

2) Hydroxyapatite column chromatography: Hydroxyapatite assays are based on the ability of double-stranded DNA to bind to hydroxyapatite more tightly than single stranded DNA. Under certain conditions, such at 0.12 M sodium phosphate buffer (pH 6.8), at 60°C, single strands will pass through a column of hydroxyapatite while double strands are retained. Double stranded material can then be removed by increasing the phosphate concentration (e.g. to 0.4 M sodium phosphate buffer, pH 6.8), or by converting it to single strands by raising the column temperature. The minimum duplex length required for binding is usually considered to be about 50 base pairs (9).

Hydroxyapatite slurry was poured into double jacketed glass columns (15 x 2.5 cms) to a bed volume of approx 16 ml. The columns were equilibriated with 0.12 M sodium phosphate buffer (pH 6.8) and maintained at 62°C with a circulating water bath. The sonicated

samples (550 base pairs) ranging in concentration from 25-300 µg/ml (lower concentrations used for studies at lower Cot values) in 0.12 M sodium phosphate buffer (pH 6.8) were sealed in glass tubes and denatured by boiling in a water bath at 100°C for 10 min. These were then incubated to specific Cot values at 62°C and loaded on hydroxyapatite columns. Single stranded material was eluted with 0.12 M sodium phosphate buffer (pH 6.8) and double stranded material with 0.4 M sodium phosphate buffer (pH 6.8).

The relative amounts of single and double stranded DNA were determined by absorbance at 260 nm in a Shimadzu double beam spectrophotometer (Model 210A) assuming 1 absorbance unit (A_{260}) is equal to 50 µg DNA/ml. The extent of reassociation was calculated using the formula

$$\% R = \frac{\text{µg of DNA eluted with 0.4 M sodium phosphate buffer}}{\text{µg of DNA eluted with 0.12 M sodium phosphate buffer} + \text{µg of DNA eluted with 0.4 M sodium phosphate buffer.}} \times 100 \quad (10)$$

In experiments at Cot values higher than 1×10^{2} M x sec/l, DNA samples were incubated in 0.36 M sodium phosphate buffer (pH 6.8) in order to accelerate the rates of reassociation. Samples incubated in 0.36 M sodium phosphate buffer (pH 6.8) were diluted to 0.12 M sodium phosphate buffer (pH 6.8) before loading on to the

hydroxyapatite columns. A corresponding correction was applied to the reassociation rate (6). The reassociation kinetics of E.coli DNA were carried out under identical conditions and used as standards.

Computer analysis of reassociation kinetics data

The reassociation of a eukaryotic DNA is generally plotted on log scale, since it facilitates handling data obtained at several different DNA concentrations, and since eukaryotic DNAs may reassociate over 5-10 decades. The values of reassociation as a function of Cot obtained by both optical and hydroxyapatite methods were plotted together on a semilogarithmic paper. The individual components of the reaction are usually not well separated and various computer fitting procedures can be used to model these curves into two or three (or more) ideal second order components.

Second order reassociation kinetics can be described by the equation

$$\frac{C}{Co} = \frac{1}{1 + K\,Cot}$$

where Co = initial DNA concentration in moles of nucleotides per litre.

C = concentration of free single strands after t seconds of reassociation

K = rate constant of reassociation.

When $C/Co = 1/2$, $K = 1/Cot\ 1/2$.

The least squares analysis of the reassociation data was carried out by employing a standard optimization subroutine STEPIT (Subroutine STEPIT written by J.P. Chandler and distributed by QCPE, Indiana University, Bloomington, Indiana, USA). The error is given by the equation

$$\text{Error} = \sum \left\{ \left(\frac{C}{C_o}\right)_{\text{expt.}} - \left(\frac{C}{C_o}\right)_{\text{fit}} \right\} \quad \ldots\ 1$$

and was minimized by a direct search procedure. The form for C/Co was assumed to be

$$\frac{C}{C_o} = T + \sum \frac{F_i}{1 + K_i C_o t} \quad \ldots\ 2$$

where $\frac{C}{C_o}$ = fraction denatured

T = fraction of genome that fails to reassociate.

F_i = fraction of genome reassociating in the i^{th} component

K_i = rate of reassociation for the i^{th} component.

The parameters T, Fi and Ki were all allowed to free float to minimize the error in equation 1.

RESULTS

Figure 3.1 shows the reassociation kinetics of 550 base pairs long fragments of all the four DNAs. The data points are obtained using a combination of optical reassociation method (upto Cot 5 x 10^0 M x sec/l) and hydroxyapatite method (Cot 0.1 M x sec/l onwards). The curves drawn through the data points are obtained by a nonlinear least squares analysis of the data, assuming second order kinetics, and allowing all parameters to free float (11). The results of such analysis are summarized in Table 3.1.

From the rate constant data in Table 3.1, the proportions of total repetitive DNAs are estimated to be 51% in sponge gourd, 59% in ridge gourd, 48% in ash gourd and 25% in ivy gourd. A detailed examination of the data in Figure 3.1 and Table 3.1 reveals the presence of three repetitive components in sponge gourd and two components in the remaining three species. The first fraction is classified as very fast reassociating component and includes fold back sequences and/or satellite sequences and classes of very highly repeated DNA (12). Based on the frequency of repetitition, the remaining two fractions have been classified as fast reassociating and slow reassociating components.

Fig. 3.1 : Reassociation kinetics of 550 base pairs long DNA of four Cucurbitaceae species, by optical (0 - 0) and hydroxyapatite (● - ●) method. The solid line through the data points represents the least squares fit, allowing all the parameters to free float. The lower dashed curves represent the predicted reassociation kinetics of pure components. All the DNA reassociation experiments were carried out at 62°C and the DNA concentrations used were 20-50 µg/ml upto Cot 5.0 M x sec/l and 200-300 µg/ml in experiments higher than Cot 5.0 M x sec/l. The molarity of sodium phosphate buffer was raised from 0.12 M to 0.36 M for experiments at Cot > 100.0 M x sec/l and a correction to the reassociation rate was made according to Britten et al (6).

a : Sponge gourd (Root Mean Square error 0.0457) and E.coli (Root Mean Square error 0.0408, ▲ - ▲).

b : Ridge gourd (Root Mean Square error 0.0333).

c : Ash gourd (Root Mean Square error 0.0552)

d : Ivy gourd (Root Mean Square error 0.0502).

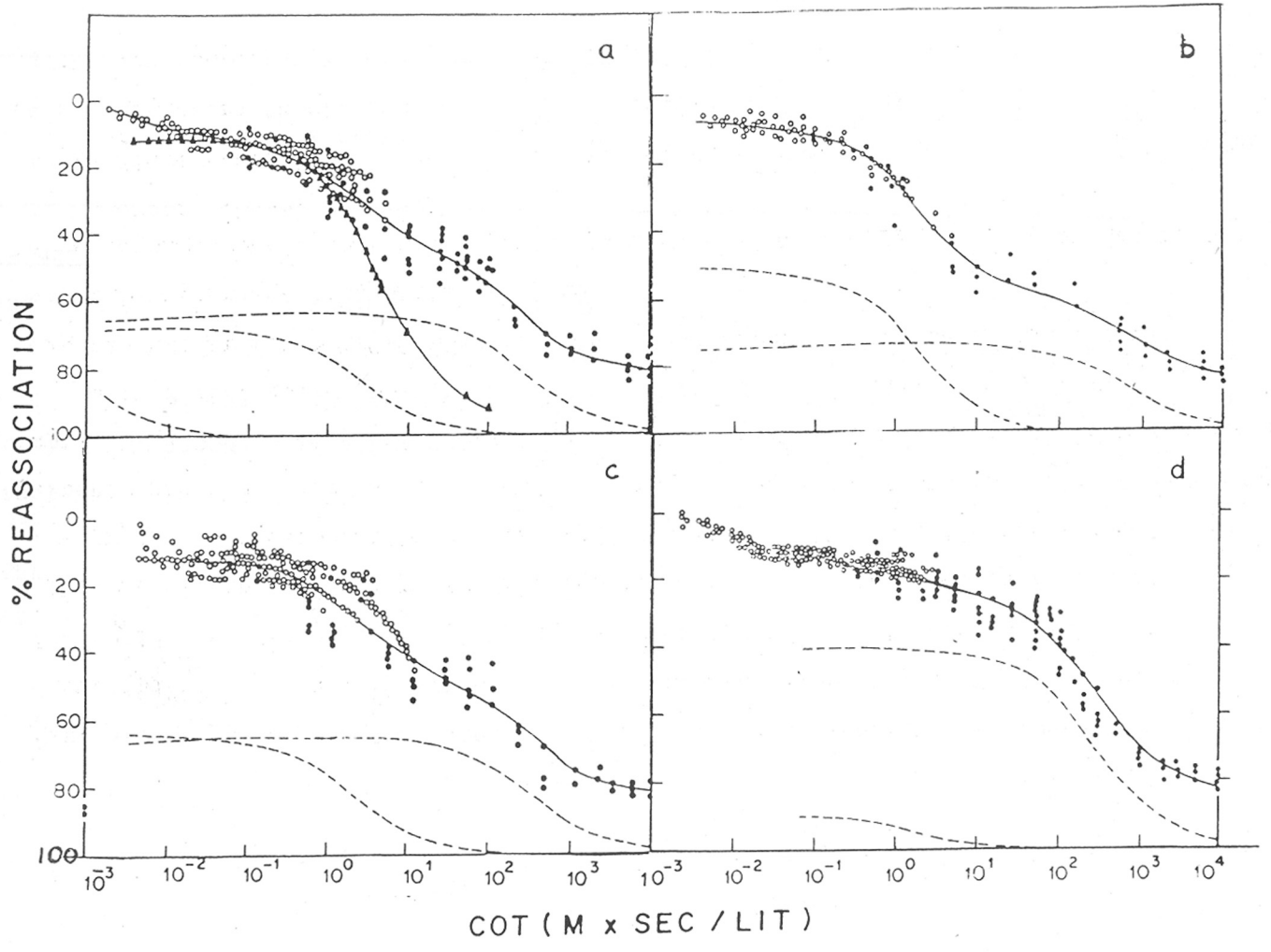

a
b
c
d
% REASSOCIATION
0
20
40
60
80
100
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
COT (M x SEC / LIT)

Table 3.1 Kinetic analysis of four Cucurbitaceae genomes

Species	Range of reassociation Cot M x sec/l	Fraction of the genome	Cot 1/2 observed [a]	Cot 1/2 pure [b]	K pure [c]	Kinetic complexity in base pairs [d]	Number of copies per 1C genome
Sponge gourd							
Very fast	$< 1.80 \times 10^{-3}$	0.02	-	-	-	-	-
Fast	1.80×10^{-3} to 4.20×10^{-1}	0.16	4.40×10^{-3}	7.00×10^{-4}	1.43×10^{3}	7.00×10^{2}	5.29×10^{4}
Slow	4.20×10^{-1} to 5.00×10^{1}	0.33	2.56×10^{0}	8.50×10^{-1}	1.18×10^{0}	8.50×10^{5}	9.08×10^{1}
Single copy	5.00×10^{1} to 1.00×10^{4}	0.36	2.32×10^{2}	8.37×10^{1}	1.20×10^{-2}	8.37×10^{7}	1
Final fraction unreassociated	$> 1.00 \times 10^{4}$	0.17	-	-	-	-	-
Ridge gourd							
Very fast	$< 4.40 \times 10^{-3}$	0.08	-	-	-	-	-
Slow	4.40×10^{-3} to 5.00×10^{1}	0.51	1.96×10^{0}	1.00×10^{0}	1.00×10^{0}	1.00×10^{6}	3.19×10^{2}
Single copy	5.00×10^{1} to 1.00×10^{4}	0.27	6.25×10^{2}	1.69×10^{2}	5.90×10^{-3}	1.69×10^{8}	1
Final fraction unreassociated	$> 1.00 \times 10^{4}$	0.15	-	-	-	-	-

Table 3.1 (contd.)

<u>Ash gourd</u>							
Very fast	$< 4.00 \times 10^{-3}$	0.12	-	-	-	-	-
Slow	4.00×10^{-3} to 2.50×10^{1}	0.36	1.85×10^{0}	6.70×10^{-1}	1.49×10^{0}	6.70×10^{5}	1.42×10^{2}
Single copy	2.50×10^{1} to 1.00×10^{4}	0.35	2.63×10^{2}	9.21×10^{1}	1.10×10^{-2}	9.20×10^{7}	1
Final fraction unreassociated	$> 1.00 \times 10^{4}$	0.17	-	-	-	-	-
<u>Ivy gourd</u>							
Very fast	$< 5.00 \times 10^{-2}$	0.14	-	-	-	-	-
Slow	5.00×10^{-2} to 1.00×10^{1}	0.11	1.54×10^{0}	1.70×10^{-1}	5.88×10^{0}	1.70×10^{5}	1.76×10^{2}
Single copy	1.00×10^{1} to 1.00×10^{4}	0.59	2.70×10^{2}	1.59×10^{2}	6.30×10^{-3}	1.59×10^{8}	1
Final fraction unreassociated	$> 1.00 \times 10^{4}$	0.16	-	-	-	-	-

a : Values obtained from the respective reassociation curves (Fig. 3.1, a to d)

b : Cot 1/2 values observed x fraction of genome

c : K pure = $(\text{Cot } 1/2 \text{ pure})^{-1}$

d : Values obtained by comparing the Cot 1/2 (4.5 M x sec/l under our experimental conditions) and genome size (4.5×10^{6} base pairs) of <u>E.coli</u> (21)

e : Cot 1/2 values of unique DNA divided by Cot 1/2 values of the given fraction.

The proportion of the very fast DNA component is 2% to 14% and it includes DNA sequences reassociating by Cot 1.80×10^{-3} to 5.00×10^{-2} M x sec/l. Since we do not have the data points less than the above Cot range, this DNA fraction could not be resolved further. The proportion of fold back sequences in plants is approximately 2 - 7% (13,14). It is, therefore, likely that the substantial proportion of this fraction may consist of satellite DNA sequences. An interesting situation is observed in ivy gourd, where the very fast fraction could not be fitted with second order reassociation kinetics, in spite of the large number of data points. This fraction, comprising more than half of the total repetitive DNA (25%) of ivy gourd may, therefore, be composed of sequences reassociating with first order kinetics.

The fast component, present only in sponge gourd, accounts for 16% of the total DNA. It consists of a relatively short stretches of DNA, with a high reiteration frequency of 5.29×10^{4}. The presence of satellite DNA has been reported in this species (15). The proportion of very fast reassociating DNA in sponge gourd is only 2%. It is, therefore, likely that the fast DNA component may actually include satellite sequences in this species.

The slow component is observed in all the four species and comprises 11-51% of the genome. It consists of

DNA sequences with a rather low copy number in the range of 9.08×10^{1} to 3.19×10^{2}.

The single copy DNA sequences represent 27 to 59% of the genomes with rate constants of 5.9×10^{-3} to 1.20×10^{-2} and kinetic complexity of 8.37×10^{7} to 1.69×10^{8} base pairs.

About 15-17% of the genome fails to reassociate even at Cot 1×10^{4} M x sec/l. This may be due to denaturation resulting from long incubations (12), or due to presence of sequences which require still longer times to reassociate (16)

DISCUSSION

The basic information about the overall genomic heterogeneity and complexity is a prerequisite for further studies such as DNA sequence organization and interspecies DNA sequence homologies.

Since the accuracy of DNA reassociation kinetics studies depends upon a number of parameters, care was taken to maintain the optimum conditions for reassociation.

Reassociation kinetics of E.coli DNA were monitored under identical conditions and the rate of reassociation of E.coli DNA was found to be comparable to the reported values (17,18). Using a computer fitting procedure, reassociation curves have been resolved into two or three ideal curves. However, it is possible that each of these

components may be a mixture of several different repetitious families differing in copy number. The values of reiteration frequencies in Table 3.1 are, therefore, average estimates.

It is generally observed that hydroxyapatite fractionation often gives over estimates of the repeated DNA sequences. This is essentially because optically monitored reassociation as against that by hydroxyapatite measures the actual fraction of reassociated duplexes. However, we have observed that the extent of over-estimation is very limited in the lower Cot values and hence optical and hydroxyapatite method reassociation results have been analysed together.

The proportion of repetitive DNA component in higher plants varies from 35% to 90%, with the lowest reported value of 35% in Vigna radiata (13). In the present work, ivy gourd contains 25% repetitive DNA and hence is the first higher plant species reported so far with such a low level of repetitive DNA.

The repetitive DNA component of higher plant species shows a very heterogeneous spectrum of repetition frequencies (a few to several hundred thousand copies) and kinetic complexity (10^2 to 10^6 base pairs) (19, 20). Based on the frequency of reiteration, the repetitive DNA is generally classified into fast and slow components. In the present

studies, the fast component is present only in sponge gourd while the slow component is the major repetitive component in the remaining three species. Long repetitive DNA sequences generally constitute a major fraction in plant chromosomes. The functional significance of these DNA sequences is not yet known. It is, however, likely that they play an important role in determining the modes of DNA sequence organization in plants.

The Cucurbitaceae species under present investigation show about two-fold variation in the content of repetitive and single copy DNA sequences. Large differences are also noticed in the percentage, copy number and kinetic complexity of different repetitive fractions. Though sponge gourd and ridge gourd belong to the same genus Luffa, the fast component with a reiteration frequency of 5.29×10^4 is present only in sponge gourd. Moreover, the kinetic complexity of single copy DNA in ridge gourd is twice as much as that of sponge gourd. It is, therefore, possible that these molecular variations may contribute to species isolation in Cucurbitaceae which otherwise is extremely conservative in its morphological and cytological features.

REFERENCES

1. Ranjekar, P.K., Pallotta, D. and Lafontaine, J.G. (1978) Can.J.Biochem. 56, 808-815.

2. Bendich, A.J. and Taylor W.C. (1977) Plant Physiol. 59, 604-609.

3. Marmur, J., Rownd, R. and Schildkraut, C.L. (1963) Progr.Nucl.Acid Res. 1, 231-300.

4. Walker, P.M.B. and McLaren, A. (1965) J.Mol.Biol. 12, 394-409.

5. Tiselius, A., Hjerten, S. and Levin, O. (1956) Arch.Biochem.Biophys. 65, 132-155.

6. Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) In: Methods in Enzymol. (Grossman, G.L. and Moldave, K. eds). Vol. 29E, Academic Press, New York, pp. 363-418.

7. Britten, R.J. and Kohne, D.E. (1968) Science 161, 529-540.

8. Kohne, D.E. and Britten, R.J. (1971) In: Procedures in Nucleic Acids Res. (Cantoni, G.L. and Davies, D.R., eds.) Vol. 2, pp. 500-512.

9. Wilson, D.A. and Thomas, C.A. (1973) Biochim.Biophys.Acta 331, 333-340.

10. Ranjekar, P.K. and Murthy, M.R.V. (1973) J.Neurochem. 20, 1257-1264.

11. Gupta, V.S., Godre, S.R. and Ranjekar, P.K. (1981) Biochim.Biophys.Acta. 656, 147-154.

12. Cullis, C.A. (1981) Biochim.Biophys.Acta. 652, 1-15.

13. Murray, M.G., Palmer, J.D., Cuellar, R.E. and Thompson, W.F. (1979) Biochemistry 18, 5259-5266.

14. Wenzel, W. and Hemleben, V. (1979) Plant Syst.Evol. Suppl. 2, 29-40.

15. Ingle, J., Pearson, G.G. and Sinclair, J. (1973) Nature New Biol. 242, 193-197.

16. Pearson, W.R., Smith, S.L., Wu, J.R. and Bonner, J. (1978) Plant Physiol. 62, 112-115.

17. Wimpee, C.F. and Rawson, J.R.Y. (1979) Biochim. Biophys.Acta 562, 192-206.

18. Zimmerman,G.L.and Goldberg, R.B. (1977) Chromosoma (Berl.) 59, 227-252.

19. Thompson, W.F. and Murray, M.G. (1981) The nuclear genome: Structure and function, In: Marcus, A (ed.) The Biochemistry of Plants, A Comprehensive Treatise, Vol.6, Academic Press, New York, pp. 1-81.

20. Wenzel, W. and Hemleben, V. (1982) Plant Syst.Evol. 139, 209-227.

21. Cairns, J. (1963) Cold Spring Harbor Symp. Quant.Biol. 28, 43.

CHAPTER IV

SEQUENCE DIVERGENCE AND NATURE OF REPEATED DNA SEQUENCES

SEQUENCE DIVERGENCE AND NATURE OF REPEATED DNA SEQUENCES

SUMMARY

Thermal denaturation studies of repetitive sequences isolated at different Cot values have shown the presence of a moderate sequence divergence of 2-8%. Reassociation kinetics of isolated total repetitive DNAs show absence of any discrete classes of fast repeats and presence of a large moderately repetitive component. Reassociation kinetics of total DNAs were studied at different temperatures ranging from Tm -30°C to Tm -5°C for allowing the fractionation of repeats according to their relatedness. With an increase in reassociation temperature, the proportion of the repetitive fraction decreases in all the four species. However, an increase is observed in their reassociation rates. Shorter but more perfect duplexes, with high copy number, reassociate at the increased stringency of reassociation. This suggests that the repeat families are essentially homogeneous in nature, in spite of the vast differences observed in their properties like proportions, kinetic complexities and reiteration frequencies (Chapter III). The high resolution thermal denaturation profiles of total repetitive DNAs isolated at different temperatures also

support this idea. From the comparison of proportions of 'fossil repeats' and 'true single copy' sequences, it appears that the genome of sponge gourd is 'turning over' at a much faster rate than that of ridge gourd, ash gourd and ivy gourd.

INTRODUCTION

DNA reassociation kinetics yields valuable information regarding the structure and organization of eukaryotic genomes. Most of the reassociations are performed at a temperature about 25°C below the melting temperature of native DNA, a criterion at which optimal reassociation rates are generally obtained. However, the repetitive DNA populations often show extreme kinetic heterogeneity (1,2) and different degrees of sequence divergence, and thus comprise a frequency/divergence continuum rather than belonging to discrete classes. It is possible to fractionate the repetitive sequences according to their relatedness by studying reassociation kinetics at different temperatures. When the reassociation temperature is reduced below that of a standard criterion (such as Tm -25°C) to a more permissive criterion, more mismatched duplexes are allowed to form to give an apparently higher proportion of repeated sequences. In contrast, at higher reassociation temperatures, only the

very well-matched, perhaps conserved, sequences reassociate. Thus the terms repetitive and single copy sequences are largely operational definitions (3). Repetitive and single copy sequences are those that are present in many and one copy, respectively, at standard criterion. The term 'fossil repeats' is used for those sequences which reassociate with single copy kinetics at standard criterion but with repetitive kinetics at lower temperatures. 'True single copy' sequences are those which follow single copy kinetics at both standard and permissive criteria (3). By fractionating the genome arbitrarily into these classes, it is possible to draw conclusions about the rate of 'turnover' i.e. the processes of sequence amplification and deletion during the evolution of the genomes (4).

The nature of families of repetitive DNA can also be determined by studying reassociation kinetics at different stringency criteria. A family is defined as a group of sequences of sufficient similarity which reassociate with one another at a given criterion of stringency of temperature and salt conditions. The different families of a repetitive fraction and different members of a repetitive family may be related to one another by different degrees. Generally two types of families are recognized: a homogeneous family containing

members of sequences which are all related to the same extent and a heterogeneous family containing member sequences of varying similarity ranging from nearly perfect replicas to sequences barely similar enough to reassociate at standard stringency criterion of reassociation (5,6).

In chapter III, we have shown that the repetitive fractions in sponge gourd, ridge gourd, ash gourd and ivy gourd vary greatly in proportions, kinetic complexity and copy number. Though these differences involve repeated DNA sequences which probably have no sequence - dependent function, we feel that these sequences may have an important role in species isolation observed in Cucurbitaceae (7-9). To strengthen this view, it was necessary to study the composition of repetitive sequences and the nature of repetitive families in these species by studying reassociation kinetics at different criteria of stringency.

MATERIALS AND METHODS

Preparation of DNAs: Extraction of DNAs from seedlings of sponge gourd, ridge gourd and ash gourd and fruits of ivy gourd, shearing of DNAs to an average size of 550 base pairs, and sizing of DNAs were carried out as described in chapters II and III.

Isolation of repetitive DNA fractions: The sonicated DNAs (average size 550 base pairs) in 0.12 M sodium phosphate buffer (pH 6.8) were first denatured at 100°C for 10 min. and then incubated at 62°C to different Cot values ranging from Cot 1 M x sec/l to the repetitive Cot (Cot 50.0 M x sec/l for sponge gourd and ridge gourd, Cot 25.0 M x sec/l for ash gourd and Cot 10.0 M x sec/l for ivy gourd) (Chapter III). The sonicated DNAs were also incubated at 55°C, 69°C and 76°C to the respective limit Cot values of repetitive DNAs. The reassociated fractions were separated from the unreassociated single strands by hydroxyapatite chromatography (10). The reassociated duplexes were extensively dialysed against 0.12 M sodium phosphate buffer (pH 6.8) prior to their characterization.

Thermal denaturation: Thermal denaturation studies of total sheared DNA preparations and the repetitive DNA fractions isolated at different Cot values were performed in a Gilford 250 spectrophotometer. DNAs in 0.12 M sodium phosphate buffer (pH 6.8) were denatured at a heat rate of 1°C per min. Calculations of Tm, G+C content, hyperchromicity and base mismatch were done according to the standard procedures (10-13) described in chapter II. Total repetitive DNAs isolated at different temperatures were denatured at a heat rate of 0.25°C per min. and

absorbance values were recorded at 0.1°C increments. The first derivative curves were obtained as described earlier (Chapter II). The difference profiles were obtained by subtracting the high criterion (76°C) derivative values from the low criterion (55°C) profiles in each case (14).

Optical reassociation: Reassociation of total repetitive DNA fractions isolated at 62°C was monitored optically in the Gilford 250 spectrophotometer. Optical reassociation of sonicated (550 base pairs) DNAs was also monitored at 55°C, 62°C, 69°C and 76°C upto the respective limit Cot values of repetitive DNAs. A curve-fitting procedure was used to fit the data points to the second order kinetics in order to calculate the proportion, kinetic complexity and frequency of reiteration of different repetitive classes.

RESULTS

Thermal stability of repeat sequences: The melting properties of different repetitive DNA fractions isolated from sonicated DNAs (approx. 550 base pairs) were determined (Figure 4.1, Table 4.1). The melting curves of all the Cot fractions except Cot 1.0 DNA in ivy gourd are smooth. The Cot 1.0 DNA of ivy gourd shows a biphasic melting profile, with a major fraction (86%) of low thermal stability (Tm = 77.9°C) and a minor fraction (14%) of high thermal stability (Tm = 91.9°C).

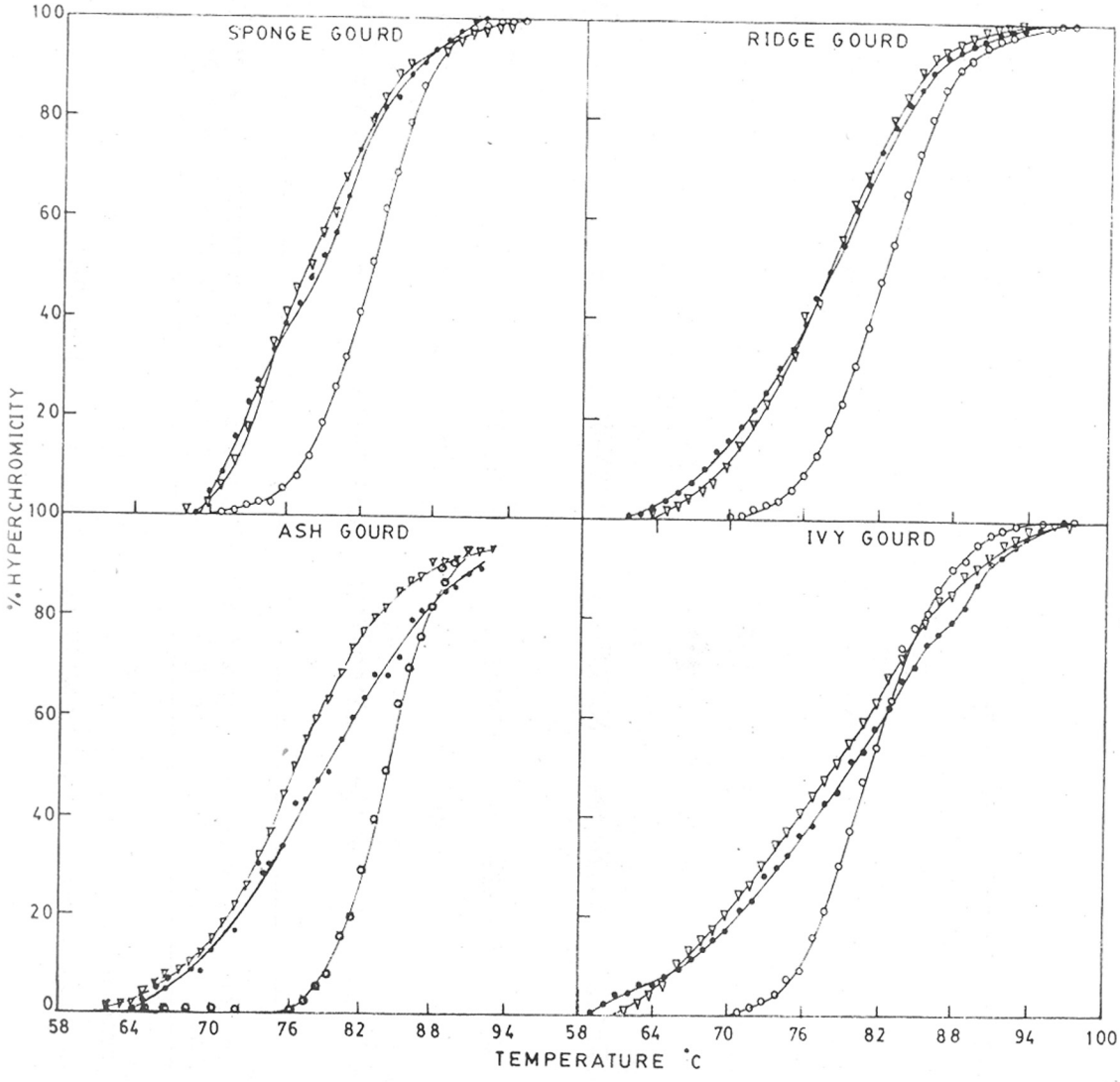


Figure 4.1 : Melting patterns of total sonicated (550 base pairs) DNAs (O - O), Cot 1.0 DNAs (● - ●) and limit repetitive Cot DNAs (▽ - ▽). Limit repetitive Cot is 50.0 for sponge gourd and ridge gourd, 25.0 for ash gourd and 10.0 for ivy gourd (Chapter III).

Table 4.1 Melting properties of total sonicated (550 base pairs) and repetitive DNAs

Species	Native sonicated DNA	Cot 1.0 M x sec/l	Cot 5.0 M x sec/l	Cot 10.0 M x sec/l	Cot 25.0 M x sec/l	Cot 50.0 M x sec/l
Sponge gourd						
Tm ^{o}C	83.44±0.90 (6)	76.14±2.74 (3)	77.00±1.00 (3)	77.50±2.21 (3)	-	77.87±0.65 (7)
% Mismatch[a]	-	7.30	6.44	5.94	-	5.57
% Hyperchromicity[b]	22.84±3.37	8.88±1.19	13.79±1.37	16.80±1.02	-	10.42±1.22
Ridge gourd						
Tm ^{o}C	82.47±0.12 (6)	78.03±0.96 (6)	77.33±0.31 (3)	77.27±0.42 (6)	-	77.67±0.42 (4)
% Mismatch[a]	-	4.44	5.14	5.20	-	4.80
% Hyperchromicity[b]	23.05±0.28	15.13±2.02	17.19±0.83	16.19±1.47	-	18.10±0.13
Ash gourd						
Tm ^{o}C	82.66±0.68 (5)	75.87±0.83 (3)	74.00±0.85 (4)	-	74.40±0.80 (3)	-
% Mismatch[a]	-	6.79	8.66	-	8.26	-
% Hyperchromicity[b]	22.41±1.88	13.43±0.64	14.43±0.89	-	14.47±1.14	-

Table 4.1 (contd.)

Ivy gourd						
Tm °C	81.55±0.45 (7)	77.19±1.19 90.91±0.92 (6)	78.10±0.53 (4)	79.15±0.85 (4)	-	-
% Mismatch[a]	-	4.36	3.45	2.40	-	-
% Hyperchromicity[b]	22.63±0.96	12.73	13.27±0.25	14.17±0.94	-	-

a : Calculated as lowering in Tm as compared to that of sonicated DNA by 1°C corresponds to 1% base mismatch (10).

b : Calculated as $\frac{A_{260}\ (98^{\circ}C) - A_{260}\ (62^{\circ}C)}{A_{260}\ (98^{\circ}C)}$ (12)

Figures in parentheses indicate the number of experiments carried out in each case.

To assess the extent of base mismatching, the Tm values of different Cot fractions were compared to the corresponding native, sheared DNA preparations in each species. From the lowering in Tm of the reassociated DNAs, the % base mismatch was estimated to be 5.6 to 7.3% in sponge gourd, 4.4 to 5.2% in ridge gourd, 6.5 to 8.7% in ash gourd and 2.4 to 4.4% in ivy gourd. This base mismatch is moderate as compared to the extensive base mismatch of repeated sequences observed in some other plant species (12,15).

Mini Cot curves: In order to determine whether any minor repetitive classes exist which are obscured in the reassociation curves of the total DNAs, the repetitive DNAs were isolated and their reassociation kinetics were studied optically (Figure 4.2, Table 4.2). The isolated total repetitive fractions of sponge gourd, ridge gourd, ash gourd and ivy gourd consist of a large (about 75%) slow reassociating component, with a reiteration frequency of a few hundred. Sponge gourd in addition shows a fast component (about 20%) with a reiteration frequency of 30000 as observed in the reassociation of total sheared DNA (Chapter III). Thus no subclasses of repetitive DNAs exist other than those observed in the reassociation kinetics of total DNAs (Chapter III). The presence of a large fraction of an early component and absence of extremely fast reassociating components makes these species suitable for further investigations (5).

Fig. 4.2 : Optical reassociation curves of total repetitive DNAs isolated at 62°C, i.e. Cot 50.0 DNA for sponge gourd and ridge gourd, Cot 25.0 DNA for ash gourd and Cot 10.0 DNA for ivy gourd. The repetitive fractions (O.D. 260 of 0.800 to 1.000) were melted and reassociated at 62°C in 0.12 M or 0.36 M sodium phosphate buffer (pH 6.8). Correction to the rate of reassociation was made according to Britten et al. (10) when the higher salt concentration was used. The solid line through the data points represents the least squares fit allowing all the parameters to free float. The lower dashed curves represent the predicted reassociation kinetics of pure components.

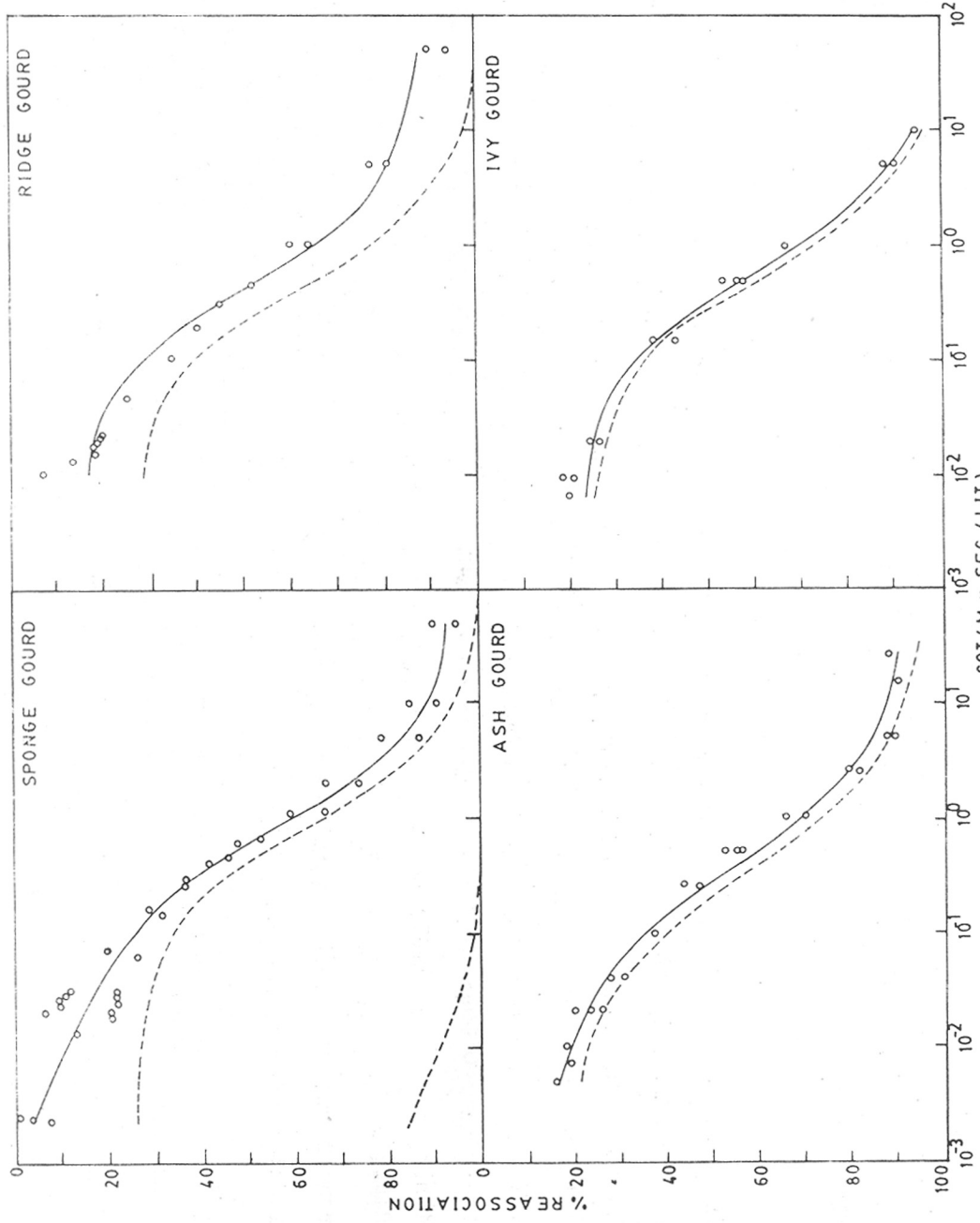

RIDGE GOURD
IVY GOURD
SPONGE GOURD
ASH GOURD
% REASSOCIATION
COT (M x SEC / LIT)

Figure 4.2

Table 4.2 Reassociation properties of isolated total repetitive DNA fractions

Species	Range of reassociation Cot M x sec/l	Fraction of genome [a]	Cot 1/2 observed [a] M x sec/l	Cot 1/2 pure [b] M x sec/l	K pure [c]	Kinetic complexity in nucleotide pairs [d]	Frequency of repetition [e]
Sponge gourd							
Very fast	< 0.0023	0.045	-	-	-	-	-
Fast	0.0023 to 0.10	0.193	0.0081	0.0016	625.000	1.60×10^3	2841.98
Slow	0.10 to 50.0	0.743	0.834	0.620	1.613	6.20×10^5	278.18
Unreassociated	> 50.0	0.064	-	-	-	-	-
Ridge gourd							
Very fast	< 0.01	0.165	-	-	-	-	-
Slow	0.01 to 50.0	0.735	0.451	0.331	3.021	3.31×10^5	1385.81
Unreassociated	> 50.0	0.116	-	-	-	-	-
Ash gourd							
Very fast	< 0.005	0.160	-	-	-	-	-
Slow	0.005 to 25.0	0.732	0.628	0.460	2.170	4.60×10^5	418.79

Table 4.2 (contd.)

Ivy gourd							
Very fast	< 0.007	0.226	-	-	-	-	-
Slow	0.007 to 10.0	0.758	0.569	0.431	2.320	4.31×10^5	474.52
Unreassociated	> 10.0	0.025	-	-	-	-	-

a : Values obtained from respective reassociation curves (Fig. 4.2)

b : Cot 1/2 observed x fraction of the genome

c : $(\text{Cot } 1/2 \text{ pure})^{-1}$

d : Values obtained from the standard relationship for *E.coli* where genome size is 4.5×10^6 nucleotide pairs (22) and Cot 1/2 is 4.5 M x sec/l (Chapter III)

e : Cot 1/2 value of unique DNA taken from chapter III, divided by Cot 1/2 of the given fraction.

Optical reassociation at different temperatures:

Reasssociation of DNA at Tm -25°C permits its sequences to reassociate if their average complementarity is at least 75% (10). When the temperature of reassociation is reduced from Tm -25°C to a more permissive criterion (temperature of incubation lower than Tm -25°C), more mismatched or diverged sequences (fossil repeats) are allowed to form duplexes, and thus the fraction of the DNA scored as repetitive increases. In contrast, increasing reassociation temperature increases sequence specificity required for reassociation and only very well-matched, recent or conserved (perhaps functional) sequences reassociate (3,4). Use of different stringency criteria for reassociation is, therefore, a useful method for fractionating populations of repeated sequences by their extent of relatedness. In the present work, the DNAs were reassociated at 55°C, 69°C and 76°C, in addition to 62°C (standard criterion). These temperatures are about 30°C to 5°C below the respective melting temperatures of the native DNAs of the four species (Chapter II). The results of these studies are depicted in Figure 4.3 and summarized in Table 4.3. The Cot 1/2 of single copy DNA in all the reassociations was considered to be the same as that at 62°C (Chapter III). The reassociation rate has been corrected for the effect of temperature and the decreased extent of reassociation (16).

Fig. 4.3 : Optical reassociation curves of total sonicated DNAs (550 base pairs) at different temperatures. Reassociations were carried out as described in the caption of Fig. 4.2.

(● – ●) Reassociation curve obtained at 55°C

(o - o) Reassociation curve obtained at 62°C.

(Δ - Δ) Reassociation curve obtained at 69°C.

(x - x) Reassociation curve obtained at 76°C.

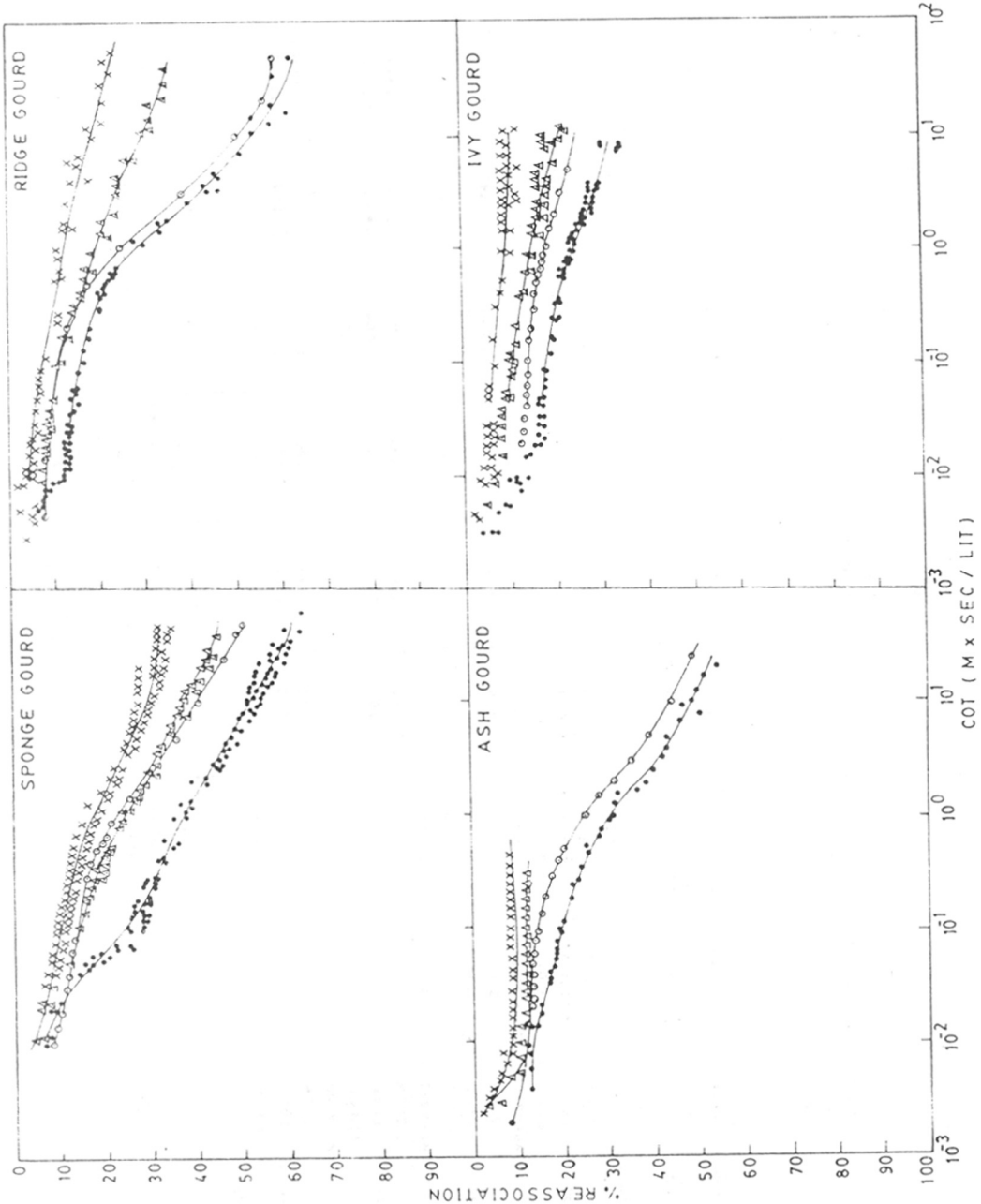

RIDGE GOURD
IVY GOURD
SPONGE GOURD
ASH GOURD
% REASSOCIATION
COT (M x SEC / LIT)

Figure 4.3

TABLE 4·3

SPECIES	COMPONENT	RANGE OF REASSOCIATION Cot M S	FRACTION REASSOCIATED (a)	Cot ½ OBSERVED M.S. (a)	RATE RELATIVE TO OPTIMUM RATE (b)	Cot ½ CORRECTED M.S. (c)	Cot ½ PURE M.S. (d)	K PURE (e)	KINETIC COMPLEXITY IN NUCLEOTIDE PAIRS (f)	FREQUENCY OF REPETITION (g)
SPONGE GOURD										
55 °C	VERY FAST	< 0·01	0·044	-	-	-	-	-	-	-
(RMS = 0·0171)	FAST	0·01 to 0·57	0·304	0·042	0·98	0·041	0·013	76·923	$1·30 \times 10^4$	5658·54
	SLOW	0·57 to 50·0 (h)	0·289	5·695	0·98	5·580	1·613	0·620	$1·613 \times 10^6$	41·58
	UNREASSOCIATED	> 50·0	0·364	-	-	-	-	-	-	-
62 °C (i)	VERY FAST	< 0·01	0·084	-	-	-	-	-	-	-
(RMS = 0·0331)	FAST	0·01 to 0·64	0·111	0·084	0·92	0·077	0·0086	116·279	$8·60 \times 10^3$	3012·99
	SLOW	0·64 to 50·0	0·345	3·745	0·92	3·445	1·189	0·841	$1·189 \times 10^6$	87·34
	UNREASSOCIATED	> 50·0	0·516	-	-	-	-	-	-	-
69 °C	VERY FAST	< 0·011	0·048	-	-	-	-	-	-	-
(RMS = 0·0196)	FAST	0·011 to 0·12	0·116	0·026	0·76	0·020	0·0023	434·783	$2·30 \times 10^3$	11600·0
	SLOW	0·12 to 50·0	0·282	2·668	0·76	2·028	0·572	1·748	$5·72 \times 10^5$	114·40
	UNREASSOCIATED	> 50·0	0·555	-	-	-	-	-	-	-
76 °C	VERY FAST	< 0·01	0·038	-	-	-	-	-	-	-
(RMS = 0·0153)	FAST	0·01 to 0·17	0·080	0·019	0·56	0·011	0·00088	1136·364	$8·80 \times 10^2$	21090·91
	SLOW	0·17 to 50·0	0·226	2·878	0·56	1·612	0·364	2·747	$3·64 \times 10^5$	143·92
	UNREASSOCIATED	> 50·0	0·656	-	-	-	-	-	-	-
RIDGE GOURD										
55 °C	VERY FAST	< 0·0042	0·113	-	-	-	-	-	-	-
(RMS = 0·0220)	SLOW	0·0042 to 50·0	0·521	2·019	0·98	1·979	1·031	0·970	$1·03 \times 10^6$	315·82
	UNREASSOCIATED	> 50·0	0·366	-	-	-	-	-	-	-
62 °C (i)	VERY FAST	< 0·0044	0·079	-	-	-	-	-	-	-
(RMS = 0·0335)	SLOW	0·0044 to 50·0	0·503	1·898	0·90	1·708	0·859	1·164	$8·59 \times 10^5$	365·93
	UNREASSOCIATED	> 50·0	0·420	-	-	-	-	-	-	-
69 °C	VERY FAST	< 0·0058	0·075	-	-	-	-	-	-	-
(RMS = 0·0167)	SLOW	0·0058 to 50·0	0·215	0·579	0·70	0·405	0·087	11·494	$8·70 \times 10^4$	1543·21
	UNREASSOCIATED	> 50·0	0·710	-	-	-	-	-	-	-
76 °C	VERY FAST	< 0·0086	0·053	-	-	-	-	-	-	-
(RMS = 0·0151)	SLOW	0·0086 to 50·0	0·141	0·870	0·50	0·435	0·061	16·393	$6·10 \times 10^4$	1436·78
	UNREASSOCIATED	> 50·0	0·806	-	-	-		-	-	-

Table 4.3 (contd.)

ASH GOURD										
55°C	VERY FAST	<0·006	0·159	–	–	–	–			
(RMS 0·0220)	SLOW	0·006 to 25·0	0·400	1·805	0·98	1·769	0·708	1·412	$7·08 \times 10^5$	148·67
	UNREASSOCIATED	>25·0	0·442	–	–	–	–	–	–	–
62°C(i)	VERY FAST	<0·004	0·120	–	–	–	–	–	–	–
(RMS 0·0230)	SLOW	0·004 to 25·0	0·360	1·850	0·96	1·776	0·639	1·565	$6·39 \times 10^5$	148·09
	UNREASSOCIATED	>25·0	0·520	–	–	–	–	–	–	–
69°C	VERY FAST	<0·003	0·040	–	–	–	–	–	–	–
	SLOW	0·003 to 0·02	0·080	–	–	–	–	–	–	–
	UNREASSOCIATED	>0·02	0·880	–	–	–	–	–	–	–
76°C	VERY FAST	<0·003	0·030	–	–	–	–	–	–	–
	SLOW	0·003 to 0·01	0·050	–	–	–	–	–	–	–
	UNREASSOCIATED	>0·01	0·920	–	–	–	–	–	–	–
IVY GOURD										
55°C	VERY FAST	<0·086	0·194	–	–	–	–	–	–	–
(RMS 0·0153)	SLOW	0·086 to 10·0	0·154	3·046	0·98	2·985	0·460	2·174	$4·60 \times 10^5$	90·45
	UNREASSOCIATED	>10·0	0·652	–	–	–	–	–	–	–
62°C(i)	VERY FAST	<0·060	0·138	–	–	–	–	–	–	–
(RMS 0·0181)	SLOW	0·060 to 10·0	0·111	1·560	0·90	1·404	0·156	6·410	$1·56 \times 10^5$	192·31
	UNREASSOCIATED	>10·0	0·754	–	–	–	–	–	–	–
69°C	VERY FAST	<0·075	0·102	–	–	–	–	–	–	–
(RMS 0·0104)	SLOW	0·075 to 10·0	0·110	1·271	0·70	0·890	0·098	10·204	$9·80 \times 10^4$	303·37
	UNREASSOCIATED	>10·0	0·789	–	–	–	–	–	–	–
76°C	VERY FAST	<0·029	0·054	–	–	–	–	–	–	–
(RMS 0·00956)	SLOW	0·029 to 10·0	0·056	0·327	0·56	0·183	0·010	100·000	$1·00 \times 10^4$	1475·41
	UNREASSOCIATED	>10·0	0·891	–	–	–	–	–	–	–

Table 4.3 (contd.)

a : Values obtained from the respective reassociation curves (Fig. 4.3)

b : Taken from Fig. 8 of Bonner et al. (16)

c : Cot 1/2 observed x rate of reassociation relative to optimum rate (at Tm -25 °C)

d : Cot 1/2 corrected x fraction of the genome

e : $(\text{Cot } 1/2 \text{ pure})^{-1}$

f, g : Values obtained as given in the caption of Table 4.2

h : Some of these data points were used earlier with data points obtained by hydroxyapatite chromatography to obtain total Cot curve (Chapter III). In this chapter, only optical reassociation data was used for curve fitting for comparative purposes, but the computer fits gave very close results to the earlier fit (Chapter III).

i : Limit Cot of repetitive DNA: Cot 50.0 M x sec/l for sponge gourd and ridge gourd, Cot 25.0 M x sec/l for ash gourd and Cot 10.0 M x sec/l for ivy gourd (Chapter III).

In all the four species, the fraction recorded as 'repetitive' reduces as the conditions for reassociation are made more stringent. However, species - specific differences are noticed in the extent of reduction of repetitive fraction. In sponge gourd, three repetitive components, very fast, fast and slow, are evident at all the four reassociation conditions. The change in proportion of very fast component is negligible at different conditions. The fast component increases almost two-fold at 55°C as compared to that at 62°C and becomes almost half in proportion at 76°C. The slow component, however, does not change considerably. The extent of reassociation of the slow component is 29% at 55°C and 33% at 62°C, suggesting the removal of certain sequences from the slow class to the fast class at 55°C. In ridge gourd, however, the proportion of very fast reassociating component decreases gradually from 55°C to 76°C. Though there is a marginal decrease in the slow component when the temperature is raised from 55°C to 62°C, it is very rapid at higher temperatures (> 62°C). In ash gourd, there is a slight increase in the proportion of both very fast and slow reassociating fractions at the relaxed criteria. However, only 12% and 8% of the genome can form stable duplexes at 69°C and 76°C respectively. This may be because the slow repetitive sequences in ash gourd are greatly diverged and hence cannot be recognized as repetitive at higher stringency conditions. In ivy gourd,

the total repetitive DNA is 35% at 55°C and 11% at 76°C. Both very fast component and slow component lose some members at elevated temperatures.

Apart from changes in the proportions of the various repetitive fractions as a function of incubation temperature, differences are also observed with respect to kinetic complexity, copy number and rates of reassociation (Table 4.3). The kinetic complexity, for example, decreases with an increase in incubation temperature in all the cases. The decrease ranges from about 4.4 fold in the slow component of sponge gourd to about 46 fold in that of ivy gourd. The copy number, on the other hand, increases from 3.5 fold in slow component of sponge gourd to about 16 fold in ivy gourd. In all the cases the reassociation rates increase with increase in incubation temperature.

Thermal denaturation of repeats isolated at different temperatures: The derivative melting profiles of repetitive fractions isolated at different temperatures were compared to determine the nature of repetitive families (14). This assay is essentially independent of copy number of different repeat families. Table 4.4 summarizes thermal stabilities of the repeat fractions isolated at different temperatures. The repeats isolated at higher temperatures show progressively lesser mismatch.

Table 4.4 Melting properties of total repetitive DNAs isolated at different incubation temperatures

Species	Incubation temperature (°C)	Tm (°C)	Hyperchromicity (%)	Base mismatch[a] (%)
Sponge gourd	55	74.18	14.56	9.26
	62	77.87	12.95	5.57
	69	81.60	10.36	1.85
	76	82.50	11.90	0.94
Ridge gourd	55	76.60	12.60	5.87
	62	77.67	12.47	4.80
	69	80.30	11.40	2.17
	76	81.45	10.92	1.02
Ash gourd	55	72.20	14.00	10.46
	62	74.40	14.47	8.26
	69	-	-	-
	76	-	-	-
Ivy gourd	55	79.13	13.82	2.42
	62	79.15	14.17	2.40
	69	81.10	11.92	0.45
	76	84.60	11.24	-

a : Calculated by comparing the Tm of repetitive fractions to that of sonicated DNAs, given in Table 4.1

Figure 4.4 shows the derivative melting profiles of total repetitive DNAs of sponge gourd and ivy gourd isolated at different temperatures. After subtracting the derivative values obtained for the fraction isolated at 76°C from those of the fraction isolated at 55°C, positive difference is obtained at lower temperatures (55°C to 76°C) and no difference is observed at higher temperatures (> 76°C). In ridge gourd, very similar results to those of sponge gourd are obtained.

An interesting observation in the derivative profiles of repetitive DNA of ivy gourd is the presence of a high melting component which is especially prominent in the 76°C profile and constitutes about 20% of this fraction. The Tm of repetitive DNA of ivy gourd isolated at 76°C is thus higher than that of native sheared DNA.

DISCUSSION

In contrast to the vast amount of information available about plant genomes (17,18), analysis of DNA reassociation kinetics at different temperatures has been studied only in a few species such as pea and mungbean (3,4,19,20), barley, daffodil, parsley fern and deer fern (6) and *Arachis* (21). Of these, barley, daffodil, parsley fern, deer fern and mungbean are found to be composed of homogeneous repetitive families, while those of pea are predominantly heterogeneous. In the

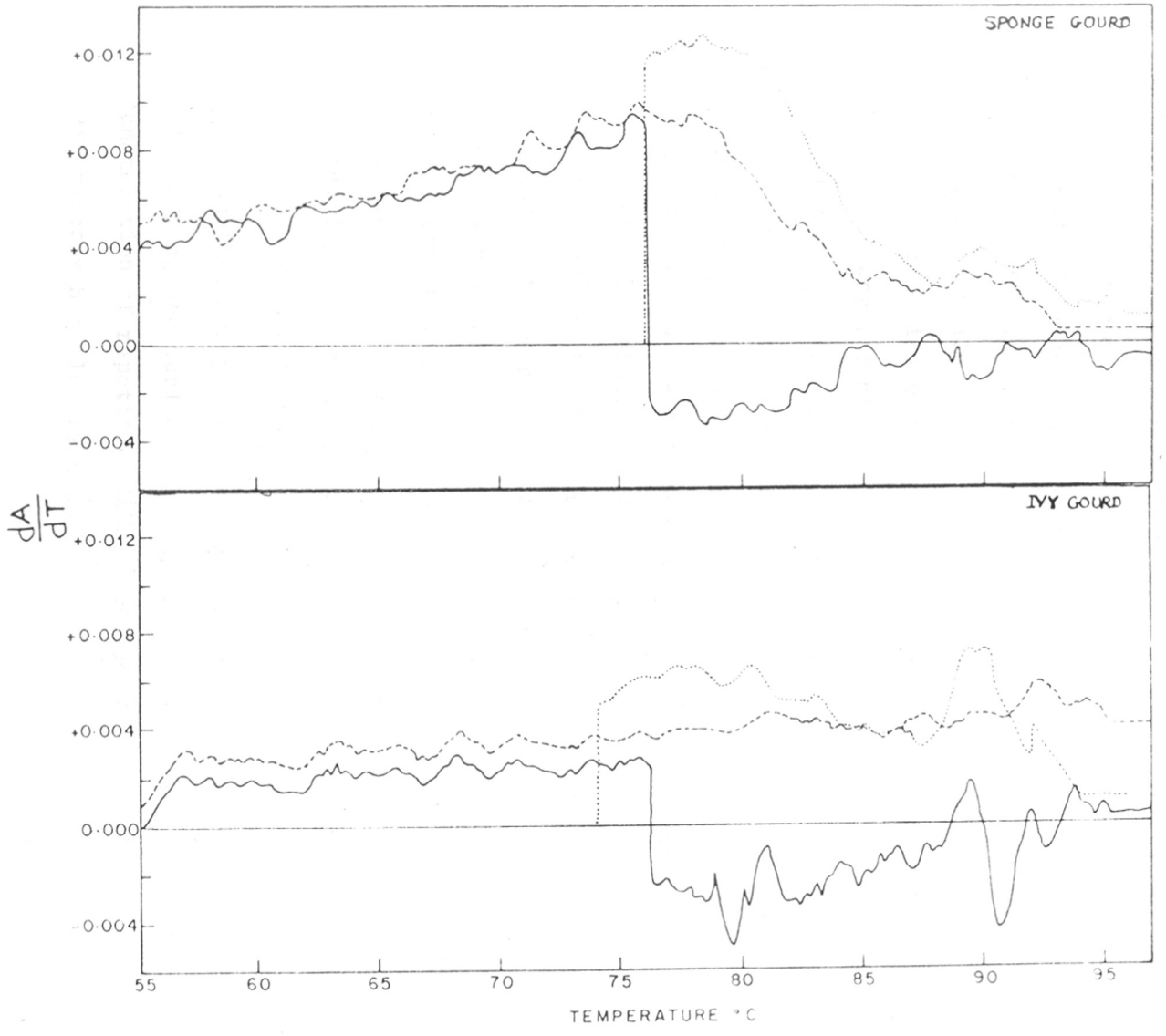
Fig. 4.4 : High resolution thermal denaturation curves for sponge gourd and ivy gourd total repetitive DNAs (Cot 50.0 and Cot 10.0 fractions respectively) isolated at 55°C and 76°C. The 76°C curves were subtracted from 55°C curves to get the difference profiles.

(----) derivative curve for fraction isolated at 55°C

(....) derivative curve for fraction isolated at 76°C

(———) difference between the curves obtained at 76°C and at 55°C.

Figure 4.4


SPONGE GOURD
IVY GOURD
+0·012
+0·008
+0·004
0·000
-0·004
dA/dT
55
60
65
70
75
80
85
90
95
TEMPERATURE °C

present studies, Bendich's definitions of homogeneous and heterogeneous repeat families have been grossly accepted (5,6). However, there are two different views regarding the composition of evolutionarily young and old repeat families. According to Bendich and Anderson (6), all repeat families, young or old, should have the same copy number, while according to Preisler and Thompson (19), the older families would have a lower number of members than the younger families due to sequence divergence and dispersion. In the present work, we observe that the stable families reassociating at higher temperatures have a higher copy number and reassociate faster than those reassociating at lower temperatures. These observations are essentially similar to those of Preisler and Thompson (19) on the homogeneous repeat families of mungbean. At lower incubation temperatures longer, imperfect repetitive stretches reassociate, while at higher temperatures, shorter but more perfect duplexes are formed. Hence it appears that the repetitive DNAs essentially consist of homogeneous families in the present four Cucurbitaceae species.

Another assay developed by Preisler and Thompson (14) for studying the derivative profiles of thermal denaturation of repetitive DNAs isolated at different temperatures has also been employed here. The difference

curves obtained for sponge gourd, ridge gourd and ivy gourd repetitive fractions isolated at 55°C and 76°C are essentially like those expected for homogeneous families. While there is a positive difference in the derivative curves at lower temperatures, it reaches zero at higher temperatures. This suggests that some families are removed in toto from the reassociating pool at higher temperatures. Thus this assay also confirms the homogeneous nature of repetitive families of the Cucurbitaceae genomes under investigation.

Applying Murray et al.'s definitions of 'fossil repeats' and 'true single copy' sequences (3), we find that in ridge gourd, ash gourd and ivy gourd there are very few 'fossil' repeats and most of the single copy DNA is 'true' (Table 4.5). On the other hand, in sponge gourd, a large proportion of single copy DNA behaves as 'fossil' repeats at lower temperatures. In absolute amounts, calculated from genomic complexity (Chapter III), ridge gourd, ash gourd and ivy gourd contain more true single copy DNA than sponge gourd. A smaller fraction of fossil repeats and a larger fraction of true single copy DNA indicates a relatively low rate of sequence amplification in the genomes of these three species. On the other hand, a larger proportion of fossil repeats, and a smaller proportion of true single copy sequences indicates a comparatively higher rate of turnover in the genome of sponge gourd.

Table 4.5 Fossil repeats and true single copy sequences in Cucurbitaceae genomes

Species	Genome size[a] 1C in nucleotide pairs	Fraction[b] of single copy DNA at standard criterion	Fraction[c] of single copy DNA at permissive criterion	% Fossil[d] repeats	% True[e] single copy DNA	Kinetic complexity of true single copy DNA in nucleotide pairs
Sponge gourd	8.37×10^7	0.53	0.36	32.08	67.92	3.01×10^7
Ridge gourd	1.69×10^8	0.41	0.37	9.76	90.24	6.25×10^7
Ash gourd	9.20×10^7	0.52	0.44	15.38	84.62	4.05×10^7
Ivy gourd	1.59×10^8	0.75	0.65	13.33	86.67	1.03×10^8

a : From Chapter III

b : Fraction remaining unreassociated at limit repetitive Cot at 62°C

c : Fraction remaining unreassociated at limit repetitive Cot at 55°C

d : Calculated as $\frac{b-c}{b} \times 100$

e : Calculated as $\frac{c}{b} \times 100$

f : Calculated as a x c

References

1. Murray, M.G., Palmer, J.D., Cuellar, R.E. and Thompson, W.F. (1979) Biochemistry 18, 5259-5266.

2. Murray, M.G., Cuellar, R.E. and Thompson, W.F. (1978) Biochemistry 17, 5781-90.

3. Murray, M.G., Peters, D.L. and Thompson, W.F. (1981) J.Mol.Evol. 17, 31-42.

4. Thompson, W.F. and Murray, M.G. (1980) In: Davies, D.R. and Hopwood, D.A. (eds.) Proceedings of the 4th John Innes Symposium, The Plant Genome, pp. 31-45.

5. Bendich, A.J. (1979) In: Rubenstein, I (ed.) Molecular Biology of Plants, Academic Press, New York, pp. 1-30.

6. Bendich, A.J. and Anderson, R.S. (1977) Biochemistry 16, 4655-4663.

7. Roy, R.P., Saran, S. and Dutt, B. (1972) In: Murthy, Y.S., Johri, B.M., Mohan Ram, H.Y. and Varghese, T.M. (eds.) Advances in Plant Morphology, Prof. V. Puri Commemoration Volume, Sarita Prakashan, Meerut, India, pp. 193-202.

8. Roy, R.P., Thakur, V. and Trivedi, R.N. (1966) J.Cytol.Genet. 1, 30-40.

9. Thakur, M.R. and Chaudhari, B. (1966) Indian J.Genet.Plant Breed. 26, 79-86.

10. Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) In: Methods in Enzymol. (Grossman, G.L. and Moldave, K., eds.) Vol.29E, Academic Press, New York, pp.363-418.

11. Mandel, M. and Marmur, J. (1968) In: Methods in Enzymol. (Grossman, G.L. and Moldave, K., eds.) Vol. 12B, Academic Press, New York, pp. 195-206.

12. Zimmerman, L. and Goldberg, R.B. (1977) Chromosoma (Berl.) 59, 227-252.

13. Wimpee, C.F. and Rawson, J.R.Y. (1979)
Biochim.Biophys.Acta 562, 192-206.

14. Preisler, R.S. and Thompson, W.F. (1981)
J.Mol.Evol. 17, 85-93.

15. Hake, S. and Walbot, V. (1980) Chromosoma (Berl.)
79, 251-270.

16. Bonner, T.I., Brenner, D.J., Neufeld, B.R.and
Britten, R.J. (1974) J.Mol.Biol. 81, 123-135.

17. Thompson, W.F. and Murray M.G. (1981) In: Marcus A.(ed.)
The Biochemistry of Plants, Vol.6, Academic Press,
New York, pp. 1-81.

18. Wenzel, W. and Hemleben, V. (1982) Plant Syst.Evol.
139, 209-227.

19. Preisler, R.S. and Thompson, W.F. (1981)
J.Mol.Evol. 17, 78-84.

20. Pearson, W.R., Smith, S.L., Wu, J.R. and Bonner, J.
(1978) Plant Physiol. 62, 112-115.

21. Dhillon, S.S., Rake, A.V. and Miksche, J.P. (1980)
Plant Physiol. 65, 1121-1127.

22. Cairns, J. (1963) Cold Spring Harbor Symp.
Quant.Biol. 28, 43.

CHAPTER V

ORGANIZATION OF REPEATED AND SINGLE COPY DNA SEQUENCES

ORGANIZATION OF REPEATED AND SINGLE COPY DNA SEQUENCES

SUMMARY

DNA sequence organization patterns have been studied in four Cucurbitaceae plant species, namely, sponge gourd, ridge gourd, ash gourd and ivy gourd. Extensive interspersion of repeat and single copy sequences has been observed in sponge gourd and ridge gourd. In ash gourd and ivy gourd, however, there is a limited interspersion of these sequences, and a large portion of the single copy DNA remains uninterspersed. In all the four species, the interspersed repetitive sequences are composed of a major class (75-80%) of short repeats (300 base pairs long) and a minor class (15-20%) of long repeats (2000-4000 base pairs long). The average length of single copy sequences dispersed among repeats is 2000-3000 base pairs. In spite of the gross similarities in the genome organization in the four species, the fraction of repeats and single copy sequences involved in short and long period interspersion patterns, and fraction of single copy sequences remaining uninterrupted by repeats are vastly different. Sponge gourd and ridge gourd show more extensive interspersion than ash gourd and ivy gourd, suggesting faster evolution of these two genomes. Between sponge gourd and ridge gourd, which

belong to the same genus Luffa, the genome of sponge gourd shows predominance of short period pattern and hence its genome could be evolving faster than that of ridge gourd. These variations at the genome organization level could be a reason for the fertility barriers which are so well established in Cucurbitaceae species.

INTRODUCTION

In contrast to satellite DNA, most repetitive sequences are interspersed to varying degrees with single copy sequence elements. The plant species investigated so far show a variety of interspersion patterns ranging from short period pattern through mixed patterns to extremely long period interspersion patterns (Chapter I). While considering the significance attributable to interspersion patterns of repetitive and single copy DNA, it will be useful to compare the data from a wide variety of organisms, in order to obtain a better indication of the limits within which different patterns may vary. These studies will also lead to a better understanding of the amplification and translocation events responsible for the origin and evolution of repetitive DNA.

After obtaining detailed information about the heterogeneity in base composition (Chapter II), reassociation kinetics (Chapter III) and rates of evolution

of repeated DNA sequences (Chapter IV), the next logical question would be the mode of arrangement of repeated and single copy DNA sequences in the four Cucurbitaceae genomes. Hence sequence organization studies were taken up to obtain some information regarding the phylogenetic relationship and the rates of evolution of these genomes.

MATERIALS AND METHODS

Extraction of DNA: DNAs of sponge gourd, ridge gourd, ash gourd and ivy gourd were extracted as described in chapter II.

Shearing of DNAs: DNAs in 0.12 M sodium phosphate buffer (pH 6.8) were sheared to an average fragment size of 550 base pairs as described in chapter III. Sonication with 1/4th inch (yellow) probe for one minute at 40 setting using the Biosonic III yielded DNA fragments of 1500 base pairs. Longer DNA fragments (2850 to 7400 base pairs) were obtained by homogenizing high molecular weight DNAs of a constant concentration (usually 1 mg in 3 ml of 0.12 M sodium phosphate buffer, pH 6.8) in a VirTis 60K homogenizer, by varying the conditions of speed and time as follows:

Table 5.1 Sonication in VirTis 60 K homogenizer to obtain different fragment sizes.

Speed (RPM)	Time (min.)	Size obtained (Base pairs)
57,000	45	2850
50,000	15	4200
48,000	7	4600
30,000	30	5200
40,000	2	6800
32,000	4	7400

Homogenizing the DNA solution in Sorvall Omnimixer (Model No.17106) for 6 min. at 25,000 rpm, or homogenizing DNA of lower concentration (500 µg in 3 ml) at 57,000 rpm for 45 min. gave a size of 2000 base pairs (Figure 5.1).

Sizing of DNAs: Fragment lengths of all the DNA preparations, and of repetitive DNA fractions obtained from Agarose A50 column (described later) were determined by agarose slab gel electrophoresis as described in chapter II.

Reassociation of longer DNA fragments: Reassociation curves of 7400 base pairs long DNA fragments were studied in the entire Cot range of 1×10^{-2} to 1×10^{4}. DNAs were reassociated either in 0.12 M or 0.36 M sodium

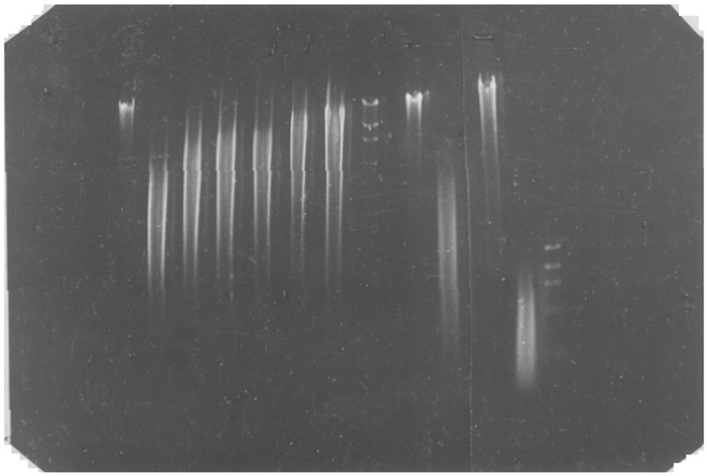


Fig. 5.1 : Determination of fragment sizes of DNA obtained by sonicating or homogenizing native unsheared DNA (> 10 kilo base pairs) as given in Table 5.1.

Lane a, i, k : native unsheared DNA

Lane b : 2850 base pairs DNA Lane c : 4200 base pairs DNA

Lane d : 4600 base pairs DNA Lane e : 5200 base pairs DNA

Lane f : 6800 base pairs DNA Lane g : 7400 base pairs DNA

Lane h : λ DNA digested with Hind III (23300, 9500, 6400, 4200, 2200 and 1800 base pairs).

Lane j : 1500 base pairs DNA Lane l : 550 base pairs DNA (Chapter III).

Lane m : ϕ X 174 RF DNA digested with Hae III (1342, 1078, 872, 606 and 310 base pairs).

phosphate buffer (pH 6.8). The corresponding correction was applied to the rate of reassociation for the higher molarity buffer (1). Double stranded DNA structures were separated from single stranded DNA molecules by hydroxyapatite chromatography. The data points were fitted by computer to calculate the size of different classes of DNA and their rates of reassociation by allowing all parameters to free-float as described in chapter III.

Thermal stability of repetitive duplexes: DNAs of different fragment lengths ranging from 550 to 7400 base pairs were reassociated to respective limit repetitive Cot values (Cot 50.0 for sponge gourd and ridge gourd, Cot 25.0 for ash gourd and Cot 10.0 for ivy gourd, Chapter III). The reassociated duplexes were separated by hydroxyapatite chromatography, dialysed against 0.12 M sodium phosphate buffer (pH 6.8) and heat denatured at a rate of 1^{o}C/min. in the Gilford 250 spectrophotometer. The S1 nuclease resistant repetitive duplexes and the various size classes of repetitive duplexes separated by Agarose A50 gel filtration (following section) were also melted in the Gilford spectrophotometer, as described in chapter II. The melting temperature (Tm), % hyperchromicity, % mismatch, duplex content and average duplex length were calculated according to the standard procedures (1-4). Single strand collapse was determined according to the procedure of

Davidson et al (5). Unreassociated, unique DNA which included DNA sequences reassociating after the respective limit repetitive Cot values was isolated by hydroxyapatite chromatography. It was melted in 0.12 M sodium phosphate buffer (pH 6.8) and the hyperchromicities were found to be in the range of 1.5 to 3%.

Isolation of S1 nuclease resistant duplexes: In order to determine the length of repeated sequences and their distribution in the total population of repeats, the unfragmented total DNAs (molecular weight > 10 kilo base pairs) were reassociated under conditions at which only repetitive sequences could reassociate (i.e. to the respective limit repetitive Cot values). Native DNA samples, dialysed against 0.18 M NaCl - 0.006 M PIPES buffer [piperazine-N-N'-bis-(2-ethane sulfonic acid)], pH 7.0, were denatured in boiling water bath and allowed to reassociate at 62°C to specific limit repetitive Cot values. Reassociations were also carried out till Cot 0.1 to study the size distribution of highly repetitive fractions. After reassociation, samples were adjusted to 25 mM sodium acetate (pH 4.5) 0.1 mM $ZnSO_4$, 25 mM β-mercaptoethanol (6,7). At least 1.5 to 2.0 mg DNA was used for each experiment. 10 units of S1 nuclease (Boehringer) per μg of DNA were added and the samples were incubated at 37°C for 10 min. The reaction was

terminated by adjusting the samples to 0.12 M sodium phosphate buffer (pH 6.8) and cooling to 4°C. The S1 nuclease resistant duplexes were separated from the hydrolysed DNA by hydroxyapatite chromatography by elution with 0.4 M sodium phosphate buffer (pH 6.8). The samples were dialysed against 0.12M sodium phosphate buffer (pH 6.8) to determine their melting properties and size distribution.

Sizing of repetitive duplexes: The size distribution of S1 nuclease resistant duplexes was determined by gel filtration on Agarose A50 column (100-200 mesh, BioRad Laboratories), previously calibrated with calf thymus DNA of known duplex length.

The agarose gel matrix (fully hydrated in 0.001 M Tris-EDTA buffer, BioRad Laboratories) was poured around glass beads of 4 mm diameter in a 92x1.5 cm column. The column was equilibriated with 0.12M sodium phosphate buffer (pH 6.8). Approximately 70-100 µg of S1 nuclease resistant duplexes, separated by hydroxyapatite chromatography, were loaded on the column and eluted with 0.12M sodium phosphate buffer (pH 6.8). The effluent was continuously monitored using LKB Uvicord at 253 nm (8,9). The exclusion limit of this column was 1500 base pairs. The column was reused a number of times by washing extensively with 0.12M sodium phosphate buffer after each chromatographic operation.

Determination of length of interspersed single copy DNA: The fraction bound to hydroxyapatite at the limit repetitive Cot with DNAs of different fragment lengths (described earlier) was plotted as a function of fragment size.

RESULTS

Presence of interspersion of repetitive and single copy DNA sequences: Presence or absence of interspersion of repetitive and single copy sequences can be determined by comparing the hydroxyapatite binding data of long and short DNA fragments over a range of Cot values. An increased binding of DNAs of higher fragment length could, however, be either due to the effect of fragment length alone or due to interspersion effects. In the absence of interspersion, the rate of reassociation will vary according to the fragment size of DNA (10). The results will be different if repeated sequences are interspersed with each other or with single copy sequences.

Figure 5.2 shows the reassociation kinetics of 7400 base pairs long DNA fragments. The computer program was allowed to find the best fit by allowing all parameters to free float. The results of this analysis are given in Table 5.2. As is evident in Figure 5.2, there is a tremendous increase in hydroxyapatite binding in case of

Fig. 5.2 : Reassociation kinetics of 7400 base pairs long DNA fragments of four Cucurbitaceae species. The solid line through the data points represents the least squares fit, allowing all the parameters to free float. (RMS error = 0.046, 0.031, 0.022 and 0.027 for sponge gourd, ridge gourd, ash gourd and ivy gourd respectively). The dashed lines represent the reassociation curve of 550 base pairs long DNAs, included for comparison (Chapter III).

Figure 5.2

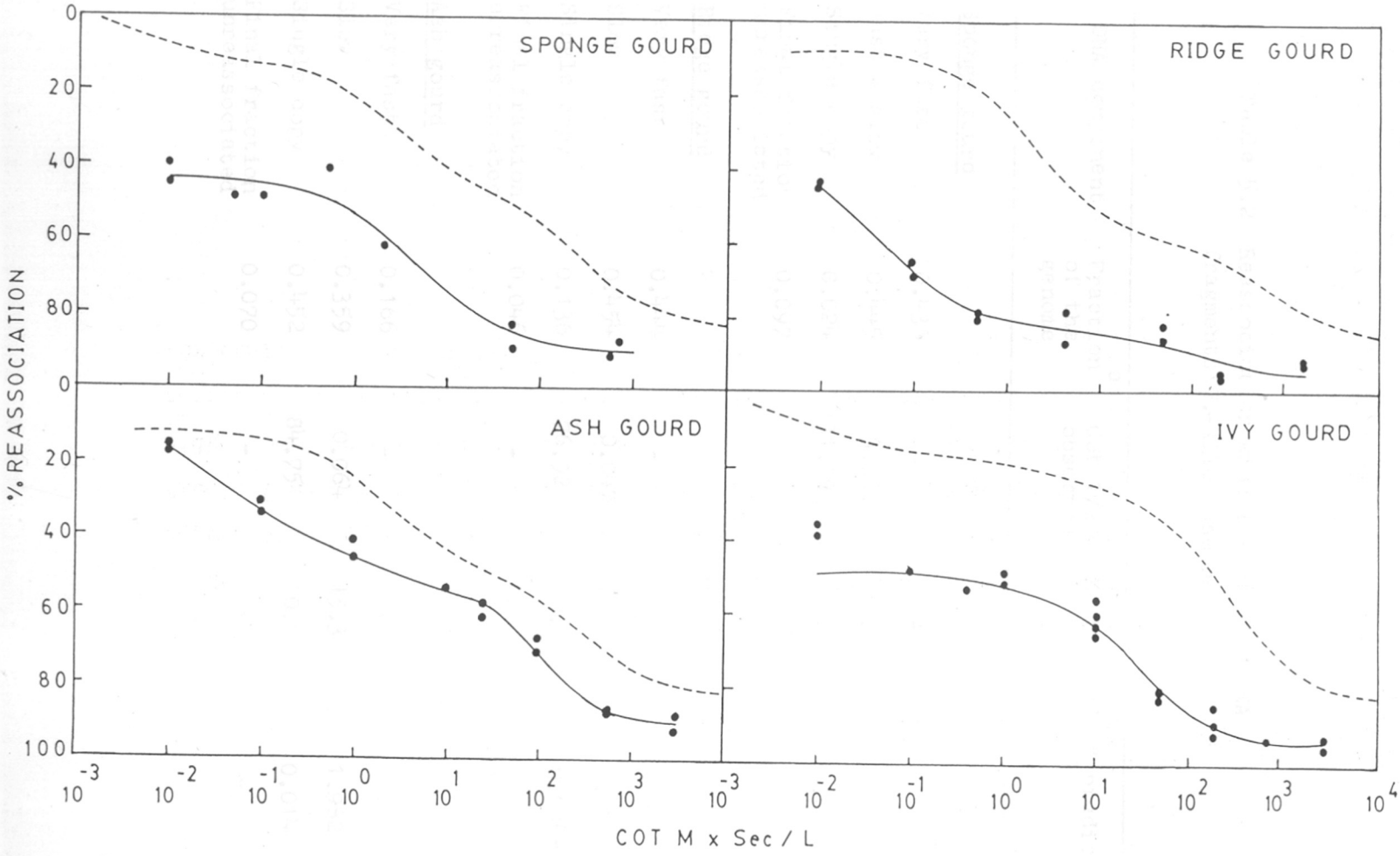

SPONGE GOURD
RIDGE GOURD
ASH GOURD
IVY GOURD
% REASSOCIATION
0
20
40
60
80
100
10^{-3}
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}
COT M x Sec / L

Table 5.2 Reassociation kinetics of long DNA fragments (7.4 kilo base pairs)

DNA component	Fraction of the genome[a]	Cot 1/2 observed[a]	K observed[b]	K predicted[c]
Sponge gourd				
Very fast	0.436	-	-	-
Fast + slow	0.445	3.62	0.276	1.43
Single copy	0.024	131.58	0.0076	0.0158
Final fraction unreassociated	0.097	-	-	-
Ridge gourd				
Very fast	0.444	-	-	-
Slow	0.454	0.047	21.41	2.15
Single copy	0.136	76.92	0.013	0.0068
Final fraction unreassociated	0.046	-	-	-
Ash gourd				
Very fast	0.166	-	-	-
Slow	0.359	0.064	15.32	1.982
Single copy	0.452	84.75	0.012	0.014
Final fraction unreassociated	0.070	-	-	-

Table 5.2 (contd.)

Ivy gourd				
Very fast + slow	0.490	-	-	-
Single copy	0.463	25.45	0.039	0.050
Final fraction unreassociated	0.049	-	-	-

a : Values obtained from the respective reassociation curves (Fig. 5.2)

b : $K = (\text{Cot } 1/2)^{-1}$

c : Calculated using the formula $K_1/K_2 = (L_1/L_2)^{1/2}$ where K_1 and K_2 are the rate constants for the reassociation of short (L_1) and long (L_2) fragments respectively (10).

sponge gourd and ridge gourd, even at the earliest Cot value of 0.01 M x sec/l. This indicates extensive interspersion of both highly repeated and/or fold back (fast-reassociating) sequences with middle repetitive (slow-reassociating) sequences. In sponge gourd, two distinct components, fast and slow, were distinguishable at a lower DNA size of 550 base pairs (Chapter III). However, at the higher DNA size, these two fractions are mixed to such an extent that their individual reassociation rates are obscured and they reassociate as a single second order component. The single copy component also reduces considerably, suggesting that single copy sequences are also interspersed with repetitive sequences. Between the two species, interspersion of repeated sequences within themselves appears to be more extensive in ridge gourd, where most of the reassociation of repetitive sequences is complete by Cot 5.0, a lower limit Cot as compared to Cot 50.0 at a size of 550 base pairs.

The presence of interspersion is further supported by comparing the rates of reassociation observed for repetitive and single copy sequences with those expected for the higher fragment size according to the formula $K_1/K_2 = (L_1/L_2)^{1/2}$ where L_1 and L_2 are the two fragment sizes under study and K_1 and K_2 are their rate constants (10). The rates observed are much different

from those expected, suggesting thereby the presence of sequence interspersion in these two species.

In ash gourd, the increase in binding to hydroxyapatite at the increased fragment size is comparatively very little. The repetitive fraction increases from 48% to 58% only. Evidence for the interspersion of very fast and slow sequences is provided by comparison of the observed and expected rate constants, wherein the observed rate is much faster than expected. The increase in reassociation is also more pronounced at low Cot values. However, the rate of reassociation of single copy DNA is not altered significantly. Thus a unique situation is encountered in ash gourd, wherein very little single copy fraction is interrupted by repeated DNA sequences.

In case of ivy gourd, the single copy fraction decreases from 75% to 40% at the higher DNA fragment size. Thus about a third of the single copy sequences are interrupted by repeats. The reassociation of repetitive component is complete by Cot 0.01, confirming the interspersion of very fast, slow as well as a part of the single copy sequences.

Hyperchromicity of repetitive duplexes of varying lengths: Another means of assessing the occurrence of DNA sequence interspersion is to study the optical melting behaviour

of reassociated duplexes isolated from DNAs of varying fragment lengths. If single-stranded tails are present on DNA which has been renatured to a repetitive Cot, the hyperchromicity of the duplex will be lower than that of native DNA and will be proportional to the fraction of DNA existing as duplex. This decrease in hyperchromicity can be used to calculate the duplex content in renatured fragments (11). Figure 5.3 shows the melting profiles of the DNAs of different fragment lengths reassociated to their respective repetitive Cot values, and the results are summarized in Table 5.3. Though in all the cases duplexes of higher thermal stability are formed, hyperchromicity of the duplexes decreases with increasing fragment size. Even the fraction bound to hydroxyapatite at the lowest size studied (550 base pairs) is not fully a duplex, meaning that most of the repetitive sequences are even shorter than 550 base pairs. The average duplex length reaches 2000-3500 base pairs at the maximum size studied, indicating that the unit of interspersion of an average single copy element with an average repetitive element is of that size.

Characterization of S1 nuclease resistant repetitive duplexes: An increase in hydroxyapatite binding and a decrease in hyperchromicity with an increase in DNA fragment size indicate the presence of extensive single

Fig. 5.3 : Thermal denaturation profiles of repetitive duplexes isolated at different fragment lengths of DNA and of S1 nuclease resistant duplexes, compared with native sheared DNAs. The fractions in 0.12 M sodium phosphate buffer (pH 6.8) were melted in a Gilford 250 spectrophotometer at a heat rate of 1°C/min. The graphs of relative absorbance [A_{260} (T_1) / T_{260} (98°C)] vs. temperature were plotted, where A_{260} (T_1) is the absorbance at the given temperature.

(▼ - ▼) Native 550 base pairs DNA

(x - x) S1 nuclease resistant repetitive duplexes.

(0 - 0) repetitive fraction isolated from 550 base pairs DNA

(Δ - Δ) repetitive fraction isolated from 2850 base pairs DNA

(▲ - ▲) repetitive fraction isolated from 5200 base pairs DNA

(▽ - ▽) repetitive fraction isolated from 7400 base pairs DNA

The thermal denaturation profiles of only a few fragment sizes have been included for the sake of simplicity.

Figure 5.3

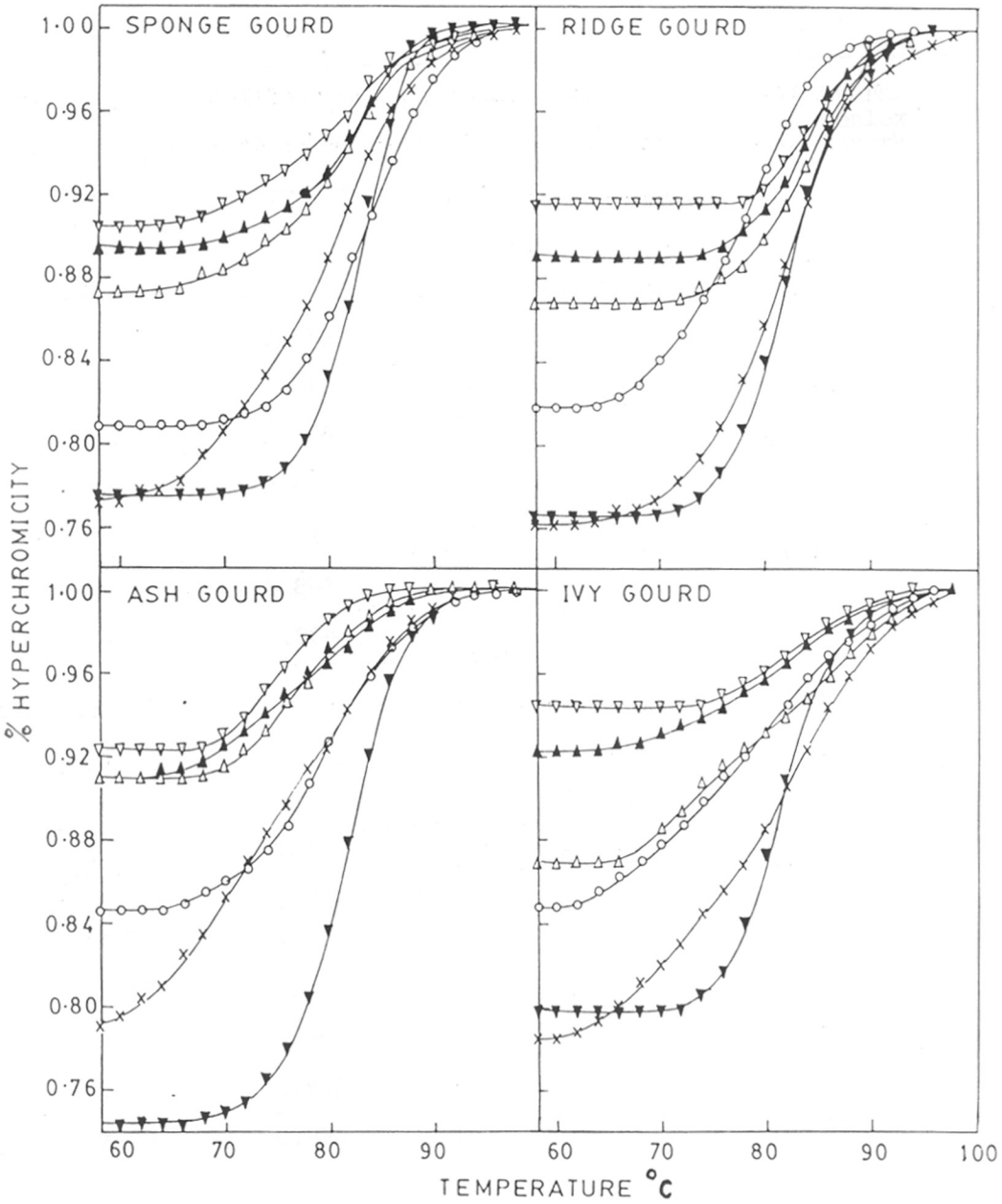

SPONGE GOURD
RIDGE GOURD
ASH GOURD
IVY GOURD
% HYPERCHROMICITY
1·00
0·96
0·92
0·88
0·84
0·80
0·76
60
70
80
90
100
TEMPERATURE °C

Table 5.3 Thermal denaturation properties of repetitive duplexes isolated from DNAs of different fragment lengths.

Species	Fraction[a] bound to hydroxy-apatite	Hyperch-[b] romicity	Tm[c] °C	Duplex[d] content	Average[e] duplex length
Sponge gourd					
Native 550 base pairs DNA	-	0.228	83.44	-	-
550	0.510	0.142	77.86	0.569	312.95
1500	0.714	0.122	80.20	0.469	703.50
2000	0.856	0.121	80.25	0.464	928.00
2850	0.808	0.118	80.30	0.449	1279.65
4600	0.838	0.115	82.28	0.434	1996.40
5200	0.841	0.106	82.20	0.389	2022.80
7400	0.864	0.102	82.70	0.369	2730.60
Ridge gourd					
Native 550 base pairs DNA	-	0.231	82.47	-	-
550	0.590	0.181	77.67	0.770	423.50
1500	0.700	0.139	82.97	0.575	862.50
2000	0.749	0.130	83.26	0.534	1068.00
2850	0.810	0.134	83.86	0.552	1573.20
4200	0.825	0.134	84.20	0.552	2318.40
4600	0.865	0.133	84.10	0.548	2520.80
5200	0.865	0.133	84.30	0.548	2849.60
7400	0.870	0.122	84.45	0.497	3379.60

Table 5.3 (contd.)

Ash gourd					
Native 550 base pairs DNA	-	0.224	82.66	-	-
550	0.480	0.145	74.40	0.627	344.85
1500	0.495	0.104	74.70	0.434	651.00
2850	0.585	0.091	77.75	0.372	1060.20
4200	0.590	0.105	78.00	0.438	1839.20
5200	0.605	0.089	78.00	0.363	1887.60
7400	0.610	0.076	79.60	0.302	2236.39
Ivy gourd					
Native 550 base pairs DNA	-	0.226	81.55	-	-
550	0.250	0.142	78.63	0.597	328.35
1500	0.430	0.123	80.40	0.506	759.00
2000	0.530	0.142	78.83	0.597	1149.00
2850	0.530	0.107	81.10	0.430	1225.51
4200	0.605	0.085	81.08	0.325	1365.00
4600	0.610	0.078	81.33	0.292	1343.00
5200	0.615	0.078	81.00	0.292	1518.40
7400	0.630	0.075	81.75	0.277	2049.80

Table 5.3 (contd.)

a : Obtained from Figs. 5.2 and 5.6

b : Calculated as $H = \frac{A_{260}\ (98^{\circ}C) - A_{260}\ (62^{\circ}C)}{A_{260}\ (98^{\circ}C)}$ (3)

c : Obtained from Fig. 5.3 or the respective melting curves.

d : Estimated using the formula

$$D = \frac{H - \text{single strand collapse}}{H_{\text{native sonicated DNA}} - \text{single strand collapse}}$$

Single strand collapse was determined as the hyperchromicity of denatured single copy DNA (4) and was 2.8% for sponge gourd, 1.5% for ridge gourd, 1.2% for ash gourd and 1.7% for ivy gourd.

e : Calculated as duplex content (D) x fragment size of DNA.

strand regions in the hydroxyapatite bound fractions. The actual amount of repetitive DNA in the genome can be assayed by the single-strand specific enzyme S1 nuclease. In all the four species, the fraction bound to hydroxyapatite at repetitive Cot reduces considerably after digestion with S1 nuclease (Table 5.4). The enzyme has been used here under such conditions that it can digest only single-stranded regions but not individual mismatched base pairs. The increased hyperchromicity of S1 nuclease resistant duplexes (Figure 5.3) is indicative of the removal of single stranded regions. The S1 nuclease resistant duplexes, moreover, show a lower Tm than that of native DNA, suggesting that the single mismatched sites are left intact after S1 nuclease action. The base mismatch in repetitive duplexes as calculated from the lowering in Tm varies from 2.7% in ivy gourd to 8.5% in ash gourd.

Fractionation of S1 nuclease resistant duplexes on Agarose A50 columns gives information about the relative abundance of the different size classes of repetitive DNA in the genome (Figure 5.4a and Table 5.5). All the four species show a major small molecular weight class (75-80% of the total repetitive DNA) having an average length of 300 base pairs, a minor high molecular weight excluded fraction (15-20%, length > 1500 base pairs) and a range of sizes (4-7%) between 300 and 1500 base pairs.

Table 5.4 S1 nuclease characterization of repetitive duplexes.

Species	S1 nuclease resistant fraction of the genome	Tm of S1 nuclease resistant fraction ^{o}C	% Base mismatch[a] in S1 nuclease resistant duplexes	% Hyper-chromicity of S1 nuclease resistant fraction
Sponge gourd	0.38 (5)	78.97±1.37 (10)	5.42	21.66±1.87 (10)
Ridge gourd	0.39 (3)	81.60±0.54 (6)	4.07	22.94±1.35 (6)
Ash gourd	0.27 (4)	75.67±1.36 (8)	8.54	20.59±2.56 (8)
Ivy gourd	0.20 (5)	80.53±1.80 (6)	2.71	22.38±0.85 (6)

a : % Base mismatch calculated as lowering in Tm by $1^{o}C$ corresponds to 1% base mismatch (1).
The Tms of native unsheared DNAs are $84.39^{o}C$, $85.67^{o}C$, $84.21^{o}C$ and $83.24^{o}C$ for sponge gourd, ridge gourd, ash gourd and ivy gourd respectively and were taken for comparison.
Figures in parentheses indicate the number of experiments conducted in each case.

Fig. 5.4a : Agarose A50 column chromatographic profiles of calf thymus DNA of known duplex lengths and of S1 nuclease resistant total repetitive duplexes. The size distribution of S1 resistant duplexes at limit repetitive Cot values was determined by gel filtration on an Agarose A50 (BioRad) column using 0.12 M sodium phosphate buffer (pH 6.8). The scans were obtained by monitoring the effluent with the help of a LKB Uvicord at 253 nm. The parallel curves are shifted vertically from each other by 2.0 units of the scale on the ordinate. The peaks marked 1 and 2 represent calf thymus DNA of known duplex length of > 1500 and 550 base pairs, respectively.

Figure 5.4a

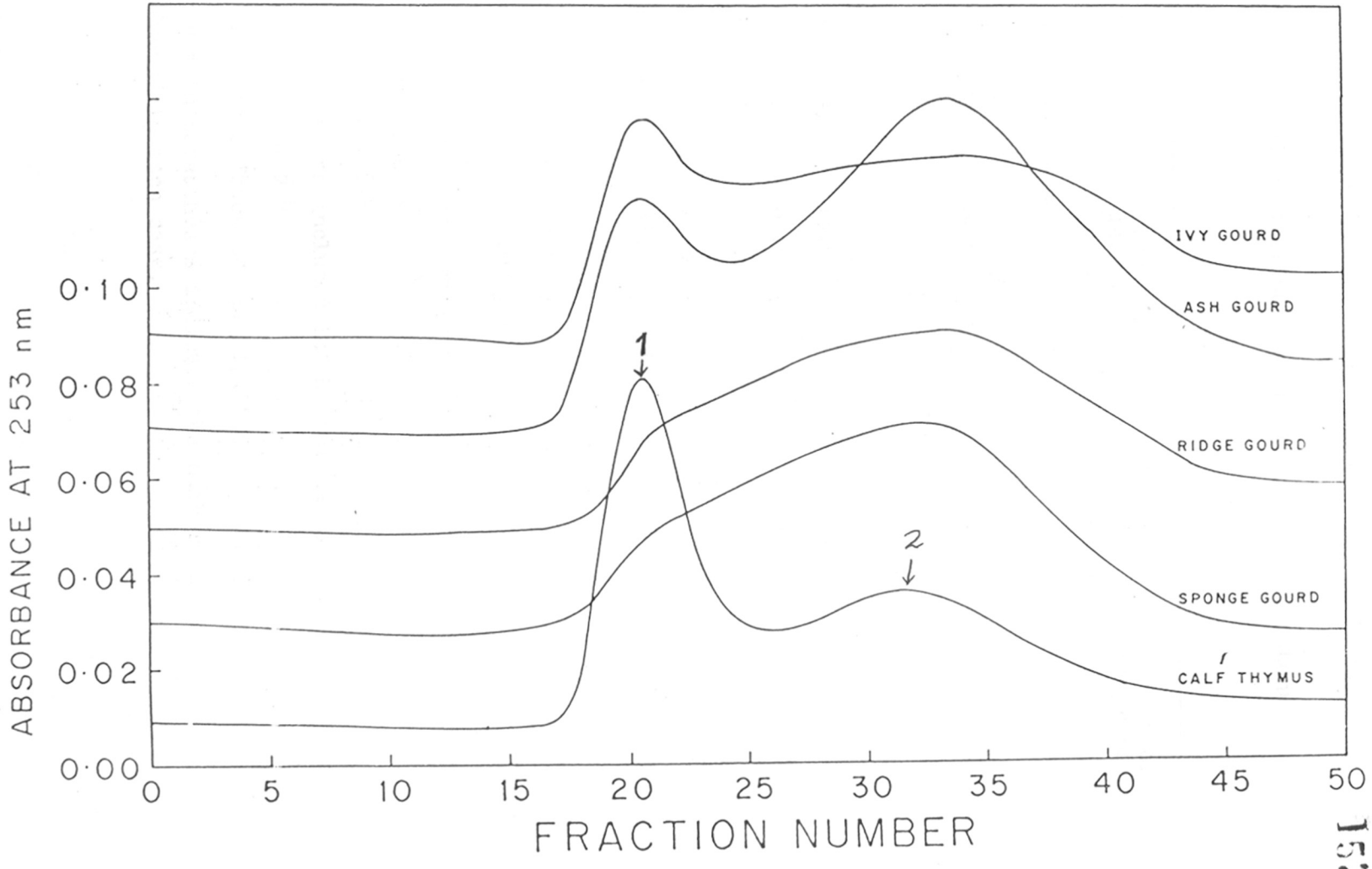

ABSORBANCE AT 253 nm
0·00
0·02
0·04
0·06
0·08
0·10
0
5
10
15
20
25
30
35
40
45
50
FRACTION NUMBER
IVY GOURD
ASH GOURD
RIDGE GOURD
SPONGE GOURD
CALF THYMUS
1
2

Table 5.5 Size distribution of S1 nuclease resistant fractions.

Species	Fraction of long repeats[a]	Fraction of repeats of intermediate lengths[b]	Fraction of short repeats[c]
Sponge gourd			
Cot 50.0 (4)	14.61 ± 2.21	6.62 ± 1.75	78.77 ± 0.47
Cot 0.1 (2)	76.24 ± 4.82	12.13 ± 3.82	11.70 ± 4.93
Ridge gourd			
Cot 50.0 (4)	15.60 ± 0.98	5.32 ± 1.37	79.08 ± 1.53
Cot 0.1 (2)	69.46 ± 6.30	2.70 ± 3.14	27.84 ± 5.10
Ash gourd			
Cot 25.0 (4)	20.67 ± 0.94	4.03 ± 1.84	75.30 ± 2.79
Cot 0.1 (2)	33.72 ± 8.81	37.92 ± 6.40	28.36 ± 4.74
Ivy gourd			
Cot 10.0 (4)	17.39 ± 4.38	4.20 ± 0.88	78.42 ± 5.26
Cot 0.1 (2)	71.07 ± 8.27	14.21 ± 1.65	14.72 ± 9.93

a : Fraction excluded from the column (Peak 1 fractions) in Fig. 5.4

b : Fractions included in the column, between peaks 1 and 2 in Fig. 5.4

c : Peak 2 fractions in Fig. 5.4

Figures in parentheses indicate the number of experiments conducted in each case.

The exact molecular weights of the two size classes were determined on agarose slab gels. The smaller size class was relatively homogeneous, with a mean size of about 300 base pairs, while the larger molecular weight class was composed of sizes ranging from 2000 to 4000 base pairs.

In sponge gourd, ridge gourd and ash gourd, the long and short repeats were melted to assess the extent of their sequence divergence (Figure 5.5). The long repeats showed a high hyperchromicity (18-20%) and Tm (76-82^{o}C) approaching that of native DNAs, indicating very little sequence divergence. The short repeats, however, exhibited very low hyperchromicities (4-10%) and Tm (64-71^{o}C), and thus a very high degree of sequence divergence.

Repetitive sequence classes reassociating at a lower Cot value: Experiments were also carried out where the DNAs of high molecular weight (> 10 kilo base pairs) were reassociated to Cot 0.1, and then treated with S1 nuclease followed by fractionated on Agarose A50 to determine the distribution of long and short repeats (Figure 5.4b, Table 5.5). The proportion of S1 nuclease resistant repeats at Cot 0.1 is 12.71%, 13.98%, 8.85% and 9.60% for sponge gourd, ridge gourd, ash gourd and ivy gourd respectively. Comparison of their size distribution

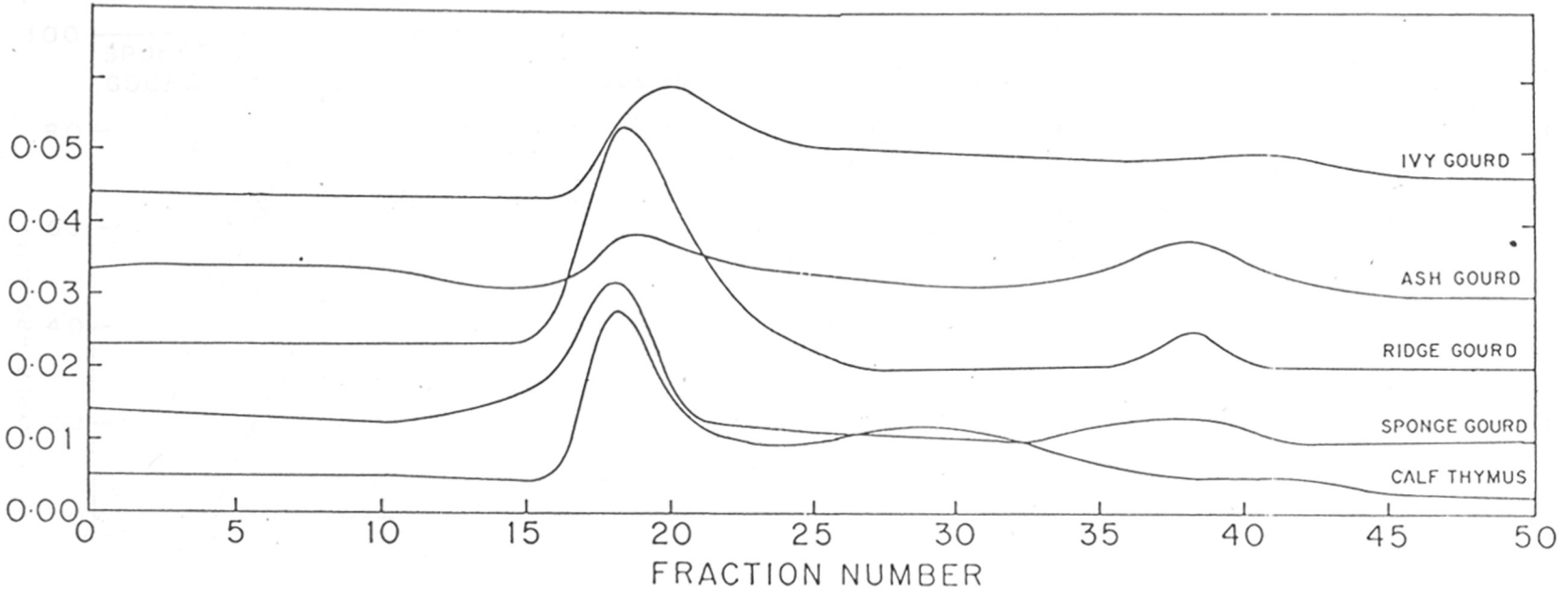


Figure 5.4b: Size distribution of S1 nuclease resistant repetitive duplexes isolated at Cot 0.1 M x sec/l. The profiles were obtained as described for Fig. 5.4a.

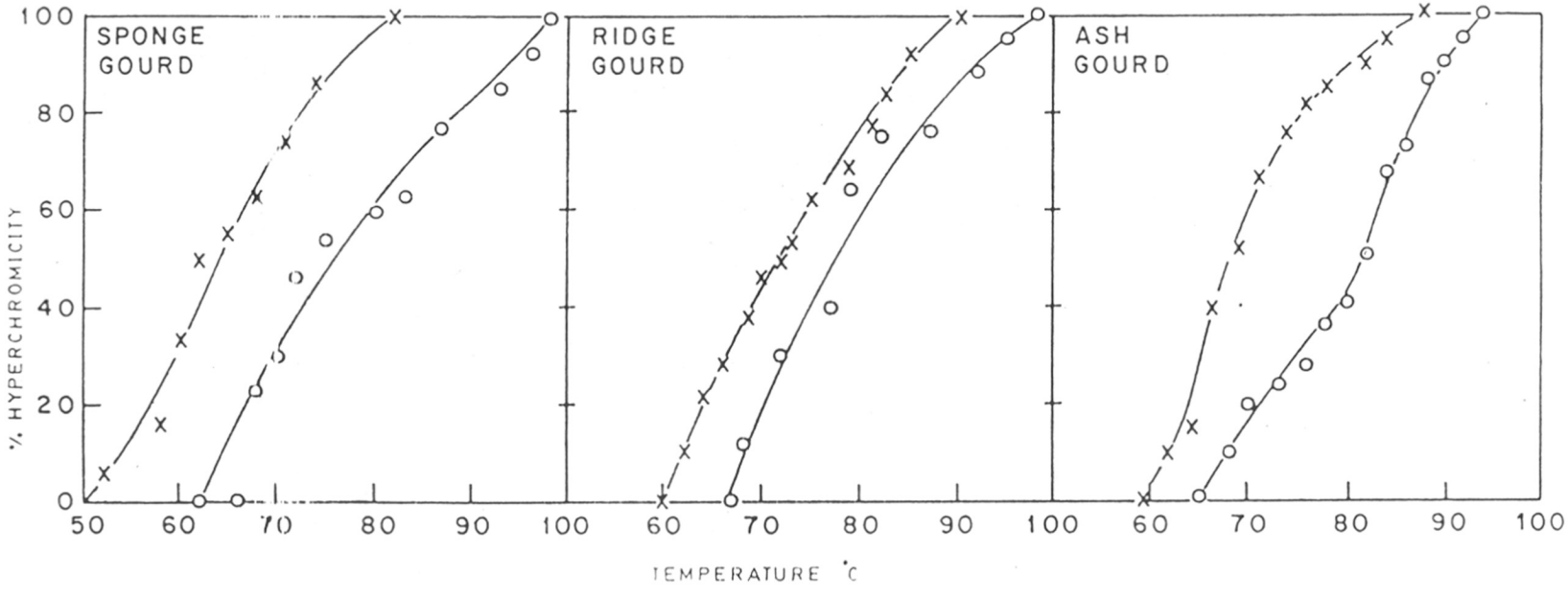


Figure 5.5 : Thermal denaturation profiles of long (0 - 0) and short (x - x) repeats isolated by Agarose A5C gel filtration.

profiles (Figure 5.4b) with those of total repetitive DNA (Figure 5.4a) shows that there is a considerable increase in proportion of long repeats and decrease in the proportion of short repeats. The latter are composed of very short sequences of about 150 base pairs. These results suggest that the fast-reassociating sequences are composed largely of long repeats (or tandemly placed short sequences) and the slow fraction is composed mainly of short repeats. The presence of lower size class in the fast fraction could be the result of interspersion of fast and slow reassociating components.

Length of interspersed single copy sequences: Interspersion of repetitive and single copy sequences will cause a certain amount of single copy DNA to get bound to hydroxyapatite at a Cot value at which only repetitive sequences are fully reassociated. Therefore, at the repetitive Cot, the amount of unreassociated single copy DNA existing as single stranded tails on reassociated duplexes will increase as a function of DNA fragment length. This will hold true only upto a fragment length which is equal to the average spacing between repetitive sequences. Thus the average length of single copy DNA can be estimated by plotting the amount of DNA bound to hydroxyapatite as a function of DNA fragment length.

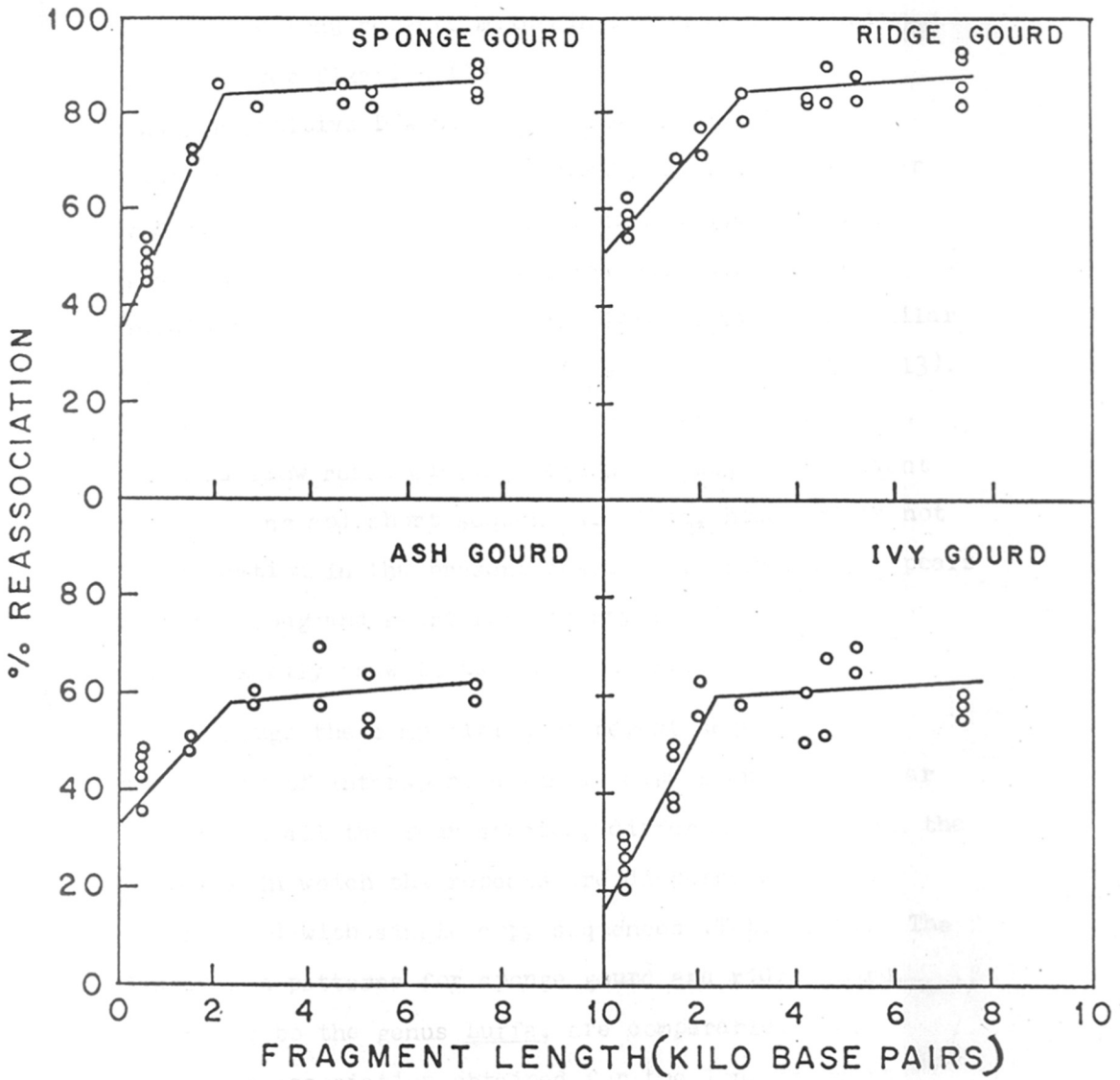
The results of such experiments are depicted in Figure 5.6. The curves for all the four species show a steep rise in the early part of the curve upto about 3000 base pairs, and then a more gradual increase in slope upto 7400 base pairs (highest size studied). An extrapolation of the point of change of slope to the X axis gives an average length of interspersed single copy DNA, which is 2100 base pairs for sponge gourd, 2900 base pairs for ridge gourd, 2400 base pairs for ash gourd and 2300 base pairs for ivy gourd.

DISCUSSION

The composition of populations of repetitive sequences in all the four species appears to be identical, with 75-80% of short repeats (about 300 base pairs long), 15-20% long repeats (2000-4000 base pairs long) and a small fraction (4-7%) with sizes ranging from 300 to 1500 base pairs. The total repetitive fractions show a moderate sequence divergence (3-8%), which can be either due to a recent origin of repeats or to a low rate of sequence divergence. The short repeats, however, have diverged more than the long repeats, which is in accordance with the result reported earlier in a few plant species (12,13). It is generally assumed that the short interspersed repeats evolve from long regions of tandemly repeated DNA. If this is so, then the long repeats should exhibit higher thermal

Fig. 5.6 : Percentage binding of repetitive DNA to hydroxyapatite as a function of DNA fragment length. DNAs of different fragment lengths were reassociated to limit repetitive Cot values (Cot 50.0 for sponge gourd and ridge gourd, Cot 25.0 for ash gourd and Cot 10.0 for ivy gourd, Chapter III). The reassociated duplexes were separated from unreassociated single strands by hydroxyapatite chromatography. A graph of % duplexes bound to hydroxyapatite vs. DNA fragment length was plotted.

Figure 5.6


SPONGE GOURD
RIDGE GOURD
ASH GOURD
IVY GOURD
% REASSOCIATION
FRAGMENT LENGTH(KILO BASE PAIRS)
0
20
40
60
80
100
2
4
6
8
10

stability either because they are relatively recent in origin and have got little chance to diverge (7) or because the long repeats maintain internal homogeneity by cross-over fixation (14). However, treatment of highly repetitive DNA with S1 nuclease and size fractionation indicate that majority of long repeats are represented in the fast reassociating class. Thus it appears that in these four genomes the short and long repeats belong to different repetitive families. Similar observations have also been made earlier in parsley (13). If the short repeats arise from the long repeats, then fast and slow reassociating sequences should be present in both long and short sequences. This, however, is not the situation in the present case. It, therefore, appears that the long and short repeats may be different evolutionarily as well as functionally.

Though the composition of repeat sequences and the length of interspersed single copy sequences appear similar in all the four species, differences exist in the patterns in which the repeats are dispersed with one another and with single copy sequences (Table 5.6). The dispersion patterns for sponge gourd and ridge gourd, belonging to the genus Luffa, are comparable. The maximum reassociation obtained for the long fragment size (7400 base pairs) at the highest Cot value studied

Table 5.6 Patterns of sequence organization in the Cucurbitaceae species

Species	Short period interspersion			Long period interspersion				Uninterspersed single copy DNA[c] % of total genome
	% of total genome	% slow repeats[a]	% Single copy[b]	% of total genome	% Fast repeats[a]	% slow repeats[a]	% single copy[b]	
Sponge gourd	70.46	28.46	42.00	15.04	8.07	1.47	5.50	14.50
Length in base pairs	-	300	2100	-	2000-4000	300	2100	>7400
Ridge gourd	46.45	26.95	19.50	40.55	8.16	3.89	28.5	13.00
Length in base pairs	-	300	2900	-	2000-4000	300	2900	>7400
Ash gourd	41.82	17.82	24.00	16.18	6.67	2.51	7.00	42.00
Length in base pairs	-	300	2400	-	2000-4000	300	2400	>7400
Ivy gourd	28.27	14.27	14.00	33.73	4.32	1.41	28.00	38.00
Length in base pairs	-	300	2300	-	2000-4000	300	2300	>7400

Table 5.6 (contd.)

a : The difference in agarose A50 size distribution profiles of S1 nuclease resistant fractions isolated at limit repetitive Cot and at Cot 0.1 gives the distribution of long and short repeats in the fast and slow reassociating components. Fast component lacks any short repeats and slow component lacks long repeats. Of the short repeats, the ones reassociating at Cot 0.1 are interspersed with fast (long) repeats and single copy sequences in a long period pattern, while those reassociating between Cot 0.1 to limit repetitive Cot are interspersed with single copy sequences in a short period pattern.

b : % Single copy DNA interspersed is calculated as % reassociation at repetitive Cot at highest fragment size studied - % of S1 nuclease resistant duplexes. Minimum % single copy DNA interspersed with fast repeats is calculated as % reassociation at higher fragment size at Cot 0.1 - total S1 nuclease resistant fraction which is the maximum possible reassociation due to total interspersion of repeated sequences only within themselves. % single copy DNA interspersed with slow repeats is equal to total interspersed single copy - that interspersed with fast repeats.

c : Fraction of the genome remaining unreassociated at the highest Cot value studied at the highest fragment size.

(Cot 1000 to 5000 M x sec/l) is about 85-90%. It is not known whether the fraction remaining unreassociated (10-15%) is composed of uninterrupted single copy sequences. However, since the maximum reassociation is obtained at lower Cot values (Cot 50 to 100 M x sec/l) itself, it is likely that the unreassociated fraction consists of degraded DNA rather than native unreassociated DNA. Thus the entire single copy fraction of these two genomes appears to be totally interspersed with repetitive fractions. Similarly, the extensive binding of DNA to hydroxyapatite at lower Cot values (Cot 0.01 M x sec/l) is indicative of extensive interspersion of various repeat classes within themselves. However, the fraction of genome of sponge gourd involved in short period interspersion is more than that in long period interspersion pattern, as compared to the genome of ridge gourd. The predominance of short period pattern in sponge gourd compared to ridge gourd suggests that the genome of sponge gourd is 'turning over' at a faster rate than that of ridge gourd (15).

In ash gourd, the interesting observation is that the genome shows very little interspersion of single copy DNA with repetitive sequences, in spite of the large amount of repetitive DNA content. The fast and slow repeats are interspersed to some extent. The maximum reassociation

obtained at the repetitive Cot for the highest fragment size is 58%, which is slightly higher than that at lower fragment size (48%). This indicates that at least 40% of the genome is composed of uninterrupted single copy sequences. This is a unique situation, considering the general tendency of repeated DNA to be dispersed throughout the genome rather than remaining in long tandem arrays. The same situation is observed in ivy gourd also, wherein about 40% of the genome is composed of long stretches of single copy DNA. It is possible that these two genomes are 'turning over' (15) at a much slower rate than those of sponge gourd and ridge gourd. High rates of 'turnover' lead to reduction in age of repetitive families (i.e. less base sequence divergence) while low rates lead to accumulation of sequence divergence. This is indeed true for ash gourd which shows more divergence in repetitive sequences than the other three species. However the situation appears quite paradoxical in ivy gourd where repeated sequences have diverged by only about 2%. Ivy gourd also shows the lowest amount of repetitive DNA (Chapter III). It, therefore, appears that the repetitive sequences in ivy gourd have undergone divergence to such an extent that they no longer reassociate with one another with repetitive rate under standard conditions (i.e. they have undergone more than 25% divergence) (1). It is also possible that due to very low amount of repetitive DNA,

translocation and dispersion events lead to interspersion of single copy DNA with single copy DNA itself and thus would not result in extensive repeat-single copy interspersion.

During the initial stages of interspersion pattern studies, mainly two types of organizations were recognized; i.e. the Xenopus type (short period) (5) and the Drosophila type (long period) (16). A large number of organisms showed the former pattern, which led to the hypothesis on the function of interspersed short repeats in regulation of gene expression (17,18). Todate, however, many more genera from phylogenetically diverse taxa have been studied, e.g. fungi (19-22), aves (23,24), nematode (25) and higher plants (9,26,27) and these studies have exhibited the predominance of long period pattern. It thus appears that variations in interspersion patterns need not be correlated with gene regulation, as one would not expect the basic mechanism of gene regulation to differ in various taxa. Rather, interspersion patterns should be more related to the dynamics of amplification and translocation of DNA sequences.

The present investigation has shown that though the four plant species are very similar in morphological and cytological features, their genomes are organized in vastly different patterns. We suggest that such differences

at the organization level could be a cause for fertility barriers which are so well known in Cucurbitaceae (Chapter I). Since a single mutation cannot bring about the origin or establishment of a species, many factors have to work hand in hand to bring about speciation. One of the factors that can promote speciation is the origin and establishment of a few to many differences in chromosome structure including submicroscopic rearrangements of nucleotides, or chromosomal repatterning (28). The biological incompatibility can be envisaged as taking effect only between two individuals that are maximally dissimilar in their repetitive families (29). This indeed seems to be the case in the present four species of Cucurbitaceae, where the species show extreme diversities in the structure and organization of repeated DNA sequences.

References

1. Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) In: Methods in Enzymol (Grossman, G.L. and Moldave, K. (eds). Vol.29E, Academic Press, New York, pp.363-418.

2. Mandel, M. and Marmur, J. (1968) In: Methods in Enzymol. (Grossman, G.L. and Moldave, K. eds) Vol. 12B, Academic Press, New York, pp. 195-206.

3. Zimmerman, J.L. and Goldberg, R.B. (1977) Chromosoma (Berl.) Chromosoma, 59, 227-252.

4. Graham, D.E., Neufeld, B.R., Davidson, E.H. and Britten, R.J. (1974) Cell 1, 127-137.

5. Davidson, E.H., Hough, B.R., Amenson, C.R. and Britten, R.J. (1973) J.Mol.Biol. 77, 1-23.

6. Goldberg, R.B., Crain, W.R., Ruderman, J.V., Moore, G.M., Barnett, T.R., Higgins, R.C., Gelfand, R.A., Galau, G.A., Britten, R.J. and Davidson, E.H. (1975), Chromosoma (Berl.) 51, 225-251.

7. Britten, R.J., Graham, D.E., Painchand, D.M. and Davidson, E.H. (1976) J.Mol.Evol. 9, 1-23.

8. Sachs, D.H. and Painter, E. (1972) Science 165, 781-782.

9. Wimpee, C.F. and Rawson, J.R.Y. (1979) Biochim.Biophys. Acta 562, 192-206.

10. Wetmur, J.G. and Davidson, N. (1968) J.Mol.Biol. 31, 349-370.

11. Vogt, V.M. (1973) Eur.J.Biochem. 33, 192-200.

12. Goldberg, R.B. (1978) Biochem.Genet. 16, 45-68.

13. Kiper, M. and Herzfeld, F. (1978) Chromosoma (Berl.) 65, 335-351.

14. Smith, G.P. (1976) Science 191, 528-535.

15. Thompson, W.F. and Murray, M.G. (1980) In: Davies, D.R. and Hopwood, D.A. (eds.) Proceedings of the 4th John Innes Symposium, The Plant Genome pp. 31-45.

16. Manning, J.E., Schmid, C.W. and Davidson, N. (1965) Cell 4, 141-155.

17. Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357.

18. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.

19. Ullrich, R.C., Kohorn, B.P. and Specht, C.A. (1980) Chromosoma (Berl.) 81, 371-378.

20. Harshey, R.M., Jayaram, M. and Chamberlain, M.E. (1979) Chromosoma (Berl.) 73, 143-151.

21. Krumlauf, R. and Marluf, G.A. (1979) Biochemistry 18, 3705-3713.

22. Hudspeth, M.E.S., Timberlake, W.E. and Goldberg, R.B. (1977) Proc.Natl.Acad.Sci. 74, 4332-4336.

23. Wagenmann, M., Epplen, J.T., Bachmann, K., Engel, W. and Schmidtke, J. (1981) Experientia 37, 1274-1276.

24. Epplen, J.T., Leipoldt, M., Engel, W. and Schmidtke (1978) Chromosoma (Berl.) 69, 307-321.

25. Schachat, F., O'Conner, D.J. and Epstein, H.F. (1978) Biochim.Biophys.Acta 520, 688-692.

26. Gupta, V.S., Gadre, S.R. and Ranjekar, P.K. (1981) Biochem.Biophys.Acta 656, 147-154.

27. Cullis, C.A. (1981) Biochim.Biophys.Acta 652, 1-15.

28. Stebbins, G.L. (1982) In: Mechanisms of Speciation Alan R. Liss New York, pp.21-39.

29. Dover, G. (1982) In: Mechanisms of Speciation Alan R. Liss New York, pp.435-459.

CHAPTER VI

HOMOLOGIES OF REPEATED DNA SEQUENCES

HOMOLOGIES OF REPEATED DNA SEQUENCES

SUMMARY

The technique of DNA-DNA hybridization has been used to assess the extent of homology of repeated DNA sequences among the four Cucurbitaceae species. In the two Luffa species, namely, sponge gourd and ridge gourd, a major proportion (95%) of the repeated DNA sequences is homologous. These sequences also have a very limited (0.65 - 1.27) base sequence divergence. Therefore, it appears that these two species have separated from each other relatively recently. However, the homology of repetitive sequences of ash gourd and ivy gourd with those of sponge gourd is limited (25-50%). The repeated DNA sequences also show a large degree of nucleotide sequence divergence (3.56 - 8.00%). These two species thus seem to have either undergone extensive sequence divergence since their origin from a common ancestor, or have arisen independently.

INTRODUCTION

Interspecies molecular comparisons have evolved from the study of biochemical properties of proteins (e.g. electrophoretic mobility, antigenecity, amino acid sequences of proteins)and finally the nucleotide sequences

of genes that code for these proteins (1-4). However, a disadvantage of this fine scale analysis is that only a minute fraction of the total genome can be compared. An alternate approach is the analysis of DNA sequence relationships of the genome by hybridization - melting technique which allows simultaneous study of a large number of DNA sequences (5). DNA sequences from two species can be reannealed in vitro and the hybrid duplexes can be subjected to thermal dissociation to assess the extent of DNA sequence homology. Such comparisons of DNAs can yield quantitative data on relatedness and divergence among organisms and thus are extremely useful in systematics (5-8).

Most eukaryotic genomes consist of repeated and single copy DNA sequences. The single copy sequences are known to have evolved essentially by random base substitutions or point mutations. The hybridization of single copy sequences would thus provide a valid measure of the extent to which base substitutions have occurred since the separation of any two lineages (9,10). Repetitive DNA sequences, on the other hand, evolve mostly by events such as amplification, reamplification, deletion and translocation. Intra - and inter-species hybridization reactions among repeated DNA sequences can, therefore, provide information about their origin and evolution (10).

In the previous chapters, we have described (i) how the repetitive sequences in cucurbits vary with respect to their proportions in the genome, copy number and kinetic complexities; (ii) what is the nature of repetitive families and (iii) what are the dispersion patterns of repeats among themselves and with single copy DNA sequences. However, in order to draw conclusions about the evolutionary origin and divergence of these sequences, it is essential to study their sequence relatedness. We have, therefore, used the technique of DNA-DNA hybridization to compare the homologies of repetitive DNAs of ridge gourd, ash gourd and ivy gourd to that of sponge gourd.

MATERIALS AND METHODS

Extraction, shearing and sizing of total unlabelled DNAs: DNAs were extracted from seedlings of sponge gourd, ridge gourd and ash gourd and fruits of ivy gourd, as described in chapter II. All the DNAs were sheared to an average size of 550 base pairs by sonication using Biosonic III (Chapter III).

Isolation of repetitive DNA of sponge gourd: The 550 base pairs long DNA of sponge gourd was incubated in 0.12 M sodium phosphate buffer (pH 6.8) to Cot 50.0 M x sec/l (limit repetitive Cot, Chapter III). The reassociated

fraction was separated from unreassociated single strands by hydroxyapatite chromatography as explained in chapter III.

Preparation of ^{32}P labelled DNAs: 2 - 3 µg of native, unsheared sponge gourd DNA (molecular weight > 10 kilo base pairs) in 10 mM Tris-HCl (pH 8.0) and 2 - 3 µg of total repetitive DNA of sponge gourd (Cot 50.0 fraction) in 0.12 M sodium phosphate buffer (pH 6.8) were labelled by using the nick-translation procedure of Rigby et al.(11). While in the former case. DNase I was added to a final concentration of 1×10^{-4} U/µg of DNA, in the latter case no DNase I was used, and the reaction volume was adjusted such that the concentration of phosphate did not exceed 0.05 M. The reaction mixture (200 µl) consisted of 50 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 20 µM each of unlabelled dATP, dCTP and dGTP. Commercial (α ^{32}P) dTTP (Amersham or New England Nuclear, specific activity 3000 Ci/mM) was added to a final concentration of 90-100 picomoles. E.coli DNA polymerase I was added to a final concentration of 2-3 U/µg and the reaction mixture was incubated at 15°C for 90 min. The progress of incorporation of labelled nucleotide was monitored in aliquots. The required specific activity was normally attained after 90 min. The reaction was terminated by adding sodium dodecyl sulphate to an effective concentration

of 0.1% and the mixture was deproteinized twice with freshly distilled phenol saturated with Tris and once with chloroform - isoamyl alcohol (24:1, v/v). The aqueous layer containing DNA was adjusted to 0.12 M sodium phosphate buffer (pH 6.8) and loaded on a hydroxyapatite column, previously equilibrated with the same buffer. The unincorporated nucleotides and single-stranded material were removed by washing the column extensively with 0.12 M sodium phosphate buffer (pH 6.8). The α-^{32}P labelled double-stranded DNA was then eluted with 0.4 M sodium phosphate buffer (pH 6.8). When unsonicated, native DNA was used for nick-translation, it was sonicated to 550 base pairs after nick-translation.

All the counting was done by Cerenkov Counting Method (12) using a Beckman liquid scintillation counter (LS 100C). The volume of solution (3 ml) taken for counting was kept constant.

Sizing of labelled DNAs: Sizing of labelled DNAs was done by agarose slab gel electrophoresis (Chapter II). The labelled DNAs (either the total DNA of sponge gourd sonicated to 550 base pairs after labelling, or the total repetitive DNA of sponge gourd obtained by reassociating the 550 base pairs long DNA to Cot 50.0 and then labelling the reassociated fraction) were loaded on agarose slab gels and the gels were run with appropriate

molecular weight markers. After electrophoresis, the mobility of DNA fragments of known molecular weight was marked. The tracks containing the labelled DNAs were cut to 1 cm pieces and counts were taken. Molecular weight of labelled DNAs was then calculated according to the mobility of the size markers.

Hybridization - melting of DNAs: By separating the two strands of DNA of one organism and recombining them with the separated DNA strands of another organism, a heteroduplex or hybrid DNA is produced. The quantity of hybrid DNA is easily measurable when one of the two strands is radiolabelled. The thermal stability of the hybrid DNA molecule is a function of the complementarity in the two component strands. The difference in melting temperatures of the homologous and heterologous DNA hybrids is a measure of the base sequence divergence between these DNAs (5,13).

We have used two complementary approaches to study the homologies in the repetitive DNA sequences in the four Cucurbitaceae species. In the first set of experiments, the labelled total sponge gourd DNA (550 base pairs) was hybridized to the cold, sonicated (550 base pairs) DNAs of sponge gourd, ridge gourd, ash gourd and ivy gourd, to their repetitive Cot values. In the second set of experiments, the labelled, total repetitive DNA

of sponge gourd (Cot 50.0 fraction) was hybridized to cold sheared DNAs of sponge gourd, ridge gourd, ash gourd and ivy gourd. Thermal elution of homologous and heterologous duplexes was carried out in each case as described in the following section.

Use of total DNA of sponge gourd as tracer

1) Homologous hybridization: a ^{32}P-labelled sponge gourd DNA (550 base pairs) was diluted about 5000-fold with unlabelled sonicated sponge gourd DNA. About 50,000 cpm (0.01 μg) of tracer DNA and 50-60 μg of driver DNA were used in each experiment. The DNA samples were denatured in sealed ampoules in a boiling water bath for 10 min. and were incubated to Cot 10.0, Cot 25.0 and Cot 50.0. The incubations were done in 0.12 M or 0.36 M sodium phosphate buffer (pH 6.8). In the latter case, the rate of reassociation was corrected (14). After incubation, the samples were diluted to 0.12 M sodium phosphate buffer (pH 6.8) and were loaded on hydroxyapatite column previously equilibrated with the same and maintained at 62°C. The counts in the initial washes with 0.12 M sodium phosphate buffer (pH 6.8) were considered to represent the unreassociated DNA. The duplex DNA bound to hydroxyapatite was recovered by stepwise raising of the column temperature in 5°C increments upto 100°C. The single stranded DNA resulting from dissociation of duplexes at

each temperature increment was eluted with 2-3 bed volumes of 0.12 M sodium phosphate buffer (pH 6.8) followed by a final wash with 0.4 M sodium phosphate buffer (pH 6.8) at 100°C. All the DNA eluting at temperatures higher than 62°C was taken as the reassociated fraction. Amount of DNA eluted as single strands at each step of thermal elution was estimated by Cerenkov Counting (12). Recovery of counts from hydroxyapatite was 95-105% of the amount loaded. The fraction reassociated (R) was calculated as:-

$$\% R = \frac{\text{Counts eluted after } 62^{\circ}C}{\text{Counts eluted at } 62^{\circ}C \text{ with 0.12 M sodium phosphate buffer, pH 6.8} + \text{counts eluted after } 62^{\circ}C}$$

The graph of cumulative counts vs. temperature gave the denaturation profiles, while the graph of counts eluted between successive temperature rises vs. temperature gave derivative profiles of the same. The temperature at which 50% of the reassociated duplexes were eluted was taken as the temperature of dissociation (Tm) of both homologous and heterologous duplexes.

A self-reassociation reaction of tracer DNA was carried out under similar conditions and the corrections were applied accordingly.

2) Heterologous hybridization: α ^{32}P-labelled total sponge gourd DNA was hybridized with total sonicated (550 base pairs) DNAs of ridge gourd, ash gourd and ivy gourd. The reassociations were carried out at the repetitive Cot of driver DNAs, i.e. Cot 50.0 for ridge gourd, Cot 25.0 for ash gourd and Cot 10.0 for ivy gourd. A self-reassociation reaction for tracer was also carried out for each experiment and the correction was applied accordingly. For each heterologous reaction, the extent of homology was determined from the extent of homologous reassociation at that Cot value. The melting temperatures and derivative melting profiles were obtained as described in the previous section.

Use of total repetitive DNA of sponge gourd as tracer

1) Homologous hybridization: α ^{32}P-labelled total repetitive DNA of sponge gourd was hybridized with cold, sonicated DNA of sponge gourd to Cot 50.0.

2) Heterologous hybridization: The labelled, total repetitive DNA of sponge gourd was hybridized with cold, sonicated DNAs of ridge gourd, ash gourd and ivy gourd to Cot 50.0, Cot 25.0 and Cot 10.0 respectively.

A tracer-tracer self reassociation reaction was carried out simultaneously with the above homologous and heterologous reactions. All values of reassociation and melting profiles were calculated as described earlier.

Control experiments: Total sonicated DNA (550 base pairs) of wheat was labelled by nick-translation and was hybridized with excess cold sonicated (550 base pairs) wheat DNA or barley DNA to Cot 50.0 (7) or Cot 90.0 (15). Self-reassociation of tracer DNA was also carried out in each case.

RESULTS

Control experiments: A control experiment was carried out using total sonicated wheat DNA as tracer. Labelled wheat DNA was hybridized with excess cold sheared (550 base pairs) barley DNA to a Cot value of 50.0 M x sec/l and 90.0 M x sec/l (7,15). Both Cot values are sufficient to allow all the repeated sequences to form duplexes. Homologous reassociation values (wheat-wheat) are 76.67% (± 3.45%) and heterologous reassociation values (wheat-barley) are 63.62% (± 3.58%). After correction for self-reassociation of the tracer (24.97% ± 2.95%), the homology between the repeated sequences of wheat and barley is estimated to be 70%. The values of homologous and heterologous reassociations and the extent of homology of repeated DNA sequences between these two species compare well with the reported values (7,15).

Use of total sponge gourd DNA as tracer

1) Homologous hybridizations: Since the limit repetitive Cot values of ridge gourd, ash gourd and ivy gourd are

50.0, 25.0 and 10.0 respectively, the self-reassociation of tracer DNA and the homologous hybridization of tracer DNA with total, cold sponge gourd DNA were carried out at these three Cot values. The tracer-tracer self-reassociation values are 5.61%, 5.60% and 9.93% for incubations upto Cot 10.0, 25.0 and 50.0 respectively. Applying these corrections, the homologous reassociations reach 47.5% at Cot 50.0, 28.5% at Cot 25.0 and 28.6% at Cot 10.0 (Table 6.1a). All the homologous reassociation values are comparable with those obtained in the reassociation kinetics studies of 550 base pairs DNA of sponge gourd (51%, 43%, 42%, Chapter III).

The thermal stabilities of homologous duplexes reassociating at Cot 10.0, Cot 25.0 and Cot 50.0 are 77.88°C, 78.24°C and 79.30°C respectively (Table 6.2a, Fig.6.1a) and thus are about 4.14 to 5.56°C less than that of native sheared sponge gourd DNA (83.44°C, Chapter IV). This indicates the presence of a sequence divergence of 4.14 to 5.56% in the repetitive sequences of sponge gourd DNA (14).

2) Heterologous hybridization: The reassociation values of heterologous hybridization of total sponge gourd DNA as tracer with ridge gourd DNA at Cot 50.0, ash gourd DNA at Cot 25.0 and ivy gourd DNA at Cot 10.0 are 39.89%, 11.88% and 12.28% respectively, after the appropriate

Table 6.1a Reassociation of labelled total sponge gourd DNA to repetitive DNAs of cucurbits.

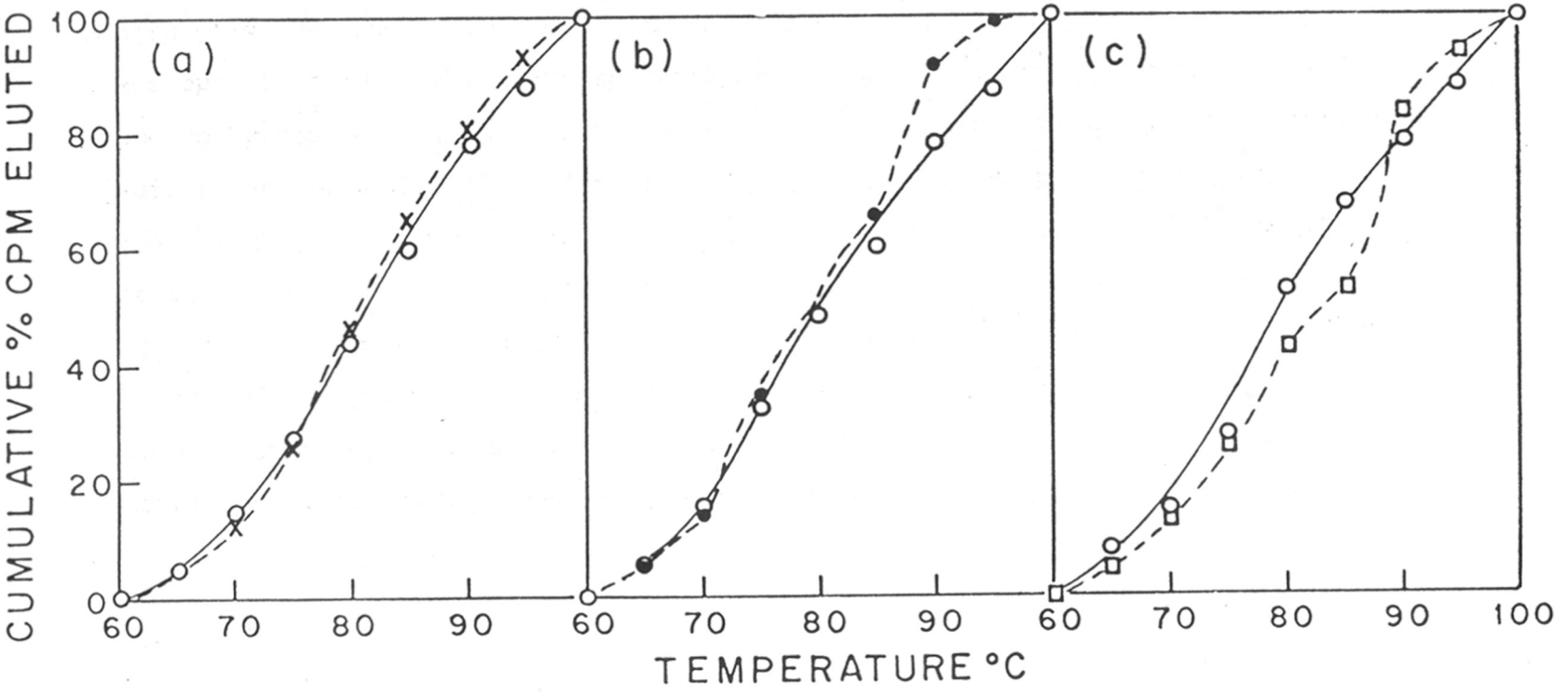
Species	Tracer incubated to Cot M x sec/l	% Reasso-ciation	% self-reasso-ciation of tracer	Correc-ted % Reasso-ciation	Average % reasso-ciation	Average % homology
Sponge gourd	50.0	49.78	12.10	37.68	47.47	100.0
		61.50	12.10	49.40		
		60.94	5.60	55.34		
Ridge gourd	50.0	44.16	12.10	32.06	39.89	84.03
		47.86	12.10	35.76		
		57.74	4.90	51.84		
Sponge gourd	25.0	33.74	5.60	28.14	28.54	100.00
		33.65	5.60	28.05		
		35.03	5.60	29.43		
Ash gourd	25.0	17.56	5.60	11.96	11.88	41.64
		17.44	5.60	11.84		
		17.45	5.60	11.85		
Sponge gourd	10.0	35.03	5.60	29.43	28.57	100.00
		33.74	5.60	28.14		
		33.74	5.60	28.14		
Ivy gourd	10.0	17.66	5.60	11.06	12.28	42.98
		17.35	5.60	11.75		
		20.63	5.60	14.03		

Table 6.1b Reassociation of isolated total repetitive DNA of sponge gourd to repetitive DNAs of cucurbits.

Species	Tracer incubated to Cot M x sec/l	% Reasso-ciation	% self-reasso-ciation of tracer	Corrected % Reasso-ciation	Average % Reasso-ciation	% Homology
Sponge gourd	50.0	79.92	12.90	67.02	65.81	100.00
		79.77	12.34	67.43		
		75.33	12.34	62.99		
Ridge gourd	50.0	73.68	12.90	60.78	62.16	94.46
		75.26	12.34	62.92		
		73.52	10.55	62.97		
Ash gourd	25.0	51.83	12.90	38.93	31.78	48.36
		41.04	12.34	28.70		
		40.06	12.34	27.72		
Ivy gourd	10.0	31.44	12.34	19.10	17.12	26.01
		26.43	12.34	14.09		
		28.89	10.73	18.16		

Figure 6.1a : Thermal dissociation profiles of duplexes formed between sponge gourd ^{32}P total sheared DNA (550 base pairs) and unlabelled sheared DNAs from sponge gourd, ridge gourd, ash gourd and ivy gourd. The samples were heat-denatured and incubated at 62°C to (a) Cot 50.0 for ridge gourd, (X - X), (b) Cot 25.0 for ash gourd (●--●) and (c) Cot 10.0 for ivy gourd (□--□). Homologous reassociations of sponge gourd DNA were also carried out to (a) Cot 50.0, (b) Cot 25.0 and (c) Cot 10.0, shown as (0 - 0). After incubations the samples were passed over hydroxyapatite at 62°C. Columns were washed with 0.12 M sodium phosphate buffer (pH 6.8) to remove single-stranded DNA. Thermal elution was then performed by raising the temperature in increments of 5°C and eluting with 0.12 M sodium phosphate buffer at each temperature, followed by a final elution with 0.4 M sodium phosphate buffer at 100°C. Results are presented as cumulative percentages of initially bound DNA eluted at each temperature.

Figure 6.1a


CUMULATIVE % CPM ELUTED
100
80
60
40
20
0
(a)
(b)
(c)
60
70
80
90
60
70
80
90
60
70
80
90
100
TEMPERATURE °C

corrections for self-reassociation of tracer DNA (Table 6.1a). Comparing these values with the respective values in homologous reactions, the extent of sequence homology with sponge gourd DNA is 85% for ridge gourd, 42% for ash gourd and 43% for ivy gourd. It thus appears that the repetitive families of ridge gourd are very similar to those of sponge gourd, while those of ash gourd and ivy gourd have very limited sequences in common with sponge gourd DNA.

The heterologous repetitive DNA duplexes of sponge gourd DNA with ridge gourd, ash gourd and ivy gourd DNAs show melting temperatures of 78.65°C, 74.68°C and 76.86°C respectively (Table 6.2a). These values correspond to a base sequence divergence of 0.65°C, 3.56°C and 1.02°C respectively when compared to the melting temperatures of the homologous duplexes. These results also indicate that there is very little sequence divergence between DNAs of sponge gourd and ridge gourd. Ash gourd and ivy gourd DNAs, on the other hand, have undergone slightly higher sequence divergence as compared to ridge gourd DNA.

Use of total repetitive DNA of sponge gourd as tracer

1) Homologous hybridization: The limit repetitive Cot value of sponge gourd DNA is 50.0 M x sec/l. The self-reassociation of tracer DNA was estimated to be 12.23% at this Cot value. Applying this correction, the homologous

hybridization between total repetitive DNA (tracer) and total DNA (driver) is 65.81% (Table 6.1b).

The Tm value of homologous duplexes is 81.07°C (Fig.6.1b, .Table 6.2b). This value is lower than the melting temperature of native sheared sponge gourd DNA by 2.37°C. Thus the homologous duplexes of sponge gourd repetitive DNA reassociating at Cot 50.0 show a mismatch of 2.37%.

2) Heterologous hybridization: The heterologous reassociation of total sheared DNAs of ridge gourd, ash gourd and ivy gourd with repetitive DNA of sponge gourd is 62.16%, 31.78% and 17.12% respectively, after correcting for self-reassociation of the tracer DNA. When these values are compared with the homologous reassociation value (65.81%), the sequence homology is estimated to be 95% between sponge gourd and ridge gourd, 48% between sponge gourd and ash gourd and 26% between sponge gourd and ivy gourd.(Table 6.1b)

The Tms of heteroduplexes of sponge gourd DNA with ridge gourd, ash gourd and ivy gourd are 79.80°C, 73.07°C and 73.53°C respectively. These values correspond to a base mismatch of 1.27%, 8.00% and 7.54% respectively as compared to homologous duplexes. This set of experiments also clearly indicates that repetitive sequences of ridge gourd and sponge gourd are very similar, while those of ash gourd and ivy gourd have diverged to a great extent from those of sponge gourd (Figure 6.1b, Table 6.2b).

Table 6.2a Sequence divergence between total DNA of sponge gourd and repetitive DNAs of Cucurbitaceae

Species	Tm (°C)	Δ Tm[a] (°C)	Base mismatch[b] (%)	Proportion (%)
Cot 50.0				
Sponge gourd	79.30	4.14	0.00	-
Ridge gourd	78.65	0.65	0.65	-
Cot 25.0				
Sponge gourd	78.24	5.20	0.00	
Ash gourd	74.68	3.56	3.56	68.40
	89.44			31.60
Cot 10.0				
Sponge gourd	77.88	5.56	0.00	70.29
Ash gourd	76.86	1.02	1.02	29.71
	91.49			

a : Calculated as lowering in Tm as compared to that of sonicated sponge gourd DNA (83.44°C).

b : Calculated as 1°C lowering in Tm = 1°C base mismatch (14)

Table 6.2b Sequence divergence between isolated repetitive DNA of sponge gourd and repetitive DNAs of Cucurbitaceae.

Species	Tm (°C)	Δ Tm[a] (°C)	Base mismatch[b] (%)	Proportion (%)
Sponge gourd				
Cot 50.0	81.07	2.37	0.00	-
Ridge gourd				
Cot 50.0	79.80	1.27	1.27	-
Ash gourd				
Cot 25.0	73.07	8.00	8.00	60.00
	87.40			40.00
Ivy gourd	73.53	7.54	7.54	60.00
Cot 10.0	87.60			40.0

a : Calculated as lowering in Tm as compared to that of sonicated sponge gourd DNA (83.44°C)

b : Calculated as 1°C lowering in Tm = 1°C base mismatch (14).

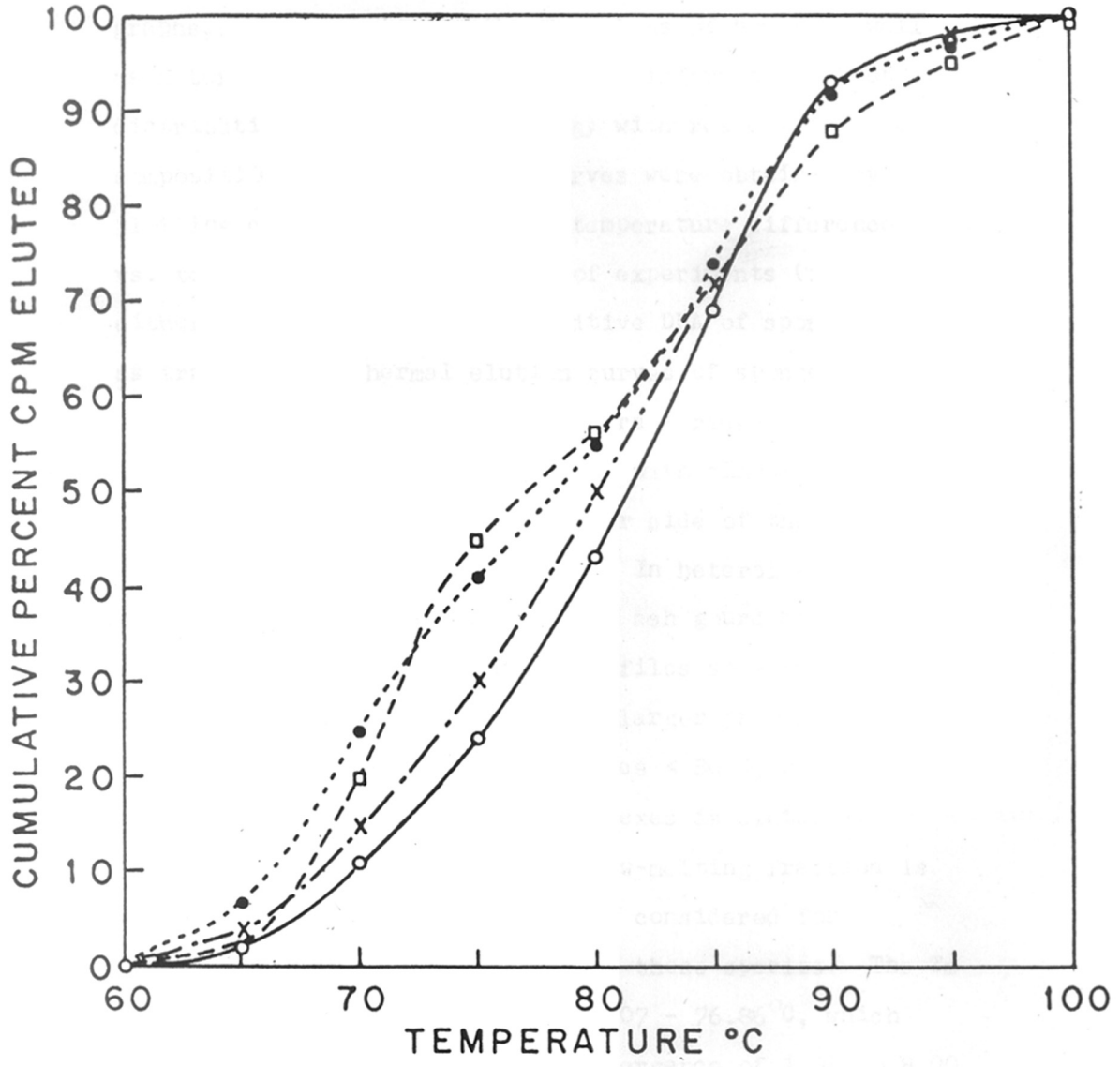


Figure 6.1b : Thermal dissociation profiles of duplexes formed between sponge gourd ^{32}P total repetitive DNA and unlabelled sheared DNAs from sponge gourd (0 — 0), ridge gourd (— x —), ash gourd (--) and ivy gourd (---). Results are obtained as described above.

Derived melting curves: From the cumulative melting graphs, we can arrive at the Tm values of homo- as well as heteroduplexes. In order to get information about distribution of sequence homology with respect to base composition, derived melting curves were obtained by plotting counts eluted at each temperature difference vs. temperature. In both sets of experiments (involving either total DNA or total repetitive DNA of sponge gourd as tracer), the thermal elution curves of sponge gourd homologous duplexes and sponge gourd - ridge gourd heterologous duplexes are uniform, with almost equal amounts of counts eluting on either side of the melting temperature (77.88°C to 81.07°C). In heterologous reactions of sponge gourd DNA with ash gourd and ivy gourd DNA, the thermal elution profiles show two distinct melting regions. A comparatively larger proportion (60 - 70%) is eluted at temperatures < 80°C, while a smaller fraction (30 - 40%) of duplexes is eluted at temperatures > 80°C. Since the low-melting fraction is a major one, its Tm value has been considered for calculating sequence divergence in these species. The Tm of the low-melting fraction is 73.07 - 76.86°C, which corresponds to a base sequence divergence of 1.02 to 8.00% when compared to the thermal stability of homologous duplexes. On the other hand, the small fraction eluting after 80°C

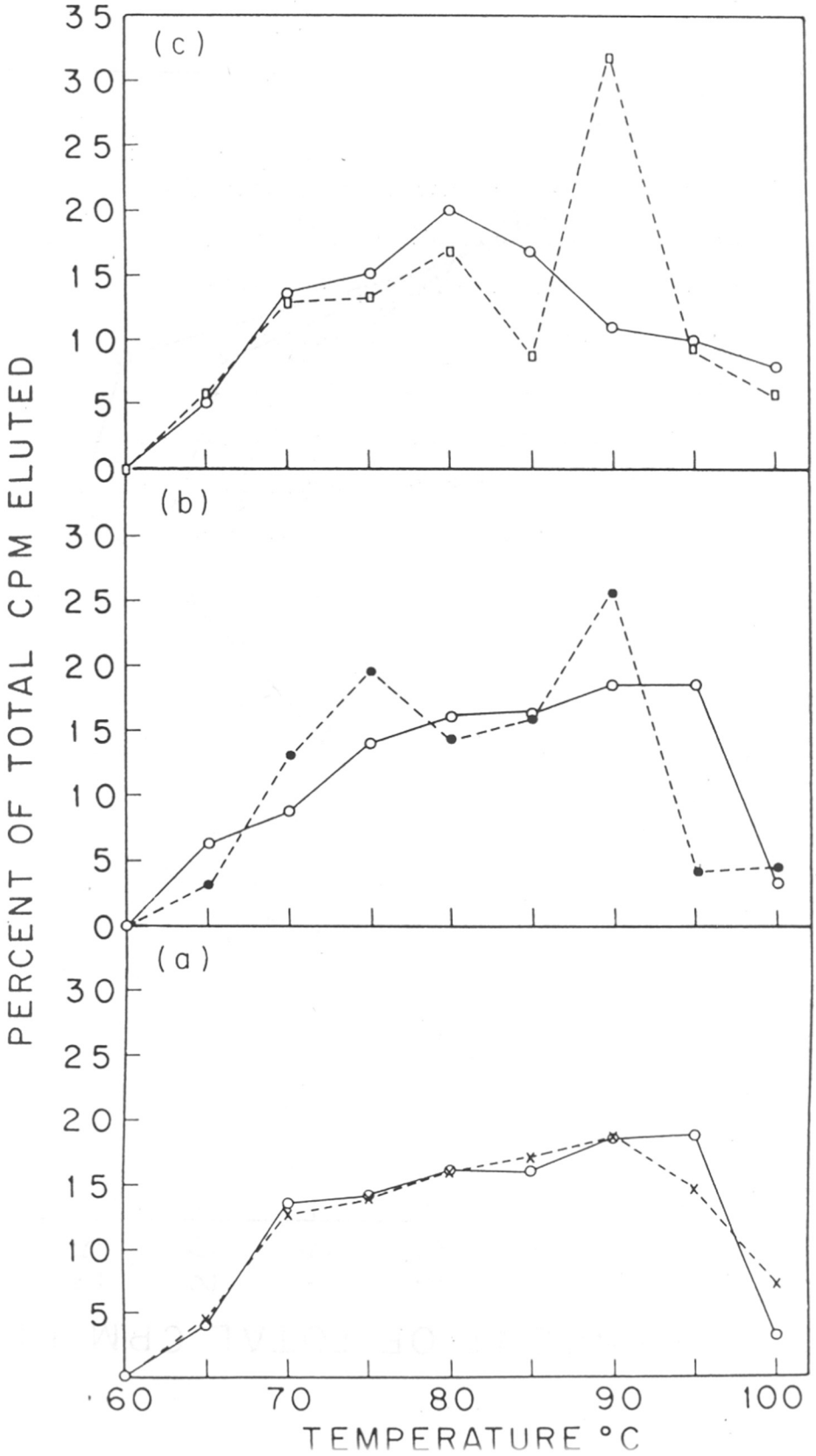
has higher thermal stabilities (Tm = 87.4°C - 91.49°C). It thus appears that the G+C rich repetitive families of these genomes have undergone less sequence divergence than the A+T rich families (Figure 6.2a,b).

DISCUSSION

In sequence homology studies, total DNAs as well as isolated repetitive and single copy fractions are commonly used as tracer DNAs (5,6,9,10,16). Both the approaches have been used in the present work, since it was expected that the results obtained by these independent methods would be complementary. Comparison of results obtained by these two approaches showed that the homology of ridge gourd repetitive DNA to that of sponge gourd increased from 84% to 95% when total repetitive DNA of sponge gourd was used as tracer. This may be due to the use of pure repetitive fraction of sponge gourd DNA and thus the removal of the single copy sequences associated with it. This could also be true for the slight increase in homology observed for ash gourd DNA. However, in ivy gourd DNA, the homology reduced from 43% to 26% when the repetitive DNA of sponge gourd was used as tracer. A plausible explanation for such a change can be that some repetitive sequences of ivy gourd are homologous to single copy sequences of sponge gourd, resulting in increased homology when total DNA of sponge gourd is used as tracer.

Figure 6.2a : Thermal elution profiles of duplexes formed between total sheared ^{32}P sponge gourd DNA and unlabelled sheared DNAs from (a) ridge gourd (X - X), (b) ash gourd (● - - ●) and (c) ivy gourd (□ - - □), plotted with the thermal elution profiles of homologous reassociations carried out simultaneously. The amount of DNA eluted at each temperature is plotted as a percentage of the total duplex DNA bound to the column.

Figure 6.2a


(c)
(b)
(a)
PERCENT OF TOTAL CPM ELUTED
TEMPERATURE °C
0
5
10
15
20
25
30
35
60
70
80
90
100

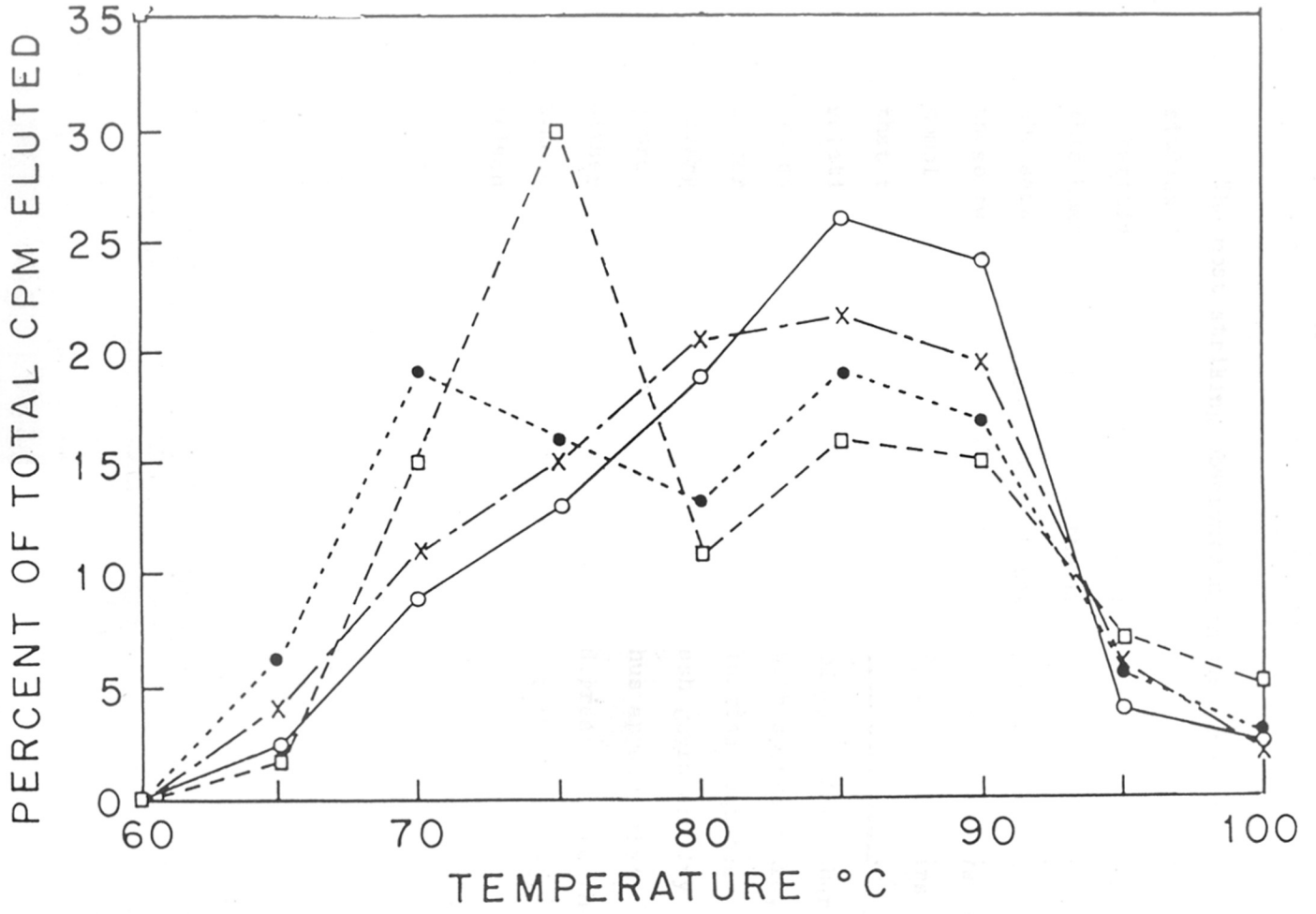


Figure 6.2b : Thermal elution profiles of duplexes formed between sponge gourd ^{32}P total repetitive DNA with unlabelled sheared DNA from sponge gourd (o — o), ridge gourd (-- x --), ash gourd (-----) and ivy gourd (---). The amount of DNA eluted at each temperature is plotted as a percentage of the total duplex DNA bound to the column.

The most striking observation in the present studies is the extensive homology (85-95%) in the repetitive sequences of sponge gourd and ridge gourd. This homology is expected as the two species belong to the same genus Luffa. The sequence divergence between these two species is also limited. From the extensive homology and the limited sequence divergence it appears that these two species have separated from each other relatively recently. The sequence homology of the other two species, ash gourd and ivy gourd, with sponge gourd, is very limited. There is also a substantial sequence divergence between repetitive DNAs of ash gourd and ivy gourd with that of sponge gourd. It thus appears that either these three genera have diverged greatly since their time of origin from a common ancestor, or they have all arisen independently.

REFERENCES

1. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-556.

2. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissman, S.M., Slighton, J.L., Blechel, A.F., Smithies, O., Baralle, F.E. Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.

3. Isenberg, I. (1979) Annu.Rev.Biochem. 48, 159-191.

4. Ruiz-Carrillo, A., Affolter, M. and Renaud, J. (1983) J.Mol.Biol. 170, 843-859.

5. Zwiebel, L.J., Cohn, V.H., Wright, D.R. and Moore, G.P. (1982) J.Mol.Evol. 19, 62-71.

6. Eden, F.C., Hendrick, J.P. and Gottlieb, S.S. (1978) Biochem. 17, 5113-5121.

7. Smith, D.B. and Flavell, R.B. (1974) Biochem.Genet. 12, 243-256.

8. Bendich, A.J. and Bolton, E.T. (1967) Plant Physiol. 42, 959-967.

9. Belford, H.S. and Thompson, W.F. (1981) Heredity 46, 109-122.

10. Stein, D.B., Thompson, W.F. and Belford, H.S. (1979) J.Mol.Evol. 13, 215-252.

11. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J.Mol.Biol.113, 237-251.

12. Berger, S.L. (1984) Anal.Biochem. 136, 515-519.

13. Bendich, A.J. and McCarthy, B.J. (1970) Genetics 65, 545-565.

14. Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) In: Methods in Enzymol. (Grossman, G.L. and Moldave, K., eds.) Vol.29E, Academic Press, New York, pp. 363-418.

15. Flavell, R.B., Rimpau, J. and Smith, D.B. (1977) Chromosoma (Berl.) 63, 205-222.

16. Triantaphyllidis, C.D. and Richardson, R.H. (1982) Genetica 57, 225-229.

CHAPTER VII

GENERAL DISCUSSION

GENERAL DISCUSSION

The central theme of this thesis has been the molecular biological characterization of Cucurbitaceae genomes using five complementary approaches. Although each individual approach is not enough to provide conclusive evidence, the combined results from all the five different approaches have provided a considerable information on the molecular properties of the Cucurbitaceae genomes, which can be summarized as follows:

1. The high resolution thermal denaturation profiles have revealed a uniform distribution of A+T and G+C rich DNA sequences in ash gourd and ivy gourd. In sponge gourd and ridge gourd, on the other hand, the base composition heterogeneity is more in G+C rich DNA sequences than in A+T rich DNA sequences.

2. The content of repetitive DNA is moderate (48-59%) in sponge gourd, ridge gourd and ash gourd and very low (25%) in ivy gourd.

3. The moderately repetitive DNA shows 5-fold variation in its proportion, 3.5 fold variation in its copy number and 6-fold variation in its kinetic complexity.

4. The proportion of very fast DNA is high (12-14%) in ash gourd and ivy gourd, moderate (8%) in ridge gourd

and low (2%) in sponge gourd.

5. The base sequence divergence among repeated DNA sequences is 2-8%.

6. The repeated DNA families are essentially homogeneous in nature in all the four species.

7. The genome organization pattern is mixed type, with identical lengths of interspersed long and short repeats and interspersed single copy sequences. While the genomes of sponge gourd and ridge gourd show extensive interspersion of repetitive sequences with single copy DNA, a large proportion of single copy DNA remains uninterspersed in ash gourd and ivy gourd.

8. The repetitive DNA sequences of sponge gourd and ridge gourd show a high degree of sequence homology (85-95%) to each other and those of ash gourd and ivy gourd show limited homologies (25-50%) to sponge gourd.

All the above results have been discussed at length in appropriate Chapters. However, when they are considered together, it is possible to draw conclusions about the possible role, rates of evolution and sequence homologies of repeated DNA sequences in family Cucurbitaceae.

I. Nucleotide sequence repetition

The genomes of sponge gourd, ridge gourd and ash gourd are highly repetitive, while that of ivy gourd

contains the lowest amount of repeated DNA observed so far in higher plant species. The presence of low amount of repetitive DNA in ivy gourd can be explained in two ways: either the rate of amplification of sequences has been very low, or the repeated sequences have diverged to such an extent that they no longer reassociate with repetitive kinetics.

The proportions of the various subclasses of repetitive DNA are different in all the four genomes. A significant fraction of the total repetitive DNA consists of moderately repeated DNA sequences, with only a few hundred copies. The difference between the genomes of sponge gourd and ridge gourd is the presence of a fast component only in sponge gourd and a large middle repetitive component in ridge gourd. The fast component in sponge gourd could have arisen from either a recent amplification of a single copy sequence or a secondary amplification of a preexisting sequence.

The kinetic complexity, defined as the total length of DNA contained in different nucleotide sequences as measured by reassociation kinetics (1), provides a measure of information content not affected by reiteration frequency. In the present work, the kinetic complexity of moderately repeated sequences appears to be relatively high, and thus these sequences may have potential

information contents. On the other hand, the very fast and fast DNA components seem to have very low kinetic complexities and thus may not have any coding functions.

Apart from furnishing information about the repetitive DNA content, the DNA reassociation kinetics can also provide the genome size estimate. The kinetic estimates of genome sizes are 8.4×10^7 to 1.7×10^8 base pairs in all the four species. The genome sizes of all these four species have also been determined cytophotometrically in our laboratory (2). The 1C values determined are 1.75 pg, 2.25 pg, 1.75 pg, 2.75 pg for sponge gourd, ridge gourd, ash gourd and ivy gourd respectively. Thus the kinetic estimates of genome size are considerably lower than the cytophotometrically determined values. Such discrepancies are often noticed in the genome sizes determined by the two methods (3-6). An explanation for this discrepancy could be the absence of truly single copy sequences. Species are known where unique sequences are present in more than one identical copies (7). Inaccuracies in the determination of rate of reassociation of single copy DNA could also result due to presence of certain factors such as polysaccharides in DNA preparations that can accelerate the rate of reassociation (8,9). We have confirmed the absence of polysaccharides in our preparations. It, therefore, appears that the results of the two methods of

genome size determination are not comparable in our experiments. Since the repetitive DNA constitutes a significant portion of plant genomes, it seems unlikely that all this DNA would function in a manner requiring sequence similarity. The degree of sequence homology required by the cell for a function is not known. Presumably, families composed of closely related members should afford the advantage of more precise control over a postulated network of functions, rather than those composed of more diverged members (10). There are examples where a single base change can alter the recognition of a regulatory protein for its binding sequences (11). However, there is also an example where rather divergent operator sequences are recognized by the repressor protein (12). Thus nearly all the repetitive families would be unsuitable for purposes which require functional repetition. Perhaps only some sequences which reassociate at Tm -5°C would be sufficiently similar to be functional (10). Data from most organisms have been gathered from experiments conducted at Tm -25°C and lead to estimates of the length, spacing, number and proportions of DNA sequences which appear to be repetitious at the standard criterion. However, the functional repeats may exist as a small number of nearly perfect replicas, and their properties may be obscured by studies at standard criterion. Thus it seems essential to distinguish between the <u>biological</u>

criterion for sequence repetition and the experimental criterion. Alternatively, the repetitive DNA may not have any sequence dependent function and its use in a 'filler' capacity may have some advantages for higher organisms.

II. Speculations on genome evolution

Two approaches have been employed in the present work to arrive at conclusions about the rate of turnover in the four cucurbit genomes.

The studies on DNA reassociation kinetics at four different temperatures have revealed that the repetitive families are predominantly homogeneous in nature in all the four genomes. The proportion of DNA scored as 'repetitive' reduces with increase in stringency of reassociation. According to the terminologies coined by Bendich and Anderson (10) and Thompson and Murray (13), the sequences scored as 'single copy' at the standard criterion could have resulted from the base sequence divergence of repeated DNA sequences. When relaxed conditions of reassociation such as low incubation temperature are provided, such diverged sequences are able to reassociate at the repetitive rate. These are termed as 'fossil repeats'. The single copy sequences which do not show repetitive kinetics at reduced stringency are termed 'true single copy' sequences. A high proportion of fossil repeats and a low proportion of true single copy

is indicative of fast turnover rates. Calculations of 'fossil' repeats and 'true' single copy sequences reveal that the genome of sponge gourd is evolving faster than that of ridge gourd, ash gourd and ivy gourd.

The study of interspersion patterns of repeated DNA sequences has demonstrated that all the four cucurbit species are grossly similar with respect to the length of interspersed long and short repeats and interspersed single copy DNA sequences. However, differences are noticed in the extent to which the sequences are interspersed with each other. In sponge gourd and ridge gourd, the interspersion of single copy DNA sequences with repeated sequences is almost total. However, sponge gourd shows a predominance of short period pattern. Ash gourd and ivy gourd exhibit extensive interspersion of repeats within themselves but very little single copy fraction is involved in interspersion. The interspersion of repeats within themselves is generally achieved by mechanisms of unequal crossing over, wherein elements from tandem arrays are translocated to other regions of chromosomes. Translocation could occur as a result of recombination events between different regions of the same or different chromosomes. However, only recombination mechanisms cannot account for interspersion of millions of short repeated DNA sequences. Migration of repeats out of the tandem arrays could be

imagined to involve excision and reinsertion processes similar to those demonstrated in prokaryotes. The mobile elements first, demonstrated in corn, have properties similar to those of transposons and are known to be widely scattered (Chapter I). Plasmids or circular DNAs have also been found in some eukaryotic systems (14,15). Thus the differences in interspersion patterns need not be correlated with the phenotypic differences between them. Instead, the differences may be dependent upon the dynamics of amplification and translocation. High turnover rates would lead to reductions in the average age of amplified sequences and in the fraction of single copy DNA, resulting in short period pattern, while low turnover rates would lead to long period interspersion patterns (15). Based on our data on DNA sequence organization, it appears that the rate of turnover is low in ridge gourd, ash gourd and ivy gourd and high in sponge gourd.

III. Homologies among repeated DNA sequences

Interspecies hybridization experiments using sponge gourd DNA as tracer have revealed that repetitive sequences of sponge gourd and ridge gourd are very closely related. On the other hand, those of ash gourd and ivy gourd have diverged extensively from those of sponge gourd. The differences may be reflecting the amplification or reamplification of different sequences in different lines

subsequent to their evolutionary separation. Even among groups of repeated sequences common to all species, further amplifications could occur. Most families of repeated sequences, once introduced in the ancestral genome, are maintained in all the subsequently diverging species (16). However, even though a given family may be represented in more than one species, the products of intraspecific reassociation of that family are more precisely paired than those formed in interspecies reactions. If it is assumed that repeated sequences evolve simply by accumulating random base substitutions, the divergence observed among members of a family in one species should be the same as that among members of the same family in the different species. Our results have shown that homologous duplexes are thermally more stable than heterologous duplexes. It is thus apparent that different members of the partially homologous families were subjected to secondary amplification events in different genomes after speciation.

IV. Role of repetitive sequences in speciation

There are several possible ways in which populations may differentiate into species. These are generally classified as premating and post-mating reproductive barriers. Premating barriers are well-known in cucurbits. However, the failure of artificial hybridization suggests

that post-mating barriers should also be responsible to some extent for species isolation in cucurbits. There are ways by which one could envisage the role of repeated DNA sequences in either of these fundamental processes. It is unlikely that the continuously changing nature of sequence composition and abundance and position of repetitive families could be of no effect on the behaviour of chromosomes and their transcription (17). Such effects may not be the function of repeat families, rather they could be their consequence. Many families are finely interspersed with genes and concerted changes in their composition might affect the transcription of genes, possibly by their effects on chromatin accessibility to polymerases. Some such changes might affect the genes controlling sexual behaviour or the time of flowering, and premating barriers could ensue. The incompatibility may also be a post-mating dysgenesis between individuals from populations in which the genomes have evolved in separate ways. Whatever the nature of barriers, the incompatibility can be envisaged as taking effect only between two individuals that are maximally dissimilar in their repeat families (17). The phenomenon of molecular drive does not appear to be restricted to particular organisms with particular population structures, and would not negate the alternate modes of speciation.

REFERENCES

1. Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) In: Methods in Enzymol. (Grossman, G.L. and Moldave, K., eds) Vol. 29E, Academic Press, New York, pp. 363-418.

2. Joshi, C.P. (1982) Cytological and molecular biological studies in plant genomes. Ph.D. Thesis, University of Poona, Poona 411008, India.

3. Zimmerman, J.L. and Goldberg, R.B. (1977) Chromosoma (Berl.) 59, 227-252.

4. Walbot, V. and Dure, L.S. (1976) J.Mol.Biol. 101, 503-536.

5. Wimpee, C.F. and Rawson, J.R.Y. (1979) Biochim.Biophys. Acta. 562, 192-206.

6. Pearson, W.R., Wu, J.R. and Bonner, J. (1978) Biochemistry 17, 51-59.

7. Gnucheva, E. Yu., Kir'yanov, G.I., Polyakov, V. Yu. and Chentsov, Yu. S. (1977) Biokhimya 42, 2186-2191.

8. Kemp, J.D. and Merlo, D.J. (1975) Biochem.Biophys. Res. Commun. 67, 1522-1526.

9. Merlo, D.J. and Kemp, J.D. (1976) Plant Physiol. 58, 100-106.

10. Bendich, A.J. and Anderson, R.S. (1977) Biochemistry 16, 4655-4663.

11. Dickson, R.C., Abelson, J., Johnson, P., Resnikoff, W.S. and Barnes, W.M. (1977) J.Mol.Biol. 111, 65-75.

12. Ptashne, M., Backman, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B. and Sauer, R.T. (1976) Science 194, 156-161.

13. Thompson, W.F. and Murray, M.G. (1980) In: Davies, D.R. and Hopwood, D.A. (eds.) Proceedings of the 4th John Innes Symposium, The Plant Genome pp. 31-45.

14. Stanfield, S. and Helinski, D.R. (1976) Cell 9, 333-345.

15. Cameron, J.R., Philippsen, P. and Davis, R.W. (1977) Nucleic Acids Res. 4, 1159-1181.

16. Flavell, R.B., Rimpau, J. and Smith, D.B. (1977) Chromosoma (Berl.) 63, 205-222.

17. Dover, G. (1982) In: Mechanisms of Speciation, Alan R. Liss Inc., New York, pp. 435-459.

LIST OF PUBLICATIONS

1. Molecular analysis of Cucurbitaceae genomes. I. Comparison of DNA reassociation kinetics in six plant species.
   Mrinal Bhave, Meena Lagu and P.K. Ranjekar
   Plant Sci. Lett. 33 (1984) 127-136.

2. Molecular analysis of Cucurbitaceae genomes: II. Comparison of high resolution thermal denaturation profiles of the DNAs in seven plant species.
   Mrinal Bhave, Meena Lagu, S.R. Gadre and P.K.Ranjekar
   Indian J.Biochem.Biophys. 21 (1984) 81-84.

3. Molecular analysis of Cucurbitaceae genomes. III. Arrangement of repetitive and single copy DNA sequences in four plant species.
   Mrinal Bhave and P.K. Ranjekar (Communicated).

4. Molecular analysis of Cucurbitaceae genomes. IV. Occurrence of homogeneous repeat families evolving with different rates.
   Mrinal Bhave and P.K. Ranjekar (Communicated).

5. Molecular analysis of Cucurbitaceae genomes. V. Homology of repeated DNA sequences in seven plant species.
   Meena Lagu, Mrinal Bhave, V.S. Gupta and P.K. Ranjekar. (Communicated).

PAPERS PRESENTED AT SYMPOSIA

Genome analysis of three Cucurbitaceae species.
Mrinal Manohar and P.K. Ranjekar.
Presented at the 50th Annual Meeting of Society of Biological Chemists, India (1981).

Characterization of sequence complexity and DNA sequence organization in five Cucurbitaceae genomes.
Mrinal Bhave, Meena Lagu and P.K. Ranjekar.
Presented at the 51st Annual Meeting of Society of Biological Chemists, India (1982).

Restriction enzyme analysis of Cucurbitaceae plant DNAs.
Mrinal Bhave, Meena Lagu, D.N. Deobagkar and P.K. Ranjekar.
Presented at the 52nd Annual Meeting of Society of Biological Chemists, India (1983).

Molecular analysis of Cucurbitaceae genomes: Denaturation and reassociation properties of repeated DNA sequences in four plant species.
Mrinal Bhave and P.K. Ranjekar
Presented at the XV International Congress of Genetics, New Delhi, India (1983).

Molecular analysis of Cucurbitaceae genomes: Comparison of DNA reassociation kinetics in six plant species.
Meena Lagu, Mrinal Bhave and P.K. Ranjekar
Presented at the XV International Congress of Genetics, New Delhi, India (1983).

Organization of plant nuclear genomes
Vidya Gupta, S. Lakshmi, M. Bhave, S. Patankar, S.A. Ranade and P.K. Ranjekar
Presented at the Third International Congress on Cell Biology, Tokyo, Japan (1984).

ACKNOWLEDGEMENTS

I am extremely thankful to my research guide, Dr. P.K. Ranjekar, for his valuable guidance and advice throughout the course of this work.

I record my deep sense of gratitude to Dr. L.K. Doraiswamy, Director, National Chemical Laboratory, for providing all the laboratory facilities. My sincere thanks go to Dr. V. Jagannathan, Dr. C. SivaRaman and Dr. M.C. Srinivasan for the instructive discussions and the encouragement extended from time to time.

I gratefully acknowledge Dr. Mrs. Vidya Gupta for teaching me the various experiments in the early stages of my work and for her critical comments and timely suggestions. I am also grateful to Dr. S.R. Gadre, Department of Chemistry, University of Poona, for developing the various computer programs which have been most useful in a number of analyses. I thank Dr. V. Shankar for taking a lot of interest in our work and for going through the thesis critically.

I wish to thank my friends Shekhar, Lakshmi, Shirish, Meena, Shubhada and Mr. Hendre for extending all the help that I required during the work and for maintaining a lively atmosphere in the laboratory. Their humour, and, at times, criticism has helped me a great deal.

I am indebted to all the members of the staff of Division of Biochemistry for their kind co-operation. Mr. S.T. Dhume, Mr. M.N. Kamthe and Mr. P.V. Pendse's technical assistance is gratefully acknowledged. I am also thankful to Mr. T. Koshi for his flawless typing.

The financial assistance in the form of a Research Fellowship from National Council of Educational Research and Training, New Delhi is gratefully acknowledged.

October 1984

MRBhave
Mrs. Mrinal Bhave