May 1989

Variants and Polymorphisms of Three Repetitive DNA Families in the Human Genome

Robert M. Roudabush
East Tennessee State University

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Order Number 8917784

Variants and polymorphisms of three repetitive DNA families in the human genome

Roudabush, Robert Mackey, III, Ph.D.

East Tennessee State University, 1989

U·M·I

VARIANTS AND POLYMORPHISMS OF THREE REPETITIVE

DNA FAMILIES IN THE HUMAN GENOME

———————

A Dissertation

Presented to

the faculty of the Department of Biochemistry

Quillen-Dishner College of Medicine

East Tennessee State University

———————

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in Biomedical Sciences

———————

by

Robert M. Roudabush

May 1989

APPROVAL

This is to certify that the Graduate Committee of

<u>ROBERT M. ROUDABUSH</u>

met on the

<u>twenty sixth</u> day of <u>May</u>, 1988.

The committee read and examined his dissertation, supervised his defense of it in an oral examination, and decided to recommend that his study be submitted to the Associate Vice-President for Research and Dean of the Graduate School, in partial fulfillment of the requirements for the degree Doctor of Philosophy in Biomedical Science.

_____
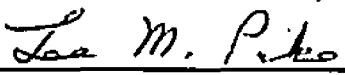Chair, Graduate Committee

_____

_____

_____

_____

Signed on behalf of
the Graduate Council

_____
Associate Vice-President for
Research and Dean of the Graduate
School

ii

VARIANTS AND POLYMORPHISMS OF THREE

REPETITIVE DNA FAMILIES IN THE HUMAN GENOME

ABSTRACT

VARIANTS AND POLYMORPHISMS OF THREE REPETITIVE DNA

FAMILIES IN THE HUMAN GENOME

by

Robert M. Roudabush

A novel 0.6 kb LINE family in human DNA, designated L2Hs, has been described (Musich and Dykes 1986). Studies employing clone N6.4, containing three 0.6 kb segments of this family, indicate that these sequences are interspersed and moderately repetitive. Two additional variant sequences of the L2Hs family, N6.1 and N6.3, have been identified. Restriction mapping of each cloned segment indicates similarities among N6.4, N6.3 and N6.1. When the cloned DNAs were cleaved with restriction enzymes and subjected to cross-hybridization, each cloned insert produced a pattern indicating that the sequences contained in N6.1 and N6.3 are represented in at least one of the three 0.6 kb segments within the clone N6.4. Hybridization of human genomic DNA digested with KpnI or KpnI+AccI reveals differences in nuclear organization for these segments. For any particular human DNA, the hybridization patterns for each of the three probes overlap. However, these differences indicate that the inserts in N6.1 and N6.3 and one of the N6.4 inserts each represents a subset of the L2Hs LINE family.

Sequence analysis of N6.1 indicates that the probability of a functional translation product from a N6.1 transcript is not high. The sequence contains stop and non-sense codons in all reading frames. However, the DNA has properties suggesting a structural, non-coding role. The N6.1 sequence contains 11 regions of alternating purine and pyrimidines which can affect the three dimensional structure and, therefore, the structural behavior of the molecule. In addition, putative binding regions for microtubule-associated proteins have been identified.

A cloned variant of the XbaI family of repetitive DNAs, PuHu7, was identified. Studies of its genomic organization showed a tandem arrangement similar to other, previously described members of this family.

The genomic organization of a previously undescribed repetitive DNA family is also reported. This family descriptor is the clone PuHu26. Hybridization of genomic DNA digested with HindIII showed that sequences homologous to PuHu26 are tandemly organized. Genomic DNA cleaved with EcoRI revealed that a subpopulation of the PuHu26 family contains EcoRI restriction sites spaced at multiples of approximately 172 bp.

DEDICATION

This work is dedicated to Karen, Jessie and Joey, without whom this work would not have been completed.

ACKNOWLEDGMENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AGM	African green monkey
BME	beta-mercaptoethanol
bp	base pair(s)
BSA	bovine serum albumin
CIA	Chloroform-isoamyl alcohol
CIAP	calf intestine alkaline phosphatase
DAPI	4',6-diamino-2-phenylindole
dH_2O	distilled water
Me_2SO_4	dimethylsulfoxide
DNase	deoxyribonuclease
DTT	dithiothreitol
EBr	ethidium bromide
E. coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EELN	electroelution
HnRNA	heterogeneous nuclear RNA
Hu	human
IRS	internal repeat sequence
kb	kilobase pair
LB	Luria-Bertani broth
MAP	microtubule-associated protein
NTE	sodium chloride-Tris-EDTA solution
OD	optical density
OLB	oligonucleotide labeling buffer
ORF	open reading frame

PEG	polyethylene glycol
PBL	peripheral blood lymphocyte
PBS	phosphate-buffered saline
RFLP	restriction fragment length polymorphism
RNase	ribonuclease
SDS	sodium dodecyl sulfate
SOG	sucrose orange G
SSC	standard sodium chloride-sodium citrate
STC	sucrose-Tris-citrate buffer
TAE	Tris-acetate-EDTA
TBE	Tris-boric acid-EDTA
TEMED	N',N',N',N'-tetramethylethylenediamine
tRNA	transfer RNA
Tris	tris(hydroxymethyl)aminomethane
$T_{10}E_1$	Tris-EDTA solution
$T_1E_{0.1}$	Tris-EDTA solution, 1/10 dilution
UV	ultraviolet
VNTR	variable number tandem repeats

CHAPTER 1

Introduction

The human genome contains a large number of repetitive elements. Some of these are genes which are found in multiple copies; examples include the genes for the ribosomal RNAs and the histones. These repetitive genes do not account for the majority of repeated DNA sequences.

Systematic studies have identified two groups of reiterated sequences which are distinguished by their organization in the genome. These are the tandemly repetitive and the interspersed repetitive DNAs. Within these two broad groups is a variety of sequences which display homology and similar copy number and, therefore, are considered to be members of the same family.

The investigations reported here have identified two novel DNA families and variant sequences from other previously described repetitive families. Cloned DNA segments of the two new families have been characterized and represent a tandemly and an interspersed repetitive DNA. The clone PuHu-26 is the descriptor of the tandemly repetitive family. The interspersed family is represented by the clones N6.4.39, N6.1 and N6.3.

The N6 series of clones are variants of each other and describe the LINE 2 (L2Hs) family of human interspersed DNAs. These cloned representatives have revealed several

2

interesting genomic characteristics of the L2Hs family, including quantitative and qualitative polymorphisms.

Characterization of the L2Hs, PuHu26 and other repetitive families have elucidated potential functions for repetitive DNAs. Interspersed repetitive DNAs are believed to assist in the regulation of transcription, while some tandemly repetitive DNAs are thought to provide a specific centromeric organization necessary for chromosome alignment during meiosis and mitosis. Answers to some of the following questions may provide additional insights to the functions of these repeated sequences.

1) How do the genomic organization of the L2Hs and PuHu-26 sequence families compare to the genomic organization of other repetitive families?

2) Do hybridizations with each of the L2Hs variants identify different subpopulations within the same genome?

3) Does the L2Hs polymorphic pattern vary between individuals?

4) Is there an internal sequence periodicity in the L2Hs repeats?

5) Does the N6.1 sequence have a segmented structure indicating a potential regulatory function?

Before answering these questions, a brief review of the characteristics of tandemly repetitive and interspersed repetitive DNAs is necessary. In describing these characteristics it will become evident that the human genome

is very dynamic and exhibits a variety of DNA organizations some of which are only now being identified.

Tandemly Repeated DNAs

Several types of tandemly repetitive DNAs have been isolated from the human genome and characterized. These are: a) tandemly repetitive DNAs isolated by cesium salt centrifugation, b) tandemly repetitive DNAs identified by gel electrophoresis of limit digests of genomic DNA with restriction enzymes and c) short, tandemly-arranged, simple sequences which flank genes.

Repetitive DNAs isolated by cesium salt centrifugations and those identified by restriction enzyme digests share several general traits (for review see Beridze 1986). Sequences of each are represented 10^5 to 10^6 times per haploid genome. As a result, these sequences are highly repetitive. They consist of clustered repeat units, minimally interspersed with other sequences, and arranged in long, tandem arrays. Highly repetitive DNAs are associated with heterochromatin at the telomeres and centromeres of chromosomes.

Highly repetitive DNAs isolated by cesium salt centrifugation are called satellite DNAs. This term reflects the different buoyant densities of satellite DNAs and the main component DNAs during CsCl centrifugation. Four human satellites, I thru IV, have been isolated (Corneo et al. 1967; 1970). Each is composed of a simple core

sequence of 5 to 150 base pairs (bp). The core sequences represent a major component of their respective arrays (Prosser et al. 1986), but are not the only sequence in the array. Mitchell et al. (1979) have shown homologies between these satellites. Components of satellites III and IV are homologous. A major component of satellite III is homologous to minor constituents of satellites I and II, although no homology between satellites I and II has been observed. In situ hybridization also displayed a complex chromosomal relationship among these satellites. Gosden et al. (1975) reported hybridization to the same chromosome for each of the four satellites. Further studies (Cook and Hindley 1979; Deninger et al. 1981) showed that satellite III contains two distinct minor sequence subsets: One is a Y chromosome-specific sequence and the other is a chromosome 1-specific sequence.

The second group of human highly repetitive sequences were identified with restriction endonucleases. These families are often named for the restriction endonuclease used to characterize them. The first of these human sequences to be isolated and sequenced was the EcoRI dimer (Manuelidis 1976; 1978; Manuelidis and Wu 1978), followed by the XbaI tetramer (Gray et al. 1985) and, more recently, by the Sau3A pentamer (Kiyama et al. 1986). Analysis of these three repetitive families has identified several common features: a) The repeating unit of each of these families

is more complex than the repeating unit of the human cesium salt satellite DNAs. There is no internal periodicity, although there are short sequences of alternating purines and pyrimidines. b) The unit repeat for each of the three families is a multiple of approximately 172 base pairs (bp). The organization of the basic repeat unit is different for each of the families. The 849 bp Sau3A pentamer repeat is composed of five segments of 171, 171, 167, 169, and 171 bp. The 682 bp unit of the XbaI tetramer family consists of 171, 170, 171, and 170 bp repeats, while the EcoRI dimer family is composed of one 172 bp and one 169 bp segment. c) A 70% sequence homology exists among the families. d) Within each family a high degree of sequence polymorphism has been noted (Furlong et al. 1986; Kiyama et al. 1986; Gray et al. 1985). e) In situ hybridizations have localized the EcoRI and the XbaI families to the centromeres of several chromosomes. One can speculate that because of their similarities to the XbaI and EcoRI families, the Sau3A sequences may be centromerically located also. Recently, studies employing rodent-human hybrid cell lines have identified a human chromosome-specific organization of the EcoRI dimer sequences (Willard 1985; Willard and Wayne 1987; Jorgensen et al. 1986). Future studies may identify chromosome-specific organizations for the XbaI and Sau3A families as well.

Other repetitive sequences may be closely associated with some genes. These sequences are 15 to 200 bp long and are positioned within introns or near the termini of a gene. Two forms of sequence organization have been identified: 1) Regions of alternating purine and pyrimidine dinucleotides (Higgs et al. 1986; Treco and Arnheim 1986) and 2) a repeated core sequence (Jefferys et al. 1985). Both forms have been implicated in recombinational events resulting in sequence polymorphisms (Bullock et al. 1986; Jefferys et al. 1985). Many of these polymorphisms result from single base changes, small deletions or insertions which create or destroy a particular restriction endonuclease site and give rise to restriction fragment length polymorphisms (RFLPs). In addition to RFLPs, there is another class of polymorphisms. This class is a result of differences in copy number or a variable number of a tandem repeat sequences (VNTR). The acronym VNTR (Nakamura et al. 1987) has been coined to describe this class of polymorphisms, first identified by Wyman and White (1980). Subsequently, other similar regions have been described near the beta and alpha globin gene complex (Antonarkis et al. 1982; Higgs et al. 1986), the insulin gene (Rotwein et al. 1986) and the c-Ha-Ras-1 protooncogene (Goldfurb et al. 1982). VNTRs have been shown to segregate in a Mendelian fashion as alleles of a single locus (Jefferys et al. 1985; Vassart et al. 1987). As the number of alleles identified

is high, it is probable that an individual will be heterozygous at several loci. A high heterozygosity index is important for defining a marker's usefulness in linkage studies. Because VNTRs have a high index they are being considered for genetic markers to establish linkage maps of inherited diseases (Botstein et al. 1980; O'Connell et al. 1987).

Interspersed DNAs

The interspersed DNAs have been classified into two groups, the LINES and the SINES (Singer 1982a, 1982b). The sequences which represent each of these groups are called families. Members of each family are homologous, but not identical and sometimes share restriction site periodicities. The SINES represent families of short interspersed repeat sequences with repeats of less than 600 bp and the LINES represent repeats with lengths over 1 kilobase (kb).

The SINES can be classified into discrete families each with a different number of individual members (Jelinek and Schmid 1982). The Alu family represents the dominant class of human SINES. The approximate length of an Alu family member is 300 bp, consisting of two 130 bp direct repeats, one of which contains an additional 31 bp sequence.

The KpnI family of DNAs represents the human LINE sequences. Four elements comprise the basic family. The lengths of these elements are 1.2, 1.5, 1.8, and 1.9 kb.

The individual KpnI sequences may be clustered to form longer sequences. Hattori et al. (1985) isolated a 6 kb KpnI family member which shows a restriction map similar to a consensus map of the KpnI family (Manuelidis and Biro 1982; Manuelidis 1984; Sakaki et al. 1983). The dominant human LINE family is the 6 kb sequence and is abbreviated LlHs for LINE 1 Homo sapiens (for nomenclature see Singer and Skowronski 1985).

A definitive cellular function for the SINES and LINES has not been demonstrated. There exists a body of circumstantial evidence to support two possible roles. One potential role is as a regulatory sequence in transcription. Both Alu and KpnI sequences have been found in genes (Ullu 1980; Singer and Skowronski 1985) and Alu sequences have been identified within introns and regions flanking genes (Ullu 1980). Alu and KpnI homologous sequences have been identified in heterogeneous nuclear RNA (HnRNA) and poly A$^+$ RNA (Calabretta et al. 1981; Calabretta et al. 1982; Kole et al. 1983). Presumably, they are cleaved out in the formation of functional mRNA molecules but they occur in some cytoplasmic poly A$^+$ RNA molecules (Skowronski and Singer 1985).

An alternative function for Alu and KpnI DNAs is as mobile genetic elements (Singer and Skowronski 1985). Both families have sequences with direct repeats of 5 to 20 bp flanking their termini. Similar repeats have been

identified in transposons (Shapiro 1979). If these DNAs are mobile elements, it would explain their abundance and interspersion in the human genome.

To gain an understanding of the possible roles of interspersed and tandemly repeated DNAs in the human genome, it is necessary to isolate and characterize representatives of these families. The specific aim of this research is to: 1) Clone representatives of tandemly repetitive DNAs, 2) characterize the organization of these tandemly repeated DNAs, 3) determine the organization of a variant of the interspersed L2Hs family by restriction enzyme cleavage analysis, 4) sequence one of the L2Hs variants and 5) identify possible biological roles for the L2Hs variant.

CHAPTER 2

Materials and Methods

Materials

N',N',N',N',-tetramethylethylenediamine (TEMED),
trishydroxymethylaminomethane (Tris), ampicillin,
chloramphenicol, dithiothreitol (DTT), 2-mercaptoethanol
(BME), deoxyribonuclease I (DNase I), ribonuclease (RNase),
4',6-diamino-2-phenylindole (DAPI), ethidium bromide (EBr)
and lysozyme were purchased from Sigma Chemical Company.
Hexanucleotide extension primers and deoxynucleotide
triphosphates (adenine, cytosine, thymine, and guanine) were
purchased from Pharmacia. Ethylenediaminetetraacetic acid
(EDTA), formamide, n-butanol, dimethylsulfoxide (Me_2SO_4) and
Triton X-100 were purchased from Fisher Scientific. Namalwa
DNA (Pritchett et al. 1976), A B-cell lymphoma, was a
generous gift of Dr. McDaniel. Fetal calf serum and Eagles
minimal essential medium were purchased from Flow
Laboratories. Phenol (loose crystals) was purchased from
Mallinckrodt. Restriction enzymes were purchased from
Bethesda Research Labs, Boehringer Mannheim Biochemicals,
Amersham or New England Biolabs. Agarose was purchased from
the Marine Colloids Division of the FMC Corporation.
Pronase and glycogen were purchased from Calbiochem.
Proteinase K was purchased from E.M. Biochemicals. DNA
polymerase I and calf intestinal alkaline phosphatase were

obtained from Boehringer Mannheim. T4 polynucleotide kinase and T4 ligase were purchased from Bethesda Research Labs. Polyethelene glycol 6000 (PEG), XOMAT X-ray film, GBX developer and Rapid Fix were purchased from Eastman Kodak. Gene Screen membrane was purchased from New England Nuclear. Zeta Probe membrane, acrylamide and N',N'-methylene-bis-acrylamide (BIS) were obtained from BioRad Laboratories. Nitrocellulose BA85 membrane was purchased from Schleicher and Schuell. Deoxyadenosine [alpha-^{32}P] triphosphate (635 Ci/mmole and >2000 Ci/mmole), deoxyadenosine [alpha-^{35}S] triphosphate (600 Ci/mmole), and dideoxy sequencing primers were purchased from Amersham. Adenosine [gamma-^{32}P] triphosphate (3000 Ci/mmole) was purchased from ICN Chemicals. Bactoagar, bactotryptone and yeast extract were purchased from Difco Laboratories.

Buffers and Solutions

$T_{10}E_1$ is 10 mM Tris-HCl, pH 7.4 or 8.4, with 1 mM EDTA. $T_1E_{0.1}$ is 1 mM Tris-HCl pH 7.4 or 8.4, with 0.1 mM EDTA. TBE buffer is 50 mM Tris, 50 mM boric acid, 1.25 mM EDTA pH 8.3. TAE buffer is 50 mM Tris, 40 mM acetic acid, 1 mM EDTA pH 8.4. NTE buffer is 1 M NaCl, 20 mM Tris-HCl pH 7.4, 20 mM EDTA. Phosphate-buffered saline (PBS) is 0.14 M NaCl, 50 mM sodium phosphate pH 7.4. Sucrose-Tris-calcium buffer (STC) is 0.25% sucrose, 50 mM Tris-HCl pH 7.4 and 2 mM $CaCl_2$.

SOG loading buffer is 60% (w/v) sucrose, 0.1% (w/v) Orange G. SOXB loading buffer is 60% (w/v) sucrose, 0.1% (w/v) orange G, 0.04% (w/v) xylene cyanol and 0.1% (w/v) bromophenol blue. Sequence gel loading buffer is 80% (v/v) deionized formamide, 10 mM NaOH, 1 mM EDTA, 0.1% (w/v) xylene cyanol, and 0.1% (w/v) bromophenol blue.

SSC (1 X) buffer is 0.15 M NaCl, 15 mM sodium citrate pH 7. DIGE buffer is composed of 0.03% (w/v) bromophenol blue, 100 mM EDTA pH 8.3, 30% glycerol and 0.66% (w/v) sodium dodecyl sulfate (SDS). Formamide hybridization solution is 3 X SSC, 10 mM Tris-HCl (pH 7.4), 0.5% (w/v) dried non-fat skim milk (Johnson 1984) and 50% formamide. Denhardt's solution contains 0.02% (w/v) BSA, 0.02% (w/v) Ficoll 400, and 0.02% (w/v) polyvinylpyrrolidine 360.

Kinasing buffer is 40 mM Tris-HCl pH 7.6, 20 mM $MgCl_2$, 10 mM DTT, 0.2 mM spermidine, 100 μg/ml BSA and 0.2 mM EDTA. Nick translation buffer is 50 mM Tris-HCl pH 7.4, 10 mM $MgSO_4$, 100 μg/ml BSA, 5 mM DTT or 7 mM BME and 13.3 mM each of dGTP, dCTP and dTTP. Chase solution is a mix containing 250 mM each of dGTP, dATP, dCTP and dTTP. A 5 X stock solution of oligonucleotide labeling buffer (OLB) is 250 mM Tris-HCl pH 8.3, 25 mM $MgCl_2$, 0.6 M HEPES buffer pH 6.6 and 8 mM of guanine, thymine and cytosine deoxynucleotide triphosphates. Wittig buffer is 0.1 mM $ZnCl_2$, 0.1 mM EDTA and 10 mM Tris-HCl pH 8.0.

Ligation buffer contains 50 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 10 mM DTT, 0.5 mM ATP, 0.1 mg/ml BSA. Lysozyme solution for mini-preps contains 5 mg/ml lysozyme in 25 mM Tris-HCl pH 8 and 50 mM glucose. CIA is chloroform/isoamyl alcohol (24:1 v/v). Phenol crystals were melted at 65°C and 8-hydroxyquinoline added to 0.1% (w/v); the solution was then equilibrated with $T_{10}E_1$. Pancreatic RNase was prepared at a concentration of 20,000 units/ml in 10 mM Tris-HCl pH 7.5, 15 mM NaCl, heated to 100°C for 15 minutes to inactivate contaminating DNases and allowed to cool to room temperature. Pronase was prepared at 20 mg/ml in water, brought to a pH 5.0 with HCl and incubated at 80°C for 10 minutes. The pH was adjusted to 7.0 with NaOH, NaCl added to 1 M and the pronase allowed to autodigest at 37°C for 20 minutes to destroy any contaminating nuclease activity. DNase I was prepared at a concentration of 1.25 mg/ml in 50 mM Tris-HCl (pH 7.4), 10 mM magnesium acetate, 1 mM DTT, 50 μg/ml BSA and 50% (v/v) glycerol.

A liter of minimal medium was prepared by mixing the following sterile components: 100 ml of 10X M9 salts (60 g Na_2HPO_4, 30 g KH_2HPO_4, 50 g NaCl, 100 g NH_4Cl per liter), 10 ml of 20% glucose, 0.1 ml of a 1% thiamine solution, 1.0 ml of 0.1 M $CaCl_2$, 2.0 ml of 1 M $MgSO_4$, and dH_2O to one liter. Luria-Bertani broth (LB) medium contained 10 g bactotryptone, 5 g bacto-yeast extract and 5 g NaCl per liter. LB medium was sterilized by autoclaving. Filter-

sterilized ampicillin, added to a concentration of 50-100 μg/ml, was used to select for cells containing plasmids having the ampicillin resistance gene.

Bulk Isolation of Genomic DNA

Lymphocytes, frozen in Eagles minimal medium - 10% (v/v) fetal calf serum - 5% (v/v) Me_2SO_4 at a concentration of 10^8 to 10^9 cells/ml, were quick thawed in a 37°C water bath. The cells were pelleted by centrifugation at 900 x g for 5 minutes at 4°C. The supernatant was discarded and replaced with an equal volume of ice-cold PBS. The cells were resuspended by repeated pipetting and recentrifuged. The supernatant was discarded and replaced with 5 ml of ice-cold STC buffer. The solution was pipetted gently several times to resuspend the cells and transferred to a Dounce homogenizer. Triton X-100 was added to 0.5% and the solution was placed in an ice bath for 10 minutes. The cells were ruptured with several strokes of a tight-fitting pestle releasing the nuclei. The quality and the quantity of nuclei released was determined by phase contrast microscopy. Nuclei were collected by centifugation at 5,000 x g at 4°C, then resuspended in 70 mM Tris-HCl pH 7.4. Lysis was achieved by the sequential addition of the following solution components (final concentration listed within parentheses): EDTA pH 8.0 (0.01 M), pronase (0.4 mg/ml), NaCl (1 M), and SDS (1%). Between the addition of each component the lysis solution was gently swirled to

prevent the formation of an insoluble DNA mass. The proteins were digested with pronase by incubating the solution at 65°C for 2 hours with gentle swirling at 20 minute intervals. It was then extracted with an equal volume of phenol equilibrated with 1 M NaCl. The separation of phenolic and aqueous phases was achieved by centrifugation at 12,000 x g at room temperature for 10 minutes. The aqueous phase was collected with a wide bore pipet and extracted with an equal volume of CIA followed by centrifugation at 7800 x g for 5 minutes at room temperature. The aqueous phase was removed and the DNA was collected by one of two methods. The aqueous phase was overlayed with 2.5 volumes of -20°C 95% ethanol; the DNA was spooled from the ethanol onto a glass rod, rinsed in 95% ethanol, air dried for 10 minutes and resuspended in $T_1E_{0.1}$. Alternatively, an equal volume of room temperature 100% isopropanol was mixed with the aqueous phase by repeated tube inversion; the DNA precipitate was removed from solution with a hooked glass rod and resuspended in $T_1E_{0.1}$.

Bulk preparations of high molecular weight genomic DNA were further purified by the addition of pancreatic RNase to 100 units/ml and incubation at 37°C for 30 minutes. The genomic DNA was dialyzed against two changes of 200 volumes of $T_{10}E_1$ (pH 7.4). Concentration of the DNA was accomplished by alcohol precipitation and resuspension in a

smaller volume or by extraction with two volumes of n-butanol.

Digestion of DNA with Restriction Enzymes

Table 1 contains a list of restriction enzymes, their buffer components and their recognition-restriction sequences. DNA was digested to completion by dilution into the appropriate enzyme buffer, adding 1 - 10 units of restriction enzyme per μg of DNA, and incubating the reaction mix for 4 - 16 hours at 37°C. BSA (100 μg/ml) and BME (6 mM) were added to all reactions.

Preparative Isolation of DNA Segments

DNA restriction fragments were subjected to electrophoresis on appropriate agarose or polyacrylamide gels. After electrophoresis, one of the following three methods were used to extract the desired DNA fragments from the gel.

Elution onto Anion Exchange Paper. A slit was cut in the gel in front of the DNA band to be isolated. DE-81 paper (Whatman DEAE cellulose) was pretreated in 50 mM TBE and inserted into the slot. The DNA was transferred by electrophoresis onto the paper at 100 volts (Banner 1982). The paper with bound DNA was placed into a 0.5 ml microfuge tube whose bottom had been previously punctured with a 23 gauge needle and fitted with a glass wool plug. The DE-81 strip was washed twice by the addition of 200 μl of 150 mM

TABLE 1

List of Restriction Enzymes, Recognition Sequences

and Buffers[a]

Enzyme	Cleavage Site	Restriction Endonuclease Buffers[b]
AccI	$GT(^A_C)(^G_T)AC$	10 mM Tris-HCl pH 7.5, 60 mM NaCl, 7 mM $MgCl_2$
AluI	AGCT	6 mM Tris-HCl pH 7.6, 50 mM NaCl, 6 mM $MgCl_2$
Asp718	GGTACC	6 mM Tris-HCl pH 7.8, 75 mM NaCl, 6 mM $MgCl_2$
BamHI	GGATCC	20 mM Tris-HCl pH 7.5, 100 mM NaCl, 7 mM $MgCl_2$
EcoRI	GAATTC	100 mM Tris-HCl pH 7.2, 50 mM NaCl, 5 mM $MgCl_2$
HaeII	PuGCGCPy	50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$
HaeIII	GGCC	50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$
HindIII	AAGCTT	6 mM Tris-HCl pH 7.5, 50 mM NaCl, 6 mM $MgCl_2$
HinfI	GANTC	6 mM Tris-HCl pH 7.5, 50 mM NaCl, 6 mM $MgCl_2$
HpaI	GTTAAC	10 mM Tris-HCl pH 7.4, 30 mM NaCl, 10 mM $MgCl_2$
HphI	$GGTGA(N)_8$	10 mM Tris-HCl pH 7.5, 6mM KCl, 10 mM $MgCl_2$
KpnI	GGTACC	6 mM Tris-HCl pH 7.5, 6 mM NaCl, 6 mM $MgCl_2$
PstI	CTGCAG	20 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 50 mM $(NH_4)_2SO_4$
RsaI	GTAC	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$
TaqI	TCGA	50 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM $MgCl_2$
XbaI	TCTAGA	10 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$

[a] Incubations were performed at 37°C for all reactions, except TaqI which was at 65°C.

[b] Buffers contained 6 mM BME and 100 μg/ml BSA.

NaCl, 10 mM Tris-HCl pH 7.4, followed by a 10 second centrifugation. The DNA was then eluted from the paper by incubation with 300 μl of 1 M NaCl in $T_{10}E_1$ pH 7.4, at 37°C for 12 hours. The tube was placed in a 1.5 ml microfuge tube and centrifuged for 10 seconds, followed by two rinses with 1 M NaCl in $T_{10}E_1$ pH 7.4. The NaCl eluants were pooled and the DNA was ethanol precipitated.

Freeze-thaw Isolation. A gel slice containing the desired DNA fragment was excised with a spatula over UV-transillumination (Tautz and Renz 1983). The gel slice was then equilibrated in 10 volumes of 0.3 M sodium acetate, 1 mM EDTA pH 7.0 for 30 minutes and transferred to a 0.5 ml microfuge tube that was previously punctured with a 23 gauge needle and fitted with a silinated glass wool plug. The tube containing the gel slice was frozen at -70°C for 12 hours then placed inside a 1.5 ml microfuge tube and centrifuged for 10 minutes at 12,000 x g in a Fisher 235B microfuge. The exudate was made 10 mM with $MgCl_2$ and 40 μg/ml with glycogen to improve precipitation of DNA (Tracey 1981). DNA samples were then ethanol precipitated.

Electroelution of DNA Fragments. A gel slice with the desired DNA fragment was excised from the gel and placed in a previously boiled dialysis bag containing 0.1 X TAE buffer with 10 μg/ml tRNA. As described by Maniatis et al. (1982), the bag was closed with a clamp and submerged in 0.1 X TAE

in a horizontal electrophoresis chamber in a 4°C cold room.
The DNA was electrophoresed out of the gel slice at 400
volts for 90 minutes. The buffer within the dialysis bag
was then removed and adjusted to 0.5 M NaCl, 10 mM Tris-HCl
pH 7.4 and 18% polyethylene glycol 6000 to precipitate the
DNA. The sample was place at 4°C for 12 hours. This
solution was centrifuged 12,000 x g for 15 minutes at 4°C to
collect the DNA precipitate which was resuspended in $T_{10}E_1$
(pH 7.4). The solution was extracted with an equal volume
of CIA and centrifuged for 3 minutes at 12,000 x g. The
aqueous phase was removed and ethanol precipitated.

A second electroelution protocol utilized the IBI
electroeluter (International Biotechnologies, Inc.). This
protocol calls for placing the gel slice into an elution
well. The DNA was electroeluted out of the gel slice onto a
bed of 7.5 M ammmonium acetate. The DNA in the ammonium
acetate was then alcohol precipitated.

DNA Cloning

The strategy for the construction of recombinant DNAs
is outlined in Figure 1. Genomic fragments from HindIII
digested human tonsil DNA were isolated by procedures
described above.

Ligations. The plasmid vector Puc8 was cleaved at the
unique HindIII site and subsequently treated with CIAP to
remove the terminal phosphates. The DNA was then ethanol

Figure 1. Scheme Used for Preparing a HindIII Library of Small HindIII Fragments of Human DNA. A more complete description of the methods used for each of the steps is presented in the text.

Human genomic DNA   Plasmids pUC8 or

                                                  pTZ19U

          ↓   ↓

Digested with HindIII   Digested with HindIII

          ↓   ↓

Fractionate fragments,   Alkaline phosphatase

collect those ≤ 2.0 kb.   to remove 5' phosphates

                      Ligation of inserts into vector

                                  ↓

                         Recombinant plasmids

                                  ↓

                       Transformation of host cells

                              with plasmids

                                  ↓

                  Select ampicillin-resistant colonies

                                  ↓

              Grow colonies in broth culture overnight

                                  ↓

              Isolate plasmids and characterize inserts

precipitated and resuspended in $T_1E_{0.1}$. In the ligation reaction 0.2 pmoles of insert were ligated with 0.2 pmoles of vector. The DNAs were heated to 65°C for 5 minutes and an equivalent volume of 2 X ligation buffer with 2 units of T4 ligase were added. The reaction was incubated for 12 to 16 hours at 16°C. The ligated DNAs were stored at -20°C.

Bacterial Transformations. Transformations were performed using a modification of the procedure described by Maniatis et al. (1982). Competent cells were prepared by growing Escherichia coli (E. coli) in LB at 37°C to an OD_{600} of 0.5-0.6. The cells were chilled in an ice bath, then centrifuged at 4°C at 8,000 x g. The pellet was resuspended in 1/4 the culture volume of cold, sterile 100 mM $MgCl_2$ and placed in an ice bath for 10 minutes. The suspension was centrifuged again for 5 minutes. The pelleted cells were resuspended in 1/50 culture volume of sterile, ice-cold 50 mM $CaCl_2$ and incubated on ice for 60 minutes to make the cells competent. A 2-10 μl aliquot of ligation mix was added to 0.1 ml of competent cells and placed in an ice bath for 60 minutes. The tube containing the cells was heat-shocked at 42°C for 2 minutes and 1 ml of LB broth was added. The culture was incubated at 37°C for 30 minutes. Ampicillin was added to 100 μg/ml and the incubation continued at 37°C for 30 minutes. These cells were plated on LB agar plates containing 100 μg/ml of ampicillin and incubated overnight at 37°C. In order to

grow on the plate the cells must have been transformed and were further analyzed to determine if they contained recombinant plasmids.

Isolation of Plasmid DNAs

Recombinant plasmids were isolated using a modified mini-scale alkaline lysis procedure of Maniatis et al. (1982). Bacterial cultures were grown in 20 ml of LB broth supplemented with 100 μg/ml ampicillin. Cultures were chilled on ice, then centrifuged at 900 - 1000 x g at room temperature for 10 minutes. The cells were resuspended in 1 ml of lysozyme buffer, and transfered to a 1.5 ml microfuge tube, then pelleted at 13,000 x g for 20 seconds. The bacterial cells were resuspended in 200 μl of lysozyme buffer by vortexing. A second, 200 μl aliquot of lysozyme buffer containing 5 mg/ml of lysozyme was added. After incubation for 5 minutes at room temperature, 400 μl of freshly-prepared alkaline-SDS solution (0.2 N NaOH, 1% SDS) was added, the solution was mixed gently and incubated at 65°C for 10 minutes. Three hundred and fifty microliters of 5 M potassium acetate, pH 4.8, was added and the solution mixed and chilled to 0°C to precipitate cellular proteins and chromosomal DNA. The insoluble material was removed from solution by centrifugation at 4°C for 15 minutes. The plasmid DNA was precipitated from the supernatant using polyethelene glycol or alcohol.

Isolation of plasmid DNA from bulk cell cultures followed a modified alkaline lysis protocol. One liter bacterial cultures were centrifuged at 4°C, 10,000 x g for 10 minutes. Cell pellets were washed by resuspension in ice cold 1 M NaCl - 0.01M EDTA (pH 8.0), centrifuged and resuspended in lysozyme buffer by vortexing. Bacterial cells were converted to spheroplasts upon the addition of an equal volume of lysozyme buffer with lysozyme at 8 mg/ml. The solution was incubated on ice for 15 minutes. Two volumes of freshly prepared alkaline-SDS were added. The solution was inverted gently several times and placed in a 65°C water bath until the cells were solubilized. The plasmids were neutralized and further processed as described by Maniatis et al. (1982).

Preparation and Isolation of Single-Stranded Recombinant DNA

Recombinant single-stranded plasmid DNA was produced from a culture of E. coli cells, strain MV1304F$^+$, harboring recombinant pTZ plasmids containing an M13 origin of replication. E. coli were grown in M-9 minimal media containing 50 µg/ml ampicillin. Starter and subsequent cultures were grown overnight at 37°C in a rotary shaker. A 1 ml aliquot of a starter culture was used to inoculated 50 ml of fresh LB broth containing ampicillin. Twenty minutes after starting the culture, a second inoculum containing 3.0 X 10^9 plaque forming units of M13KO7 helper phage was added.

The helper phage provides kanamycin resistance to the cells
and allows for the production of a single strand of the
plasmid and its packaging into a M13 phage particle.
Kanamycin was added to 70 µg/ml forty minutes from the
start of the culture and incubation was continued for 12 -
16 hours. Culture fluids were chilled on ice, centrifuged
twice for 15 minutes at 17,000 x g and the resulting pellets
discarded. One-fourth volume of 27% PEG in 3.3 M NaCl was
added to the supernatant. The solutions were mixed by
inversion and chilled in an ice-water bath for 1 hour. The
precipitated phage was collected by centrifugation at 4°C
for 15 minutes at 17,000 x g. The phage pellet was
resuspended in 2 ml of $T_{10}E_1$ pH 7.4 by vortexing and spun at
12,000 x g to pellet cells and debris. The phage was then
reprecipitated from the supernatant in the cold by adding
one-fourth volume of 27% PEG-3.3 M NaCl. After chilling for
30 minutes, the precipitate was collected by microfuge
centrifugation at 4°C. The pellet was resuspended in $T_{10}E_1$
pH 7.4. Phenol/CIA and CIA extractions were done. The
single-stranded DNA was alcohol precipitated and resuspended
in distilled water. The orientation of the insert within
the single-stranded phage DNA was determined by performing a
C-test (Messing and Vieira 1982). Recombinant single-
stranded DNA from separate isolates were mixed in DIGE
buffer, incubated at 65°C to allow complementry strands to
hybridize and assayed on agarose gels. Complementary

inserts were identified as high molecular weight DNAs migrating slower than single-stranded controls.

Polyethylene Glycol Fractionation of DNA

Fragments released from genomic or recombinant plasmid DNAs by restriction cleavage were PEG fractionated to separate low molecular weight fragments (less than 1600 bp) from higher molecular weight fragments. Higher molecular weight fragments were precipitated for 12-16 hours with 5% PEG at 37°C in 0.5 M NaCl, 10 mM EDTA and 10 mM Tris-HCl pH 7.4 (Schleif and Wensink 1981). The shorter DNA fragments were collected from the supernatant by precipitation with 18% PEG at 4°C. These fragments were suitable for end-labeling or further fractionation on gels.

Alcohol Precipitations of DNA

This procedure was used to collect and concentrate DNA. Ammonium acetate (5.0 M) was added to a final concentration of 0.3 - 2.5 M and either 2.5 volumes of 95% ethanol or 1 volume of isopropanol were added. If ethanol was preferred, the solution was chilled for 10 minutes in an ethanol-dry ice bath. If isopropanol was employed, the precipitation solution was chilled in an ice-water bath for 10 minutes. The DNA precipitate was collected by centrifugation at 12,000 x g using a Fisher Model 235B microfuge. Centrifugation of ethanol precipitates was performed at 4°C while centrifugation for isopropanol

precipitates was performed at room temperature. The precipitates were rinsed with room temperature 70% ethanol and dried under vacuum for 10 minutes.

DNA Fractionation by Centrifugation

To remove contaminating proteins and host DNA, some plasmid DNAs were purified by cesium chloride density gradient ultracentrifugation. DNA was dissolved in 15 ml of $T_{10}E_1$ (pH 8.0); 15.8 g CsCl were added and dissolved before the addition of 0.75 ml of 10 mg/ml EBr. The refractive index of the solution was adjusted to 1.3905. Samples were centrifuged in a 70Ti rotor for 36 to 44 hours at 41,000 rpm (120,000 x g) at 20°C. The plasmid DNA forms I, II and III were collected and extracted with water-saturated n-butanol to remove the EBr. The extracted DNAs were dialyzed against several changes of $T_{10}E_1$ (pH 7.4) and alcohol precipitated.

Measurements of DNA Concentrations

For most preparations, DNA concentrations were calculated from absorbance at 260 nm. An absorbance of one at 260 nm in a one centimeter pathlength cell corresponds to 50 μg/ml (Volkin and Cohen 1954).

For small quantities of DNA the concentrations were measured by using the enhancement fluorescence of DAPI when complexed with DNA (Brunk et al. 1979). This procedure was used only when the DNA size was greater than 1.0 kb and it

was not extensively nicked. Samples were added to the DAPI

solution and the fluorescence was measured with a

fluorescence spectrophotometer (Perkin-Elmer model 350-40)

at an excitation wavelength of 360 nm and an emission

wavelength of 450 nm. DNA concentrations were obtained by

extrapolating the fluorescence values from a standard curve.

The DNA used to establish a standard curve depended upon the

nature of the unknown DNA. If the sample DNA was genomic,

the standard used was salmon sperm DNA; if the sample was

plasmid DNA, then a plasmid DNA of known concentration was

used as a standard.

Two additional methods for determining relative DNA

concentrations were employed. As in the DAPI protocol, the

DNA used as a standard must reflect the type and the form of

the sample DNA. Both methods require a series of dilutions

for the DNA standard and sample DNA. The dilutions were

either mixed with EBr and spotted onto a UV transilluminator

or subjected to gel electrophoresis in the presence of EBr.

In either case, a positive/negative photograph was taken and

the negative was scanned by a densitometer. Peak heights or

peak areas of the standard DNAs were used to generate a

curve from which the sample DNA concentration was

interpolated.

Gel Electrophoresis of DNA

Agarose Gels. Agarose was added to 100 ml of 50 mM

TBE buffer to a final concentration ranging between 0.5 and

1.2% (w/v), depending upon the size of the fragments to be separated. Agarose solutions were brought to a boil to melt the agarose, cooled to 65°C then poured into a gel mold containing a gel comb and allowed to solidify at room temperature. In some cases EBr was added to the buffer and the melted agarose matrix to a final concentration of 0.5 μg/ml. (This compound, when complexed to DNA, increases a DNA fragment's apparent molecular weight and is not used when a true measurement of molecular weight is desired.) Prior to gel loading the DNA was mixed with 1/5 volume of SOG loading buffer. Gels were electrophoresed at 100 volts (constant voltage) until the tracking dye reached the desired end-point. After electrophoresis and EBr staining, the DNA was visualized by UV-transillumination and photographed with Polaroid type 55 positive/negative or types 57 or 665 high speed films.

Polyacrylamide Gels. Stock solutions of polyacrylamide (19/1 or 30/1 acrylamide:bis in deionized water) were diluted to the desired final concentration in 50 mM TBE buffer and degassed for 2 minutes. Ammonium persulfate and TEMED were added to 0.15% (w/v) and 0.04% (v/v), respectively, and the gel solution poured between gel casting plates (1.5 x 16 x 18 cm). The gel was allowed to polymerize for at least 1 hour. After prerunning the gels at 150 volts for 1 hour, DNA samples (mixed with 1/5 volume of SOXB loading buffer) were pipetted into the wells.

Electrophoresis was performed at 200 volts (constant voltage) until the marker dyes migrated to the desired positions in the gel. DNA was visualized by post-staining the gels in 0.5 µg/ml EBr followed by UV transillumination and photography.

Sequencing Gels. Sequencing samples were electrophoresed on 0.2 or 0.4 mm thick denaturing gels. The gels contained 6 or 8% polyacrylamide (19/1 acrylamide/bis), 8.3 M urea and 50 mM TBE buffer. The gel solution was filtered and degassed. Ammonium persulfate and TEMED were added to 0.07% (w/v) and 0.03% (v/v), respectively. The solution was mixed and poured between glass plates with spacers of the appropriate thickness. After adding the sample well-forming comb the gel was allowed to polymerize for at least one hour before pre-electrophoresis at 2000 V for a minimum of 15 minutes. Samples containing sequencing loading buffer were heated for 5 minutes at 90°C, then chilled in an ice bath. Aliquots of 2 to 6 µl were loaded into the wells and electrophoresis was continued at 2000 V until the tracking dyes had migrated to the desired positions. The gel was soaked in 5% acetic acid/5% methanol (v/v) for 15 minutes to remove the urea before drying.

Determination of DNA Molecular Weights

A BASIC computer program (Duggleby et al. 1981; Lowe

1986) was used to determine the sizes of DNA fragments from their relative electrophoretic mobilities. Fragments of known sizes were used to establish a standard curve from which the program calculated the molecular size of the sample fragments. Standards included bacteriophage lambda DNA restricted with EcoRI+HindIII, the plasmid pBR322 restricted with HinfI or HaeIII and the 171 bp alphoid ladder of rhesus genomic DNA partially cleaved with HindIII. Nucleotide sequence analysis was performed using the Molgen Jr. Programs of Lowe (1986).

Transfer of DNA from Agarose Gels to Solid Supports

DNA fragments which were fractionated by electrophoresis in agarose gels were transferred to solid supports according to Southern (1975). Gels were soaked in four volumes of 0.5 M NaOH, 1.5 M NaCl for 30 minutes with gentle shaking. After a brief rinse in deionized water, the gel was neutralized for 30 minutes in 4 volumes of 0.5 M Tris-HCl pH 6.8, 3.0 M NaCl. After neutralization, the gel was transferred to a capillary blotting apparatus containing 12 X SSC. A sheet of either nitrocellulose, Gene Screen, or Zeta Probe membrane, wetted sequentially in dH_2O, 2 X SCC, then 12 X SSC, was placed on top on the agarose gel. Two sheets of wetted 3MM filter paper were placed on top of the transfer membrane and a two inch stack of dry paper towels was placed on top of the filter paper. This allowed the DNA to be drawn into the solid support via the buffer flow

induced by capillary action. After the transfer was complete, the solid support membrane was rinsed in 100 ml of 2 X SSC and then baked in a vacuum oven at 80°C for three hours.

Slot Blotting of DNA to Solid Supports

DNA samples were immobilized on Zeta Probe, Gene Screen and nitrocellulose membranes by filtration through a slot-blotting manifold (SRC 072/9, Schleicher and Schuell). Preparation of the sample DNA depended upon the membrane to be used. Genomic DNA and plasmid DNA were partially depurinated at room temperature in 0.2 N HCl for 15 minutes. (This allowed the conversion of these circular DNAs to linear forms by alkaline cleavage at the depurination sites in the next step.) The DNA was then denatured by adding NaOH to 0.3 N and incubating at 65°C. After a 15 minute incubation, the samples were chilled on ice. Aliquots of 300 μl of the denatured sample were slotted directly onto Gene Screen or Zeta probe membranes. If nitrocellulose was employed, prior to slotting the samples were neutralized with an equal volume of 5 M ammonium acetate. Gene Screen and Zeta Probe membranes were dried at 60°C for 2 hours. Nitrocellulose filters were dried at 80°C in a vacuum for 3 hours.

Genomic and Recombinant DNAs Employed as Hybidization Probes

DNAs used as probes in this work were classified in two catagories, those which were isolated from genomic DNAs and those which were isolated as recombinant DNAs. In general, genomic probes contain a collection of DNAs, some of which are repetitive while others may be single copy in nature. The African green monkey (AGM)/HindIII alpha probe, for example, is highly enriched for alpha DNA (P.R. Musich, personal communication). Alpha and non-alpha DNAs in genomic probes will identify homologs in hybridization reactions. Probes developed from recombinant plasmids or the inserts of recombinant plasmids will identify homologs of only those inserts. Table 2 provides a list of the probes used in this work. In addition, the restriction enzyme used to release the fragment(s), the method of isolation and the method of labeling are identified.

Preparation of Radioactive Probes

Radioactive DNA probes were prepared by one of four methods: Nick translation, 5' end-labeling, 3' end-labeling (all as described by Maniatis et al. 1982) or hexanucleotide primer labeling (Feinberg and Vogelstein 1984). For nick translation reactions, approximately 2 μg of DNA was diluted in nick translation buffer with 220 μCi of high specific activity [alpha-^{32}P]-dATP. Labeling was started by the addition of 10 units of DNA polymerase I and

TABLE 2

DNA Probes

Probe	Enzyme used for fragment release	Method for fragment isolation[a]	Type of radioactive labeling[b]	DNA Source
Genomic[c]				
AGM	HindIII	PEG PPT, or EELN or SuGr	3' or 5'	This work
Rh	"	"	"	"
Hu	"	"	"	"
Recombinant Plasmids				
pE1	–	–	N. T.	Gray et al. 1985
pBα(1°)5	HindIII	EELN	3', 5' or N. T.	Shafit-Zagardo et al., 1982
EcoR1-2°	EcoRI	"	"	"
PuHu7	HindIII	"	"	This work
PuHu25	"	"	"	"
PuHu26	"	"	"	"
PuHu32	"	"	"	"
PuHu33	"	"	"	"
N6.1	KpnI or Asp718	"	"	Musich and Dykes 1986
N6.3	"	"	"	"
N6.4	"	"	"	"

[a] DNA was isolated by PEG precipitation (PEG PPT), electroelution (EELN) or sucrose gradients (SuGr)

[b] 3' or 5' represents 3' end-labeling or 5' end-labeling; N.T. represents Nick Translation

[c] Abbreviations for genomic DNA indicate the source of the DNA used as a probe; AGM - African green monkey; Rh - Rhesus monkey; Hu - Human

0.125 ng DNaseI; the reactions were incubated at 18°C for approximately 40 minutes and terminated by the addition of EDTA to 10 mM. The DNA was purified by ethanol precipitations.

Prior to 5' end-labeling, the 5' terminal phosphates were removed from the DNA. The DNA was diluted in Wittig buffer with 1 to 5 units of CIAP at 37°C for 30 to 90 minutes. To stop the reaction, EDTA was added to 10 mM and the solution was heated to 65°C for 10 minutes. CIAP was removed from the solution with a CIA extraction followed by ethanol precipitation of the DNA. For labeling, the DNA fragments were incubated in kinase buffer containing approximately 220 μCi of [gamma-^{32}P]-ATP and 10 units of T4 polynucleotide kinase at 37°C for 15 minutes. The reaction was stopped with the addition of an equal volume of 2.5 M ammonium acetate. The labeled DNA was purified by ethanol precipitation.

For 3' end-labeling, the DNA was resuspended in nick translation buffer with 200 μCi of high specific activity [alpha-^{32}P]-dATP. The reaction was started by the addition of 2 units of the Klenow fragment of DNA polymerase I and incubated at room temperature for 15 minutes; cold chase deoxynucleotides were added to fill out the ends. The reaction was incubated for a second 15 minute interval before stopping with an equal volume of 5 M ammonium acetate, EDTA to 2 mM and 1 μg of carrier tRNA. The DNA was collected by ethanol precipitation.

A gel slice containing the DNA to be labeled by the
hexanucleotide primer method was cut from a SeaPlaque
agarose gel. After equilibration against and a 3-fold
dilution with dH_2O the gel slice was heated to 96°C for 10
minutes to melt the agarose and denature the DNA. After
mixing with oligolabeling buffer containing [alpha-^{32}P]-dATP
(220 μCi) and 40 ng/ul BSA, the reaction was started by the
adding Klenow DNA polymerase I to 0.1 units/ul. After room
temperature incubation for 3 hours the reaction was
terminated by freezing.

Determination of Radioisotope Incorporation

Prior to precipitation or freezing of the labeling
reactions, a 1 μl aliquot was removed, added to 4 μl of 5 M
ammonium acetate and briefly chilled in an ice bath.
Ninety-four microliters of sonicated salmon sperm DNA (700
μg/ml) and 1 μl of 0.5 M EDTA pH 8.0 were added. One
microliter of this dilution was removed, spotted onto a
Whatman GF/C glass fiber filter and the radioactivity
determined. The counts represent 1% of the total counts
present in the diluted sample. To the remaining 99 μl of
diluted reaction mixture, 1 ml of cold HCl solution (1 N
HCl, 1% NaH_2PO_4 and 1% sodium pyrophosphate) was added. The
solution was chilled on ice for 5 minutes and filtered
through a GF/C filter, which was rinsed with 7 ml of cold
HCl solution. The radioactivity of the filter was counted
for the acid-precipitable counts. A ratio of precipitable

counts to total counts yielded the efficiency of incorporation.

Filter Hybridization

An aliquot of radiolabeled DNA (0.5×10^6 to 1.0×10^6 cpm) was used as a probe in hybridization reactions. DNA probes prepared by hexanucleotide primer labeling were heated to 95°C for 10 minutes to liquify the agarose. An aliquot of the labeling reaction was mixed with an equal volume of 100% formamide. This solution was heated to 65°C for 20 minutes to insure complete DNA denaturation, then added to a hybridization solution containing 50% formamide.

DNA probes prepared by 3' or 5' end-labeling and by nick translation were denatured by adding sodium hydroxide to 0.3 N and heating for 15 minutes at 65°C. The denatured probe was then added to the hybridization solution.

Solid support membranes used in Southern blotting were pretreated prior to hybridization. Each filter was washed twice for 20 minutes at 65°C in 0.1 X SSC with 0.5% SDS, then rinsed three times at room temperature with 3 X SSC. After this pretreatment the filters were prepared for hybridization in either Denhardt's solution (Denhardt 1966) or formamide hybridization solution. If hybridization in Denhardt's was selected, a 3-hour pretreatment was performed in 20 ml of 10 X Denhardt's solution containing 3 X SSC, 10 mM Tris-HCl pH 7.4 at 65°C. Denatured salmon sperm DNA was added to 50 µg/ml, SDS to 0.1%, and the incubation

continued for 3 additional hours. The denatured DNA probe was then added.

Hybridization in formamide required a pretreatment in hybridization solution at 42°C for 2 hours. At this time the pretreatment solution was replaced with fresh hybridization solution and denatured on a probe was added.

DNA fragments were allowed to hybridize at various temperatures for 12-16 hours. Temperature selection was dependent upon the type of hybridization solution employed and the desired stringency. Hybridizations in Denhardt's solution were incubated at 65°C or less, while hybridizations in formamide solutions were incubated at 42°C or less.

Filters were subjected to three posthybridization stringency washes in 3 X SSC-0.1% SDS for 30 minutes each. Wash temperatures varied according to the stringency of hybridization. The washed filters were then placed in sealed plastic bags and subjected to autoradiography.

Reuse of Solid Supports

DNA fragments transferred to nylon membranes (Zeta and Gene Screen) or nitrocellulose membranes were subjected to several cycles of hybridization with different probes. Previously hybridized probes were stripped off nylon membranes in two 20 minute washes at 42°C in 0.4 N sodium hydroxide. Probes bound to nitrocellulose filters were removed by two 20 minute washes at 65°C in 50% formamide,

2 X SSC and 0.5% non-fat skim milk. The filters were then rinsed at room temperature for 5 minutes in 2 X SSC. Stripped filters were sealed in new plastic bags and autoradiographed to insure that no radioactivity remained.

Detection of Recombinant Plasmids by DNA Hybridization

Ampicillin resistant colonies from transformation reactions were transferred to Whatman 541 filter disks (Taub and Thompson 1982). Filters containing bacterial colonies were placed on a LB-agar plate containing 100 μg/ml of ampicillin. The colonies were allowed to grow to a diameter of 2-4 mm before the filter was transfered to LB plates containing 250 μg/ml chloramphenicol and incubated for 16 hours. These colonies were lysed and their DNA denatured and immobilized on the filter by washing the filter twice for five minutes in each of the following solutions: 0.5 N NaOH, 0.5 N Tris-HCl pH 7.4, 2 X SCC pH 7.0. The filters were briefly washed in 95% ethanol, air dried and stored in a desiccator until hybridized.

Dideoxy Sequencing

The N6.1 insert in a recombinant single-stranded viral DNA molecule was sequenced using the dideoxy procedure of Sanger et al. (1977). Sequencing was facilitated using a universal M13 17-nucleotide sequencing primer. In a 10 μl reaction, 2 μg of DNA template was mixed with 5 ng of primer and 1 μl of 10 X HindIII buffer. The template-

primer mix was heated at 95°C for 5 minutes, 65°C for 10 minutes and 55°C for 10 minutes. It was then incubated in a beaker containing 250 ml of water at 55°C which was allowed to cool at room temperature for 30 minutes. Then 1 μl of 250 mM DTT, 3 μl of ^{35}S-dATP (8.1 μCi/μl), 1 μl of 2 X HindIII buffer and 1 μl of DNA polymerase Klenow fragment (5 units/μl) were added to the template-primer reaction. An equal volume of the template-primer mix was added to appropriate dideoxy- and deoxynucleotide mixtures (U.S. Biochemical Corp. Gene Scribe-Z manual). The polymerization reactions were incubated at room temperature for 15 minutes at which time 1 μl of chase solution was added to each reaction. The reaction was allowed to proceed for another 15 minutes at room temperature before stopping with the addition of 14 μl of sequence gel loading buffer.

Autoradiography

Gels and filters containing ^{32}P-labeled DNAs were exposed to Kodak XAR-5 or SB-5 film with a Dupont Cronex Lightning Plus intensifying screen placed behind the film. The film was exposed in an X-ray exposure holder (Kodak) at -80°C for 12 hours to 2 weeks, depending on the radioactivity of the sample. Gels with ^{35}S-labeled DNA were fixed and dried onto filter paper for exposure to XAR-5 film at room temperature. Films were developed for approximately 7 minutes in GBX developer, fixed in Rapid Fix for 5 minutes and then rinsed for 10 minutes in water.

CHAPTER 3

Results

The human genome is very large and contains an apparent myriad of repetitive sequences. By virtue of their numbers it would appear to be a simple task to isolate a representative of a specific repetitive DNA family. Because of the abundance of reiterated sequences, it is necessary to establish isolation protocols which select for known features of a repetitive DNA family, excluding other repetitive DNA sequences. These features may have foundations in molecular evolution in which DNA from one species is used to identify a sequence in a divergent species. Other selective features may be physical, such as a particular restriction enzyme site periodicity unique to the desired repetitive DNA family. These traits can be used with conventional fractionation procedures, such as electrophoresis, to isolate representatives of a particular repetitive DNA family.

Isolation and Characterization of Tandemly Repetitive DNAs: Cloning and Identification of Recombinant Plasmids

Genomic DNA was isolated and digested with the restriction enzyme HindIII. After several cycles of size fractionation, by PEG precipitation and sucrose density gradients, the enriched shorter fragments were ligated into the unique HindIII site of the plasmid pUC8. The

40

transformation of JM109, an ampicillin sensitive strain of
E. coli, produced over 400 ampicillin resistant colonies.
One hundred colonies were picked and plasmids were isolated
by the mini-prep procedure. Each plasmid was treated with
HindIII and electrophoresed to determine if an insert was
contained in the plasmid.

Of the 100 clones selected, 80 contained inserts, 36 of
which are represented in Figure 2A. Many of the plasmids
contain multiple inserts ranging in size from 0.17 kb to 1.6
kb sizes which result in a complex electrophoretic pattern.
The EBr intensity of some insert bands is not as intense as
others indicating a difference in the amount of DNA loaded
onto the gel and the presence of mixed recombinant plasmids.
The HindIII-digested plasmids were transferred from the
agarose gels to Gene Screen membranes by Southern blotting.
Recombinant inserts were screened by hybridization to
identify those clones most homologous to an isotopically-
labeled AGM genomic alpha DNA probe. This probe was used to
select for phylogenetically related repetitive alphoid DNA
families in the human and AGM genomes. Stringency
conditions (3 X SSC, 65°C) allowed hybridization of
molecules with a homology of 80% or greater (Musich et al.
1980). Upon autoradiography of the blot, seven human
recombinant inserts homologous to the AGM probe were
identified (Figure 2B). Single insert bands in lanes
containing clones 10, 21 and 25 were identified.

42

Figure 2. Identification of Human Recombinant Clones
Homologous to an AGM Genomic Alpha Probe. Panel A. Rows 1
and 2 show a photograph of recombinant plasmids which were
digested with HindIII and electrophoresed through a 0.7%
agarose gel containing EBr. Examples of plasmids
containing multiple HindIII inserts can be seen in row 1,
lanes 10 and 11, and row 2, lanes 4, 8 and 9. Panel B.
The DNA contained in the gel was transferred to a Gene
Screen filter by Southern blotting. The filter was
hybridized with an AGM alpha-rich probe to identify
plasmids containing homologous inserts. Homologous DNAs,
with their names in parenthesis, were identified in row 3,
lanes 2 (PuHu7), 5 (PuHu10) and 16 (PuHu21), and row 4,
lanes 1 (PuHu25), 2 (PuHu26), 8 (PuHu31) and 9 (PuHu32).
Smearing noted in rows 1 and 2 is due to the probe
hybridizing to partially cleaved plasmids. The marks found
on the right of each photograph represent fragment sizes in
kb.

FIGURE 2

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

ROW 1

-1.6
- 0.5
-0.3
-0.2

ROW 2

-2.4
-1.2
-0.6
-0.2

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

ROW 3

-0.6
-0.3

ROW 4

-1.2
-0.9
-0.6
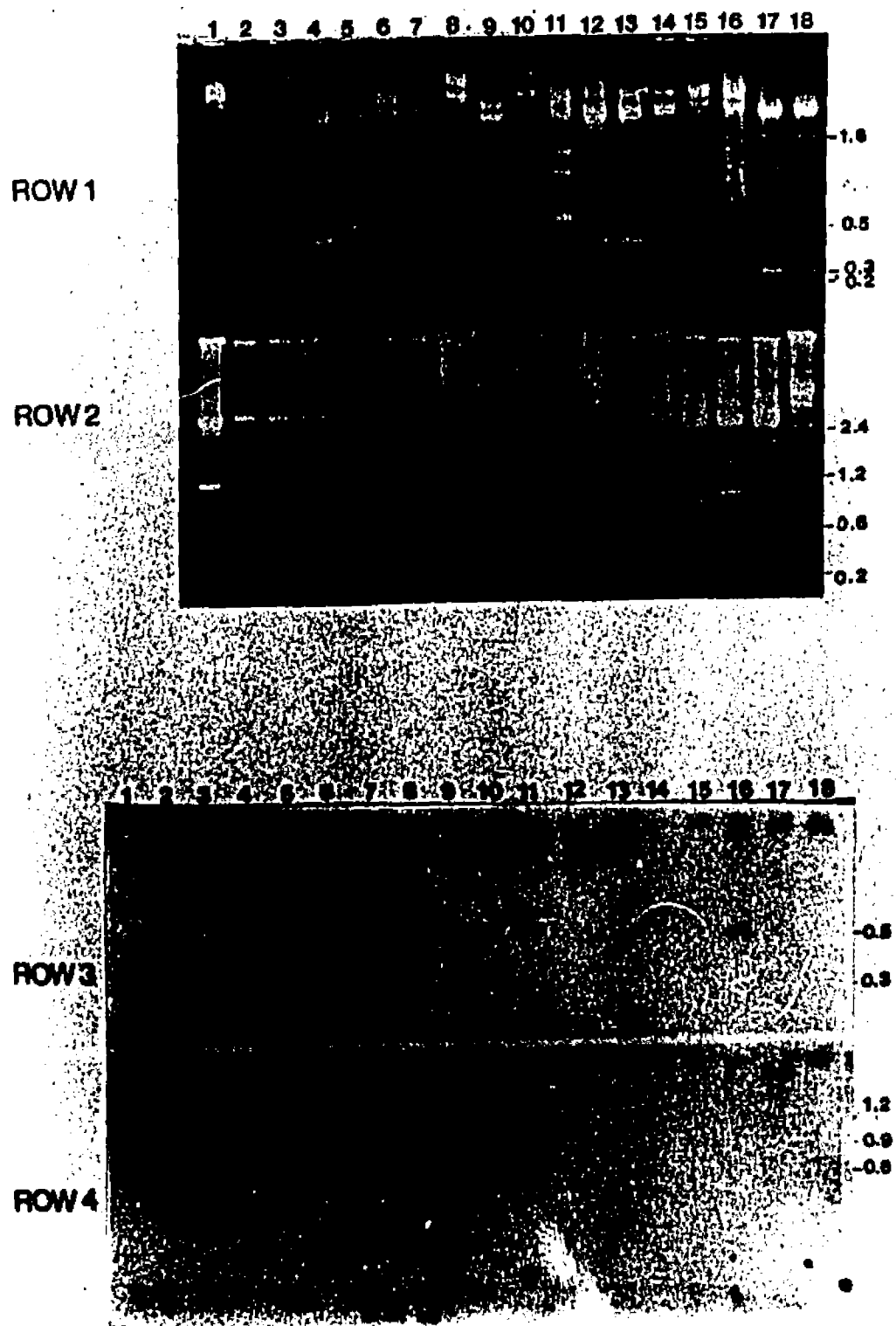
The approximate insert sizes are 0.5, 0.4 and 1.2 kb, respectively. Clones 31 and 32 contained multiple inserts ranging from 0.6 to 2.4 kb. Two inserts in each clone, of 0.9 and 1.2 kb, hybridized to the AGM alpha probe. Hybridizations with a cloned member of the repetitive AGM alpha family did not show homology between the cloned human inserts (data not shown). Additional hybridization studies were employed to determine the organization of the recombinant inserts in the human genome.

Determination of Genomic Organization. Human genomic DNA was digested with the enzymes HindIII, EcoRI, XbaI or AluI. These restriction enzymes were selected because, with the exception of HindIII, they cleave certain human repetitive DNAs families into characteristic patterns (Manuelidis and Wu 1978; Houck et al. 1979; Ullu 1980). The digested DNAs were electrophoresed through a 1.0% agarose gel (Figure 3A) and transferred to Gene Screen for hybridization analysis. Autoradiographic hybridization patterns were used as aids to assign the cloned DNA segments to previously described human repetitive DNA families.

General hybridization to genomic DNA was noted using PuHu clones 10, 25, 31 and 32 as probes (Figure 3B). Clone PuHu21 was not analyzed. This type of hybridization pattern, which mirrored the total DNA distribution in the gel, is expected with probes of interspersed repetitive families (Bober et al. 1986). With few exceptions

44

Figure 3. Genomic Hybridization Patterns for Interspersed Repetitive DNA Clones. Panel A shows the 0.7% agarose gel pattern of EBr-stained human DNA digested with restriction enzymes XbaI (lane 1), AluI (lane 2), EcoRI (lane 3) and HindIII (lane 4). Panel B is the corresponding autoradiographic image of the DNAs after transfer to Gene Screen and hybridization with nick-translated clone PuHu32 probe. The pattern seen is that of an interspersed repetitive DNA. Similar hybridization patterns were noted with probes from clones PuHu10, PuHu25 and PuHu31 which were also classified as representatives of interspersed repetitive DNAs (data not shown). The size markers between panels A and B represent kilobase pairs.

Figure 3

A

1   2   3   4

kb
1.03
0.93
0.86
0.68

0.34

B

1   2   3   4
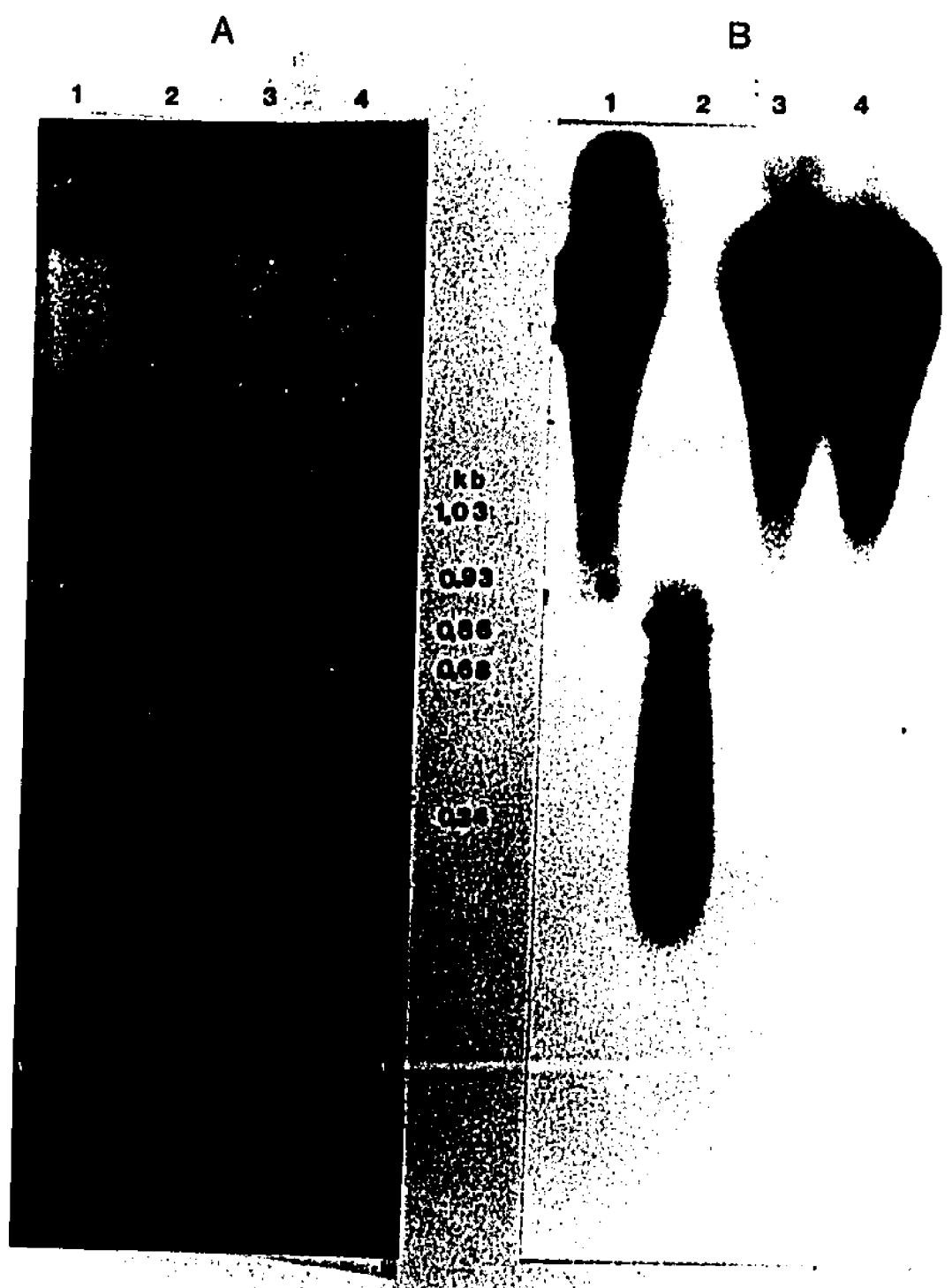
interspersed DNAs are flanked by heterologous sequences
containing a variable, heterogeneous spacing of restriction
sites. As a result, digestion of genomic DNA with
restriction enzymes will not produce a simple pattern.

Multimeric hybridization patterns were observed with
probes from clones 7 (PuHu7) and 26 (PuHu26), indicating
that these clones represent repetitive DNA families with
discrete genomic organization patterns. The PuHu7
hybridization pattern to genomic DNA digested with XbaI is
similar to that of the cloned XbaI family descriptor, pE1
(Gray et al. 1985) (Figure 4, lane 4 of panels B and C).
Each probe hybridized preferentially to the smaller segments
which form a multimeric pattern based on a 170 bp unit.
Similar hybridization patterns for these two probes were
also observed for the AluI and EcoRI digests (Figure 4,
lanes 2 and 3). However, distinct differences between PuHu7
and pE1 hybridizations are observed in the HindIII digests
(Figure 4 lane 1). For example, PuHu7 hybridizes intensely
with two DNA fragments of 0.17 kb and 0.34 kb, but only
faintly to segments between 0.4 and 1.8 kb. The pE1 probe
identifies these same segments, but hybridizes intensely
with segments of 1.0 to 1.6 kb and faintly with shorter DNA
segments. Both probes hybridize intensely with high
molecular weight DNA. These distinct hybridization
differences show that PuHu7 is a member of the XbaI family,
but represents a distinct subpopulation of it.

46

Figure 4. Identification of a Cloned XbaI Family Variant.
Panel A shows the 0.7% agarose gel pattern of EBr-stained
human DNA digested with HindIII (lane 1), EcoRI (lane 2),
AluI (lane 3) and XbaI (lane 4). Panels B and C are the
corresponding autoradiographic images of the digested DNAs
after transfer to Gene Screen and hybridization with nick
translated probes from clones PuHu7 (panel B) or the XbaI
family descriptor pE1 (panel C). Between hybridizations
the membrane was stripped to remove the previous probe.
Subsequent rehybridization of the filter allowed a direct
comparison of the genomic hybridization patterns.
Similarities in the XbaI patterns of PuHu7 and PE1 (lanes
4, panels B and C) indicated that PuHu7 is a variant of the
XbaI family. An example of the differences between the two
clones is illustrated in the HindIII digests of panels B
and C. An asterisk (*) is used to identify bands which are
less intense in panel B as compared to the corresponding
bands in panel C. The markers shown between panels A and B
represent size in kb.

A   B   C

1   2   3   4   1   2   3   4   1   2   3   4

0.95
0.86
0.68

0.34
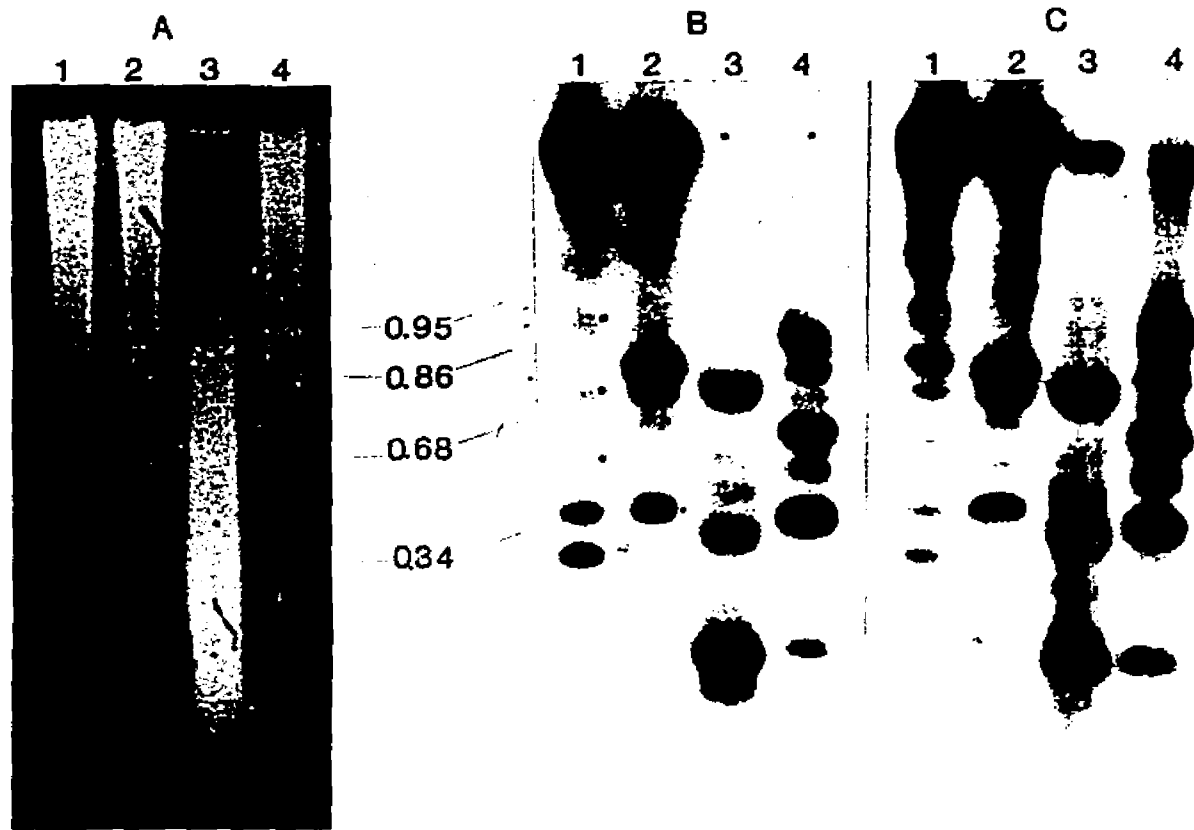
FIGURE 4

The results of hybridization with PuHu26 to genomic DNA digested with HindIII (Figure 5B, lane 1) identifies a multimeric series of bands reflecting a HindIII restriction site periodicity and repeat length of approximately 170 bp. Hybridization to the high molecular weight DNA shows that the PuHu26 genomic family contains members which resist HindIII cleavage. These HindIII-resistant sequences can be cleaved with AluI which recognizes the four internal nucleotides of the HindIII restriction site (see Table 1, page 17). AluI identifies 5 bands above a general hybridization background (Figure 5B, lane 3). This general hybridization pattern indicates that the PuHu26 genomic family contains AluI sites other than those found in the HindIII restriction sites. XbaI cleavages failed to resolve the high molecular weight, hybridizable segments (Figure 5B, lane 4), but EcoRI did produce a faint multimeric series of PuHu26-homologous segments (Figure 5B, lane 2). The intensity of hybridization to these segments, relative to those in the HindIII digests, indicates that a subpopulation of the PuHu26 genomic family contains EcoRI restriction sites at a 170 bp periodicity. Within the EcoRI pattern, a 0.86 kb band is more intense than homologous segments slightly longer or shorter in length. This may represent of a more homologous population of segments or a preferred cleavage of family sequences to 0.86 kb segments, thus increasing the amount of homologous DNAs in that band.

Figure 5. Identification of a Novel Tandemly Repetitive DNA Family. Panel A shows the 0.7% agarose gel pattern of EBr-stained human DNA digested with restriction enzymes HindIII (lane 1), EcoRI (lane 2), AluI (lane 3) and XbaI (lane 4). Panels B and C are the corresponding autoradiographic images of the digested DNAs after transfer to Gene Screen and hybridization with nick-translated probes from clones PuHu26 (panel B) or a cloned EcoRI dimer probe, RI-2* (panel C). Panel B lane 1 shows the tandem organization of genomic repetitive family described by PuHu26. A comparison of the genomic hybridizations patterns of the RI-2* (panel C) and PuHu26 shows that the PuHu26 clone represents a family of DNAs which are organized differently than the EcoRI family. Slot-blot hybridizations of the RI-2* clone with PuHu26 did not show detectable homology between the two. The markers between panel A and B represent sizes in kb.

FIGURE 5

0.98 kb
0.86
0.68
0.34
0.17
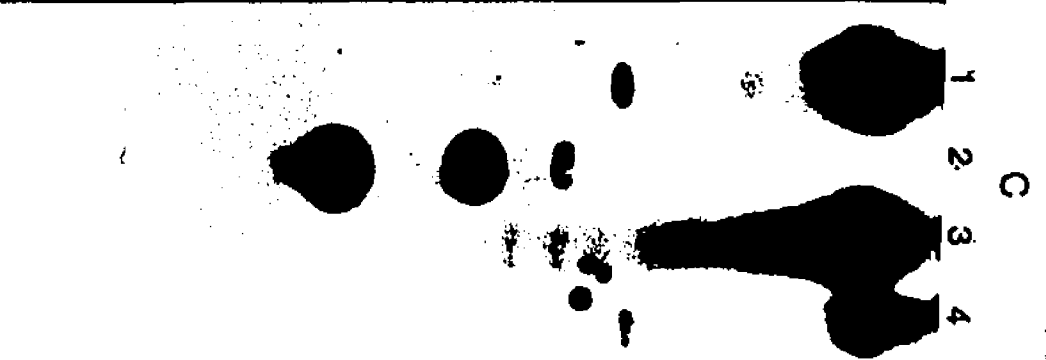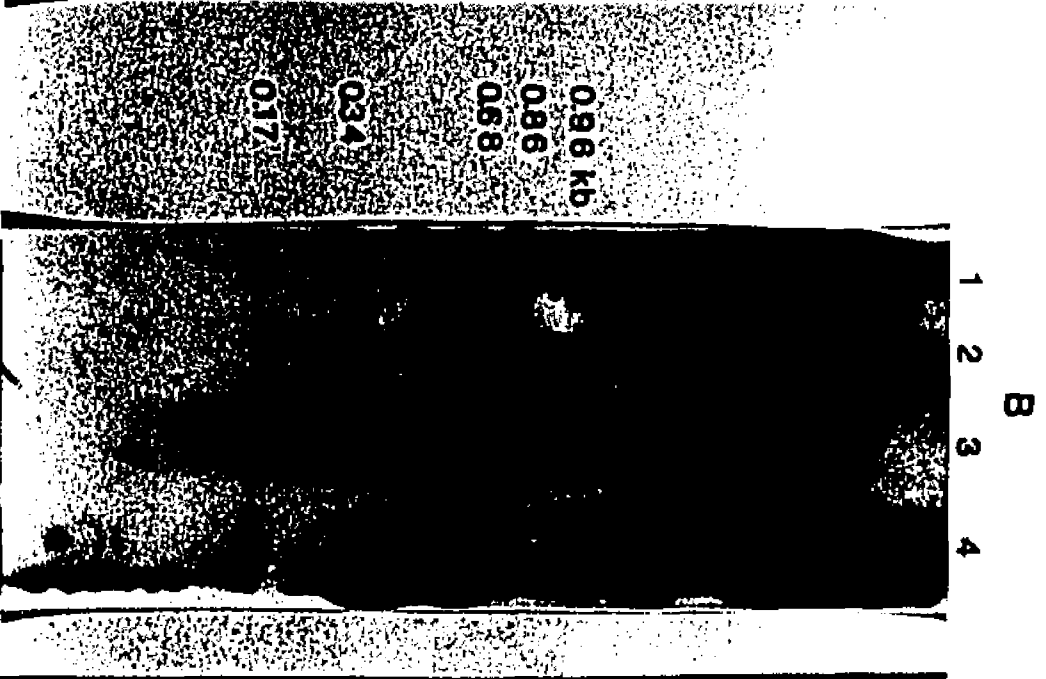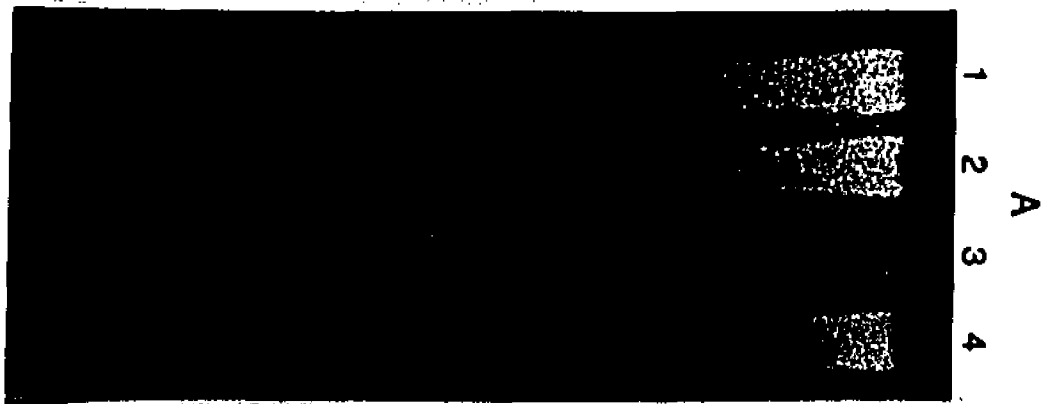
Hybridization with a cloned representative of the EcoRI dimer family [EcoRI-2'C (Manuelidis 1976)] also produced a multimeric pattern (Figure 5C, lane 2).

A comparison of genomic DNA/HindIII hybridization patterns shows that the two probes identify distinct populations (Figures 5B and 5C, lane 1). The EcoRI-2* probe hybridizes predominantly to segments of 1.0 and 1.9 kb. The PuHu26 probe identifies segments of 0.17 to 1.2 kb in increments of 170 bp. In a comparison of hybridizations to genomic/EcoR1 digests (Figure 5B and C lane 2), the PuHu26 probe identifies a multimeric series of bands. The probe identifies homologous genomic segments, in increments of 170 bp, beginning with fragments of 0.17 kb through high molecular weight DNA (Figure 5B lane 2). The EcoR1-2* probe hybridizes predominantly to bands of 0.34 kb, 0.68 kb and 1.03 kb, with slight hybridization to high molecular weight DNA (Figure 5C lane 2). In addition, slot-blot hybridization of the EcoR1-2* clone with the PuHu26 probe failed to detect homology (data not shown). Slot-blot and genomic hybridizations reveal a difference in the nucleotide sequence between the PuHu26 and the EcoRI-2* clones; therefore, the two probes describe distinct families of repetitive DNAs.

Characterization of L2Hs Variants

Recently, Musich and Dykes (1986) identified a second family of long interspersed repetitive DNAs (LINES) in the

human genome. This newly discovered LINE family was named L2Hs (2nd LINE family in Homo sapiens). Characterization of this LINE family using the L2Hs clone N6.4 identified several remarkable features: 1) The recent evolution of this repetitive DNA family and 2) its quantitative and qualitative polymorphisms within the human population. Variant sequences other than N6.4 were identified in the initial screening of a pUC19-KpnI recombinant library (Musich and Dykes 1986). Two variants, N6.1 and N6.3, exhibit features similar to those identified by N6.4. The following experiments describe both the physical features and the genomic organization of the N6.1 and N6.3 variants.

Initial characterization of the L2Hs variants N6.1, N6.3 and N6.4 showed that KpnI released 0.6 kb inserts from each recombinant plasmid. The plasmids were subjected to further restriction analysis to check for multiple inserts and to develop a restriction map. By comparing the restriction patterns of the vector, without an insert, to the cleaved recombinant plasmids, it was determined that no restriction sites for EcoRI or HindIII existed within the inserts (data not shown). Since these restriction sites flank the KpnI insertion site, a digest with EcoRI+HindIII would release any multiple inserts as one segment. Such double digests of N6.1, N6.3 and N6.4 released inserts of about 0.6, 1.2 and 1.8 kb, respectively (data not shown). As KpnI digests released only 0.6 kb inserts from each of

the plasmids, N6.3 and N6.4 must contain multiple KpnI inserts. Therefore, N6.1 contains one 0.6 kb insert, N6.3 contains two 0.6 kb inserts and N6.4 contains three 0.6 kb inserts. Digestion with HaeII showed that the inserts occurred within a 0.35 kb HaeII segment of pUC19 and that HaeII sites do not exist within the inserts of N6.1 and N6.3 (Figure 6A, lanes 1 and 2). HaeII sites do occur within two of the three inserts of N6.4 (Figure 6A, lane 3).

RsaI recognizes the four internal nucleotides of the KpnI restriction site (see Table I, page 17) and it should release the inserts from the N6 plasmids. However, RsaI digestion showed only the three vector bands for N6.1 and N6.3, while N6.4 contained two additional 0.51 kb segments (Figure 6D). The cleaved plasmids were analyzed by Southern blot hybridization. The probe used was gel purified insert from the plasmid N6.1. Hybridization results (Figure 6B and E) indicate that at least one of the inserts of both N6.3 and N6.4 are homologous to that of N6.1. Hybridization to the RsaI digests shows that none of the RsaI segments hybridize to the N6.1 probe. This indicates the presence of a high density of RsaI restriction sites in the L2Hs homologous inserts. Since the smallest DNA fragment retained by the hybridization matrix is approximately 100 bp, L2Hs fragments released by RsaI must be smaller than 100 bp. This would indicate a minimum of 7 RsaI sites within the L2Hs inserts of N6.1, N6.3 and N6.4.

52

Figure 6. Restriction and Hybridization Analysis of L2Hs
Clones. HaeII digests (panels A and B) of N6.1 (lane 1),
N6.3 (lane 2) and N6.4 (lane 3) were performed. RsaI
(panels D and E) was also used to digest N6.1 (lane 1),
N6.3 (lane 2) and N6.4 (lane 3). The cleaved plasmid DNAs
were transferred to Zeta Probe membrane by Southern
blotting and the filters hybridized with an N6.1 insert
probe (panels B and E). The hybridizations at 1.6 kb in
lane 1 and the faint bands greater than 1.6 kb of lanes 2
and 3 are a result of probe-homologous inserts remaining in
partially cleaved plasmids. The molecular weight markers
(panel C) are pBR322 digested with HinfI (lane 2) and
rhesus monkey genomic DNA digested with HindIII (lane 1).
The pBR322 segments are 1631, 517+506, 396, 344, 298,
221+220, 154 and 75 bp. The rhesus monkey segments form a
multimeric series of bands spaced at 172 bp intervals.
Molecular weight markers shown between panels B and C
represent sizes in kb.
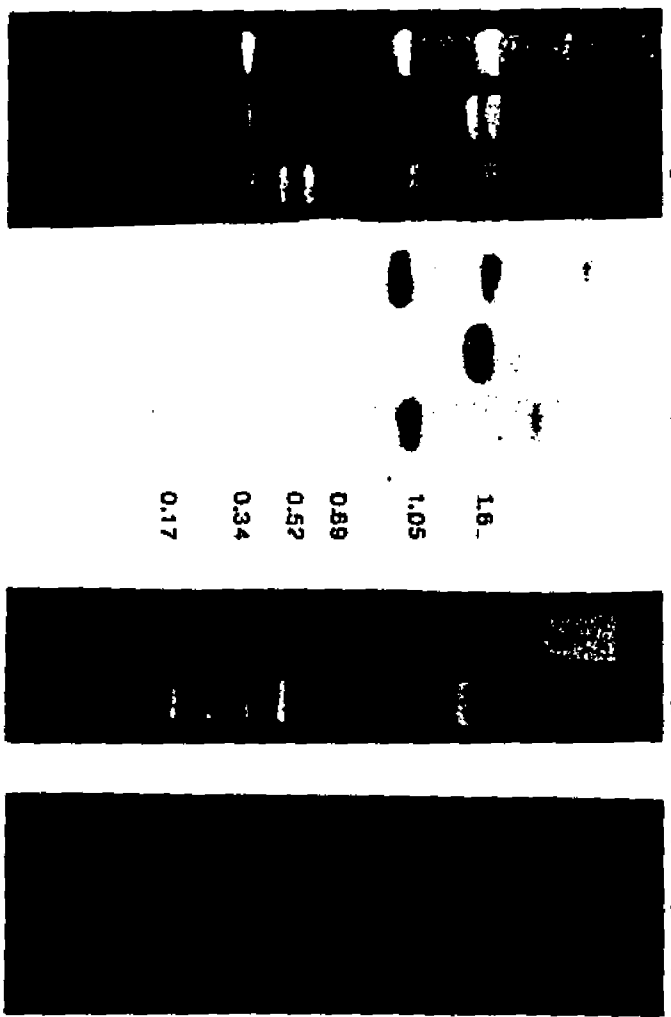
FIGURE 6

The non-homologous inserts of N6.4 were arbitrarily named D and E. Each contains a HaeII and a RsaI restriction site.

Fourteen restriction enzymes were employed to cleave the N6 inserts. Of the enzymes tested, only six cleaved segments D and E and only RsaI cleaved the L2Hs homologous insert in the N6 plasmids (Table 3). It is apparent that inserts N6.4 D and E are unique relative to the L2Hs segments. The paucity of restriction sites has not been reported as a trait for any other cloned human repetitive DNA (Manuelidis and Wu 1978; Beridze 1986; Gray et al. 1985). Thus, it distinguishes the cloned L2Hs sequence from the four human satellite DNAs, the SINE and the KpnI LINE sequences.

Genomic L2Hs Restriction Profiles

Data gathered from restriction enzyme cleavages of N6 clones describe DNA segments resistant to many restriction enzymes. It was necessary to determine if the restriction site profiles described for the N6 cleavages correctly represent genomic L2Hs DNAs. Restriction digests were designed to increase the probability of cleaving the L2Hs family by using enzymes which frequently cleave genomic DNA. Genomic DNA was digested with HinfI, HinfI+HaeIII, HinfI+HaeIII+AluI, HinfI+HaeIII+AluI+MspI and HinfI+HaeIII+AluI+MspI+MboII before electrophoresis through a 0.7% agarose gel (Figure 7A). These digests progressively deplete the high molecular weight DNA, reducing it to

Table 3

Restriction Analysis of L2Hs Recombinant Clones

Enzyme[a]	Recombinant L2Hs Plasmids					
	N6.1	N6.3		N6.4		
		A[a]	B	C	D	E
RsaI	+[c]	+	+	+	+	+
HaeIII	−[d]	−	−	−	+	−
TaqI	−	−	−	−	+	+
HinfI	−	−	−	−	+	−
Sau3A	−	−	−	−	+	+
HaeII	−	−	−	−	+	+

[a] MspI, AluI, HpaI, EcoRI, HindIII, AccI, BamHI and SalI were tested and found not to cleave the inserts of N6.1, N6.3 and N6.4

[b] A, B, C, D, E: Arbitrary assignment for each of the inserts.

[c] + indicates the presence of at least one restriction site.

[d] − indicates no restriction site within insert.

segments less than 1 kb. The DNA contained in the gel was transferred to nitrocellulose and hybridized with nick translated N6.1 probe (Figure 7B). The L2Hs segments remained longer than the bulk of the DNA indicating that they contain few sites for these restriction enzymes. The restriction profiles of genomic L2Hs sequences are, therefore, correctly represented by restriction digests of the cloned L2Hs variants.

The hybridization results also provide insight into the organization of genomic L2Hs DNAs. It can be inferred that the L2Hs family has a nonrandom organization. If these sequences were interspersed as single copy elements, the serial restriction digests should have cleaved the DNA flanking the sequences. Subsequent hybridizations would have recorded the progressive release of L2Hs-homologous DNAs to smaller molecular weight forms.

Genomic Organization of L2Hs Sequences

To determine the genomic organization of L2Hs sequences partial digests were analyzed. DNA was treated with the restriction enzyme KpnI and aliquots of the digest were removed at various times. KpnI activity was stopped by the addition of EDTA to 5 mM, SDS to 0.1% and heating to 65°C for fifteen minutes. The samples were then electrophoresed through a 0.7% agarose gel. The DNA in the gel was transferred to nitrocellulose by Southern blotting. Hybridizations with L2Hs probes indicated a minimum of two

Figure 7. Serial Restriction Digests of Human Genomic

DNAs. In panel A, genomic DNAs were digested with: HinfI

(lanes 1,6,11), HinfI+HaeIII (lanes 2,7,12),

HinfI+HaeIII+AluI (lanes 3,8,13), HinfI+HaeIII+AluI+MspI

(lanes 4,9,14) and HinfI+HaeIII+AluI+MspI+MboII (lanes

5,10,15) and electrophoresed through a 0.7% agarose gel.

Lanes 1-5 contain human tonsil DNA, lanes 6-10 contain DNA

from peripheral blood lymphocytes and lanes 11-15 contain

Namalwa DNA. Also included in lane 15 is lambda DNA

digested with EcoRI+HindIII. Lanes 16 through 20 represent

Namalwa digested with KpnI+HinfI, KpnI+HaeIII, KpnI+AluI,

KpnI+MspI and KpnI+MboII, respectively. The

electrophoresed DNAs were transferred to nitrocellulose.

Panel B shows the results of hybridization with nick-

translated N6.1 probe. The autoradiograph shows that the

L2Hs family contains few sites for the restriction enzymes

tested as the bulk of the hybridized DNA remains high

molecular weight. The 0.6 kb bands in lanes 16, 17, 19 and

20 were more intense upon a longer exposure. The KpnI

digests (lanes 16 - 20) were not complete digests. Size

markers on the right of panels represent kilobase pairs.
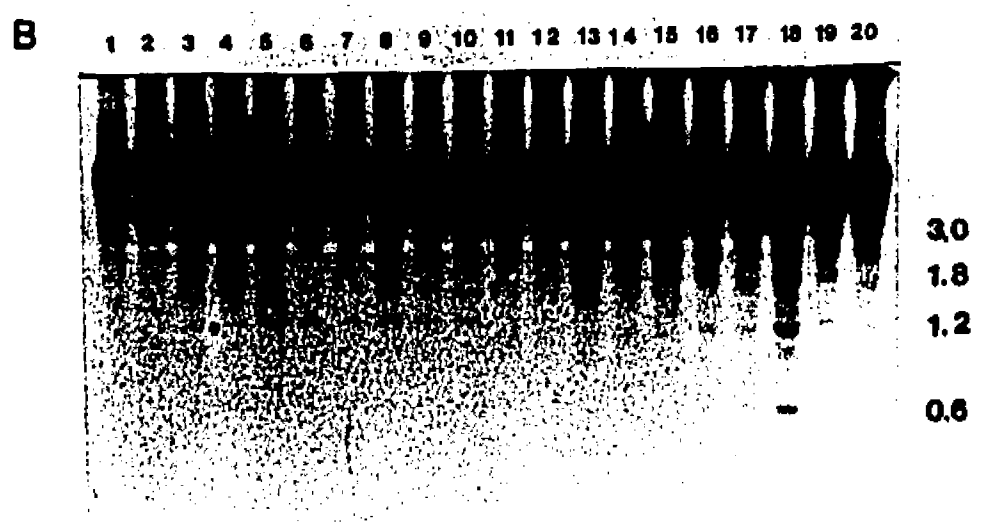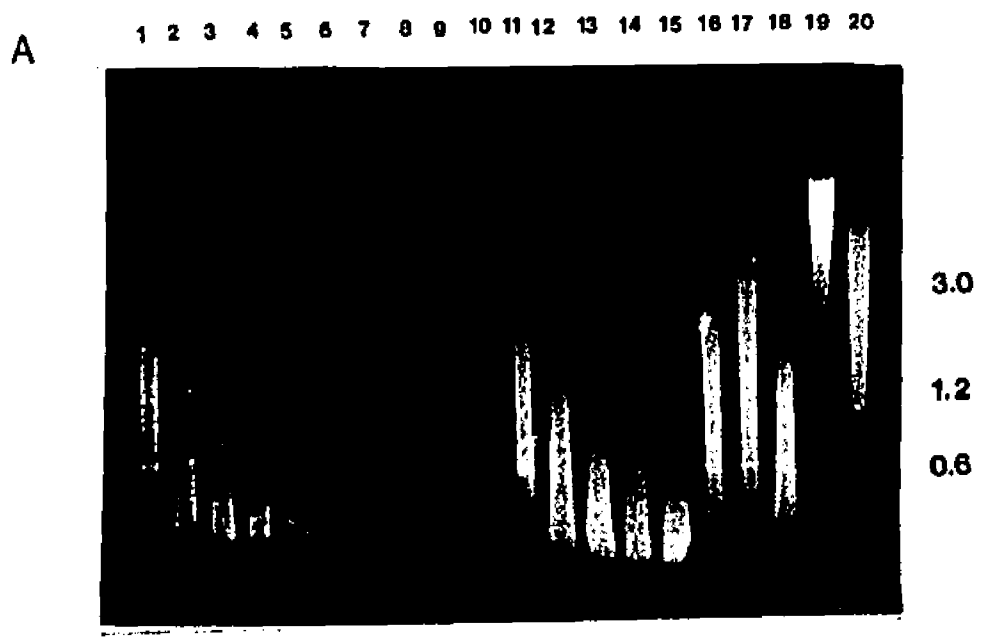
FIGURE 7

populations of family sequences. One population is resistant to KpnI, resulting in segments of 3.0 kb or greater; the second is sensitive to KpnI (Figure 8, lane h) and releases shorter segments. The N6.1 probe shows the progressive cleavage of N6.1-homologous segments at 0.6 kb intervals from 3.0 kb to 0.6 kb (Figure 8A, lanes c through e). This suggests that there is a clustering of KpnI sites at 0.6 kb intervals within some L2Hs and adjacent sequences. The additional segments hybridizing to the genomic digests with the N6.4 probe at 1.5, 1.9, and 3.0 kb are attributable to the two non-L2 inserts contained in the N6.4 probe (P.R. Musich, personal communication). Subcloning of the three N6.4 inserts reveals that one non-L2 insert identifies homologous segments at 2.0 and 3.0 kb and the other at 1.5 and 3.0 kb (data not shown). Hybridization with either N6.1 or N6.4 to the partial KpnI digests also fails to show a simple polymeric arrangement of N6 sequence repeats. If the 0.6 kb segments occur in a tandem array, the 1.2 and 1.8 kb bands should be signficantly more intense relative to the 0.6 kb band at intermediate stages of cleavage.

Polymorphisms within a Genome

In previous studies where genomic DNA was digested with 20 restriction enzymes AccI was found to cleave genomic L2Hs DNAs (P. R. Musich, personal communication). Namalwa DNA, isolated from a B-cell lymphoma cell line (Pritchett et al. 1976), was treated with the restriction enzyme AccI,

Figure 8. KpnI Partial Cleavage Patterns of Genomic L2Hs

Sequences. Namalwa nuclear DNA was treated with KpnI and

aliquots were removed at various times. The reaction was

stopped by the addition of EDTA, SDS and heating to 65°C.

The segments were resolved on a 0.7% agarose gel, blotted

to nitrocellulose and hybridized with either N6.1 (panel A)

or N6.4 (panel B) probes. a = no digestion, b = 15 minutes

digestion, c = 30 minutes digestion, d = 1 hour digestion,

e = 2 hours digestion, f = 4 hours digestion, g = 8 hours

digestion, h = 16 hours digestion. The markers between the

panels A and B represent sizes in kb.

Figure 8
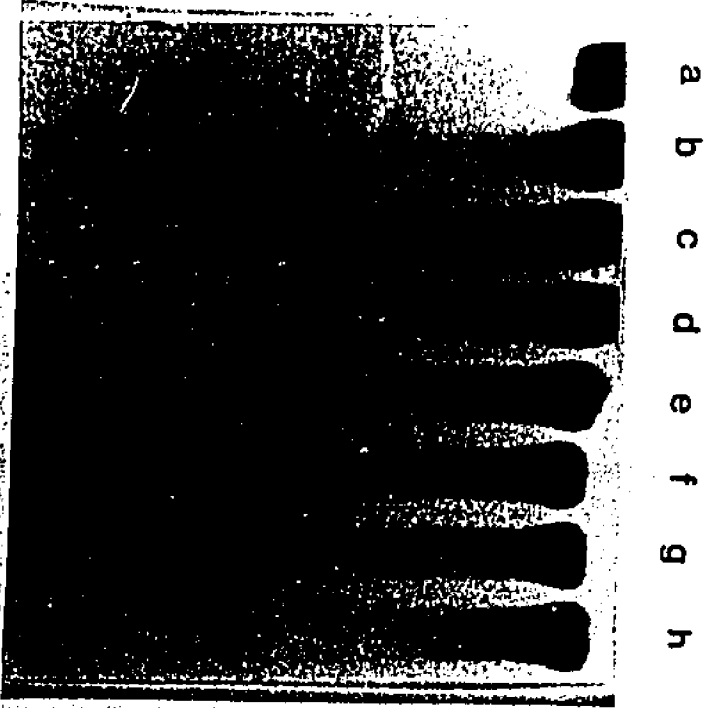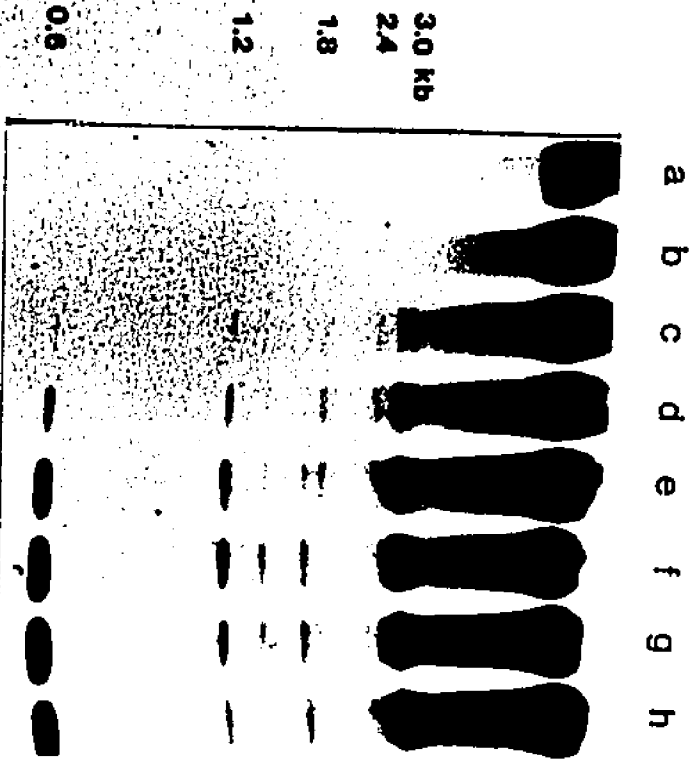
Southern blotted and hybridized consecutively with N6.1,
N6.3 and N6.4 probes. The autoradiograms of these blots
showed that a population of the L2Hs family is AccI
resistant and that each of the three probes had a
characteristic hybridization pattern. In Figure 9 the
hybridization patterns are shown and interpreted in a
schematic line drawing to illustrate these differences. The
N6.1 pattern contains a distinct band not identified by N6.3
or N6.4 (Figure 9B, block D). Absent from the N6.1 pattern
are several high molecular weight segments (Figure 9B,
blocks A and B) and numerous low molecular weight segments
(Figure 9B, blocks F and G) seen with both of the N6.3 and
N6.4 probes. The N6.3 and N6.4 patterns are very similar;
however, N6.4 hybridizes to a segment (Figure 9B, block C)
not homologous to the N6.3 probe. All three probes identify
a common set of segments (for example, Figure 9B block E),
although the distribution of hybridization intensities among
these common segments is distinct for each of the three
probes. The N6.1 hybridization pattern is the least complex
of the three patterns. Unlike the KpnI patterns (see Figure
8, page 59) there is no dominant 0.6 kb segment.

Southern blots of Namalwa DNA digested with AccI+KpnI
were hybridized consecutively with N6.1, N6.3 and N6.4
probes (Figure 10). The double digest converted some L2Hs
very high molecular weight DNAs noted in KpnI and AccI
digests to shorter DNA segments between 5 and 20 kb in

Figure 9. Polymorphisms Between Genomic Cleavage Patterns for the Cloned L2Hs Probes. Namalwa DNA was treated with AccI; the DNA segments were resolved on a 0.8% agarose gel, Southern blotted and hybridized consecutively with N6.1, N6.3 and N6.4 probes (panel A). In panel B the hybridization patterns are drawn schematically to illustrate the differences between the cloned probes. The similarities and differences between the probes are represented by blocks of bands in the interpretive drawing. Each of these blocks has been given a letter A, through G. The filters were stripped between hybridizations. Exposure and development times of the photographs were controlled to maximize the similarities between each hybridization.

A

B

N6.1   N6.2   N6.4

A
B
C
E
F
G
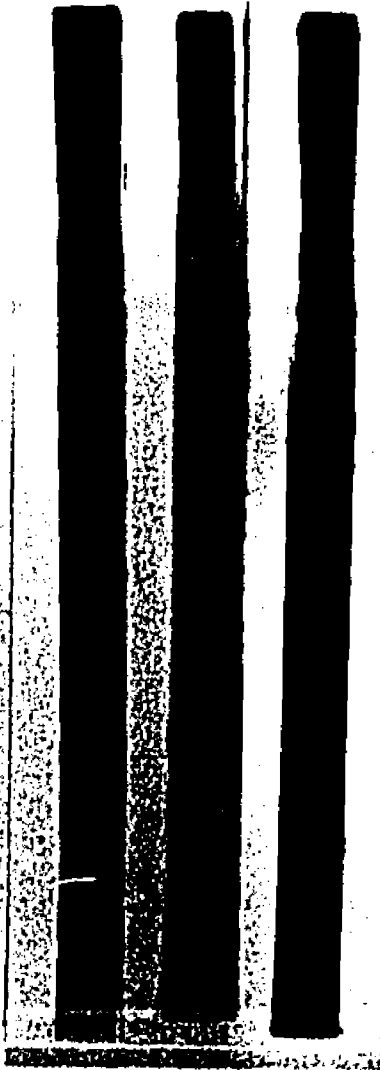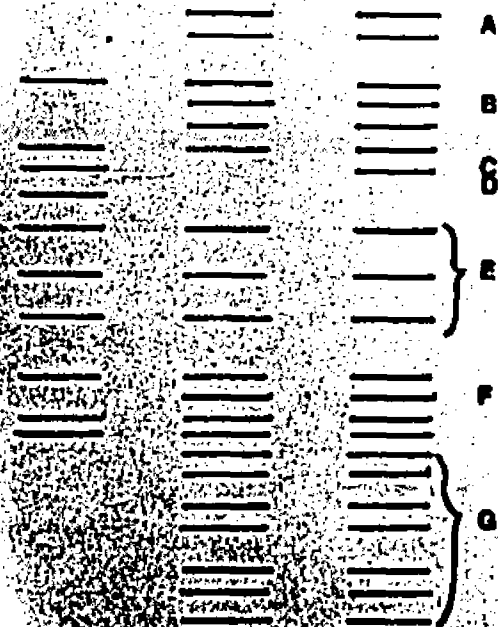
length. Each probe identifies the segments of 0.6 and 1.2 kb. The segments identified by the N6.1 probe are shared by N6.4; however, not all N6.4-homologous segments are common to N6.1 (Figure 10, blocks B, C and D). Again, the N6.1 probe displayed the least complex pattern of the L2Hs clones; the majority of hybridization occurred to segments greater than 1.2 kb. Although N6.3 and N6.4 probes hybridized to these long segments, they also identified segments less than 1.2 kb. For all three probes, L2Hs segments appear clustered within the patterns. The intensity of hybridization to these clusters is distinctive for each of the three probes. The presence of high molecular weight DNA in the AccI+KpnI double digests indicates the presence of three L2Hs populations. These three populations are KpnI sensitive, AccI-sensitive and KpnI+AccI-resistant populations.

The differences in the genomic hybridization patterns described above clearly indicated that the L2Hs inserts must contain sequences which are common to all three clones and sequences which are unique to each of the clones. To determine if the L2Hs sequences display a similar hybridization pattern in different genomes, DNA from peripheral blood lymphocytes (PBL) of 12 donors was digested with KpnI, AccI or AccI+KpnI. The samples were electrophoresed through a 0.8% agarose gel and Southern blotted to Gene Screen. Hybridization with a N6.1 probe

Figure 10. AccI+KpnI Polymorphisms Between Genomic Cleavage
Patterns for the Cloned L2Hs Probes. Namalwa DNA was
treated with AccI+KpnI, Southern blotted and hybridized
consecutively with (displayed left to right) N6.1, N6.3 and
N6.4 probes (panel A). Panel B contains interpretive
drawings in the same order as the photographs are
presented. The filters were stripped between
hybridizations. Similarities and differences have been
assigned letters, A through K. The exposure and
development times of the photographs were controlled to
maximize the similarites between each hybridization. Some
of the N6.1 hybridization bands present on the original
film were lost during photographic processing.

A

B

N6.1   N6.3   N6.4
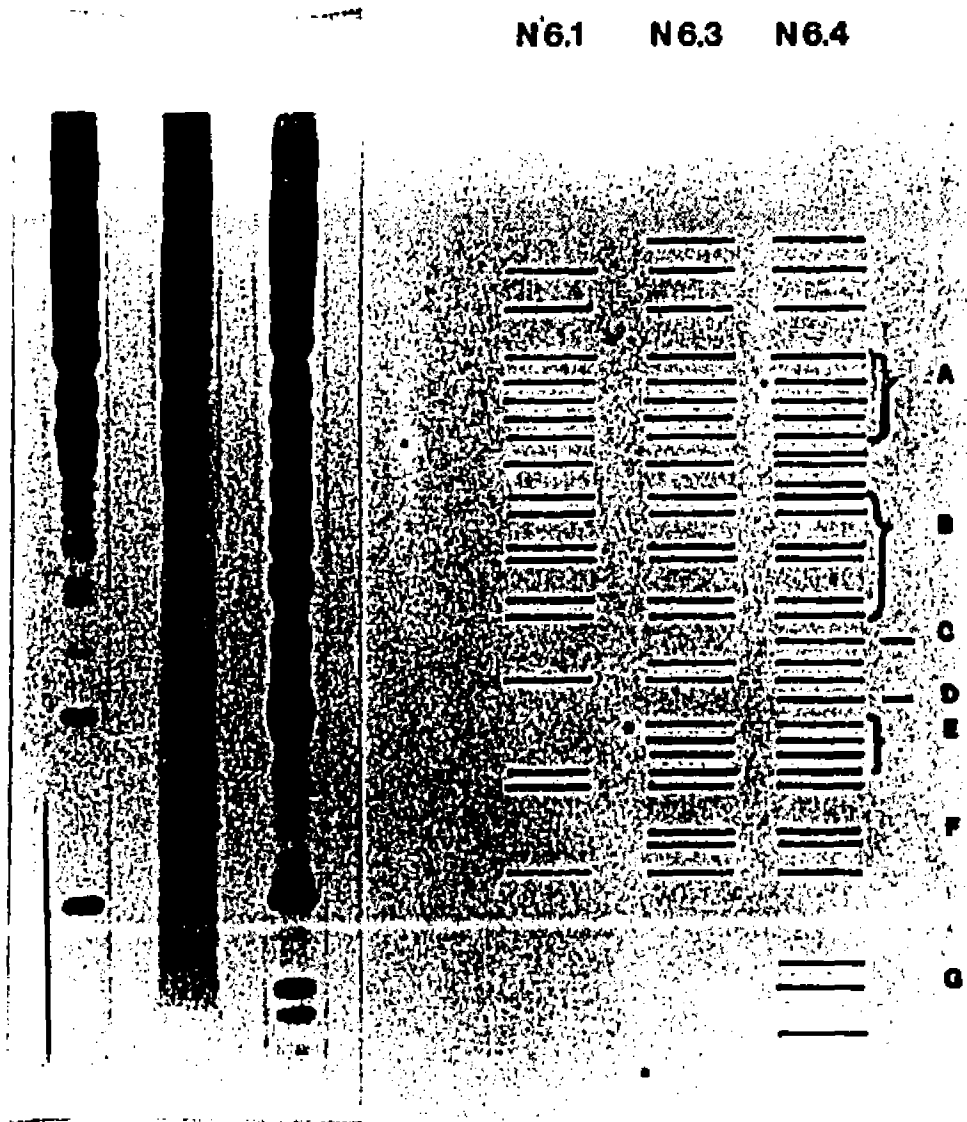
showed that the L2Hs family is polymorphic within the human population; no two restriction patterns were identical (Figure 11, panels A, B and C). There is no identifiable repeat pattern within a sample or between samples. Some bands are common within a series of samples digested with the same restriction enzyme but the relative intensity of hybridization to these segments varies from sample to sample. The extent of the L2Hs polymorphism is best illustrated in the hybridization of KpnI treated DNAs (Figure 11A). The 0.6 and 1.2 kb KpnI bands characteristic of Namalwa DNA are a variable trait. Some samples contain more 0.6 and 1.2 kb L2Hs segments than others. The data show that the organization of L2Hs DNAs is not constant between genomes.

Partial Sequence of the N6.1 Insert

The partial sequence of the N6.1 insert (Figure 12) confirmed the presence of a large number of RsaI restriction sites. Fourteen RsaI sites were identified within the first 420 bp of the N6.1 insert. The majority of the RsaI sites are spaced by 16 to 42 bp. The largest RsaI fragment expected from the data would be 80 bp. Other restriction sites identified by computer analysis are listed in Table 4.

Ten regions of alternating purine-pyrimidine or pyrimidine-purine occur within the sequence (Table 5). These regions, in conditions of high ionic strength, can

Figure 11. L2Hs Genomic Polymorphisms Between Individuals.
Namalwa DNA (lane 13) and the DNA from 12 different
peripheral blood lymphocyte samples were digested with KpnI
(panel A), AccI (panel B) or AccI+KpnI (panel C). The
digested DNAs were electrophoresed through 0.8% agarose
gels before transfer by Southern blotting to Gene Screen.
The filter was then probed with nick-translated N6.1. The
markers on the left of the panels represent sizes in kb.

FIGURE

A

1.2 kb

0.6

B

1.2 kb

0.6

C

1.2 kb

0.6
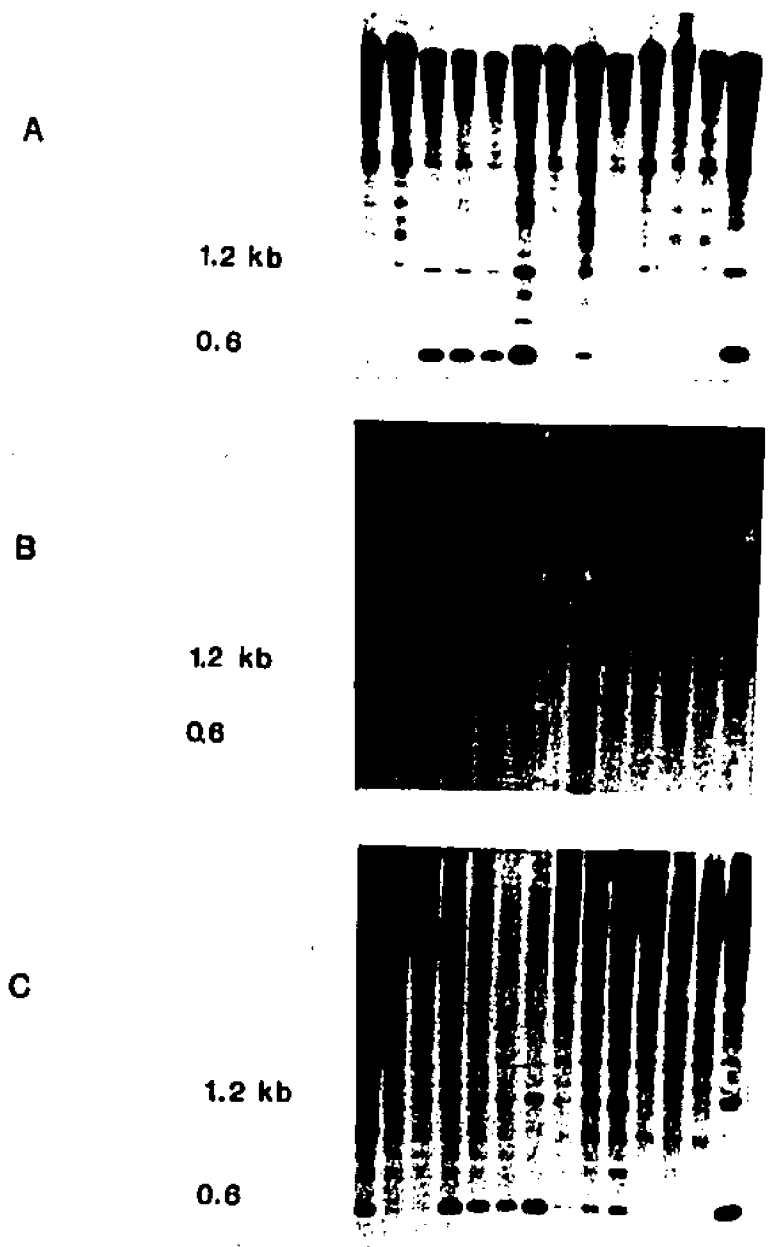
Figure 12. Partial Sequence of N6.1. The dideoxynucleotide sequencing method was employed to determine the sequence of the N6.1 insert. The 14 RsaI restriction sites are indicated by underlining. The majority of fragments resulting from a RsaI digest would be between 16 and 42 bp in length. Based on this sequencing analysis the largest RsaI fragment would be 80 bp in length. The numbers found above the sequence mark the position of every tenth nucleotide.

```
          10          20          30          40          50
AAACT ATACA TTATA TACTG TACAT AAAAT ATGAA ATTAC ATCAA ATATA

          60          70          80          90         100
TATTA TATTA GGTAC ATAAA AAATG AAAGT ACATC AAATA TAGAT TATAT

         110         120         130         140         150
ACTGT ACATA AAATA TCAAA GTACC CAAAT ATATA TCATA TACTG TACAT

         160         170         180         190         200
GAAAT ATCAA AGTTC ACAAA CTATA TATTA TATAC TGTAC ATAAA ATATC

         210         220         230         240         250
AAAGT ACCCA AACTA TACAT TATAT ACTGT ACATA AAAGA TGAAA TTACA

         260         270         280         290         300
TCAAA CATAT ATTTT ATTAG GTACG TAAAA TATGA AAGTA CATCA AATAT

         310         320         330         340         350
AGATT ATATG TGTAC ATAAA ATACA AAGTA CCCAA ATATA TATTT TATAC

         360         370         380         390         400
TCTAC TGAAA TATCA AGTTC TACAC TATAT ATTAT ATACT GACAC ATAAA

         410         420
TATCA AGTAC CAGGT ATATA TTCA
```

Table 4

Restriction Sites in N6.1 Insert DNA Identified

by Computer Analysis[a]

Enzyme	Recognition Sequence	Site Location[b]	Cleavage Site[c]
AtuII ApyI	CC / ($^{A}_{T}$)GG	410	411
MaeII	A / CGT	273	273
NlaIII	CATG /	148	151
ScrFI	CC / NGG	410	411
SnaBI	TAC / GTA	272	274
Tth111II	CAAPuCA(N)$_{11}$ /	252	268

[a] RsaI sites are indicated by underlining in Figure 12 and are not included in this Table.

[b] The position of the first base in the recognition sequence is indicated.

[c] The enzymes cleave on the right side of the base position indicated by the "/".

undergo a transition from a B-DNA structure to a Z-DNA form. Interestingly, a close spatial relationship exists between the RsaI sites and the potential Z-DNA regions. Within the first 424 nucleotides of the N6.1 insert, 7 of the 10 possible Z-DNA forming regions identified contain RsaI restriction sites.

The probability of a functional translation product from an N6.1 transcript is not high; of the six possible reading frames, only reading frame 2 contains an open reading frame (ORF) coding region with a good prospect of being translated (Figure 13 and Table 6). An uninterrupted ORF of 80 codons extends from position 77 through 316 in the sequence. The frame's only initiation codon occurs at position 149, suggesting a maximum coding capacity of 55 codons, representing a protein of 55 amino acid.

TABLE 5

Potential Regions for Z-DNA Structure Formation

In N6.1 Insert DNA

Size of Region (bp)	Location in the Sequence From	To	Base Sequence
8	19	26	TG TA CA TA
8	46	53	A TA TA TA T
8	103	110	TG TA CA TA
8	129	136	A TA TA TA T
8	144	151	TG TA CA TG
8	186	193	TG TA CA TA
8	228	235	TG TA CA TA
8	255	262	A CA TA TA T
14	305	318	TA TA TG TG TA CA TA
8	336	343	A TA TA TA T
8	414	421	G TA TA TA T

Figure 13. Possible Translations of the N6.1 Partial Sequence. Reading frames 1, 2 and 3 are translations of the experimentally derived sequence (Figure 12), while reading frames 4, 5, and 6 are translations of the computer-generated complementary strand. Stop codons are represented as ... . Each of the six reading frames (RF) is indicated to the left of the translated sequence. Both N6.1 strands may code only for an oligopeptides due to the frequency of stop codons. A search of all reading frames failed to identify potential transcripts in which valine or methionine was the first codon in an open reading frame.

A companion table, Table 6, has been compiled identifying the number of codons, the length of each ORF and the position of the first methionine or valine in each reading frame.

6 HOOC-Ser-Tyr-Met-Ile-Tyr-Gln-Val-Tyr-Phe-Ile-Ile-Asn-Cys~...~Ile-Tyr-Ile-Ile-Asn-Pro-Val-Tyr-NH₂
5 HOOC-Val-Ile-Cys~...~-Ile-Ser-Tyr-Met-Phe-Tyr-Ser-Ile-Val-Asp-Phe-Ile~...~-Ile-Leu-Tyr-Met-NH₂
4 HOOC-Phe-...~-Val-Asn-Tyr-Val-Thr-Cys-Leu-His-Phe-...~-Met-Leu-Tyr-Ile-Asn- Tyr~...~-Thr-Cys-NH₂
   5'-AAA CTA TAC ATT ATA TAC TGT ACA ATT ATG AAA TTA CAT CAA ATA TAT ATA TTA GGT ACA TA
      1::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
   3'-TTT GAT ATG TAA TAT ATG ACA TGA ATT TTA TAC TTT AAT GTA GTT TAT ATA TAA TAT CCA TGT AT
1 H₂N-Lys-Leu-Tyr-Ile-Ile-Tyr-Cys-Thr~...~-Asn-Met-Lys-Leu-His-Gln-Ile-Tyr-Ile-Gly-Thr-COOH
2 H₂N-Asn-Tyr-Thr-Leu-Tyr-Thr-Val-His-Lys-Ile~...~-Asn-Tyr-Ile-Lys-Tyr-Ile-Leu-Tyr~...~-Val-His-COOH
3 H₂N-Thr-Ile-His-Tyr-Ile-Leu-Tyr-Ile-Ser-Asn-Ile-Tyr-Tyr-Ile-Arg-Tyr-Ile-COOH

6 HOOC-Phe-Ile-Phe-Thr-Cys~...~-Ile-Tyr-Ile-Ile-Tyr-Gln-Val-Tyr-Phe-Ile-Leu-Thr-Gly-Leu-Tyr-Ile-NH₂
5 HOOC-Phe-Tyr-Ser-Leu-Val-Asp-Phe-Ile-Ser~...~-Ile-Ser-Tyr-Met-Phe-Tyr~...~-Leu-Val-Trp-Tyr-Ile-NH₂
4 HOOC-Leu-Ile-His-Phe-Tyr-Met-Leu-Tyr-Leu-Asn-Tyr-Val-Thr-Cys-Leu-Ile-Asp-Phe- Tyr-Gly-Phe-Ile-NH₂
   5'-TAA AAT ATG AAA GTA CAT CAA ATA TAG ATT ATA TAC TGT ACA ATT ATA TAC AAA GTA CCC AAA GTA TAT AT
      67::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
   3'-ATT TTA TAC TTT CAT GTA GTT TAT ATC TAA TAT ATG ACA TGT ATT TTA TAG TTT CAT GGG TTT ATA TA
1 H₂N-...-Asn-Met-Lys-Val-His-Gln-Ile~...~-Ile-Tyr-Cys-Thr~...~-Asn-Ile-Lys-Val-Pro-Lys-Tyr-COOH
2 H₂N-Lys-Tyr-Ile~...~-Lys-Tyr-Ile-Lys-Tyr-Arg-Leu-Tyr-Thr-Val-His-Lys-Ile-Ser-Lys-Tyr-Pro-Asn-Ile-COOH
3 H₂N-Lys-Tyr-Glu-Ser-Thr-Asn-Ile-Tyr-Tyr-Ile-Leu-Tyr-Ile-Gln-Ser-Thr-Gln-Ile-Tyr-COOH

6 HOOC-Asp-Tyr-Val-Thr-Cys-Ser-Ile-Asp-Phe-Asn-Val-Phe~...~-Ile-Asn-Tyr-Val-Thr-Cys-Leu-Ile-Asp-NH₂
5 HOOC-Ile-Met-Tyr-Gln-Val-His-Phe-Ile-Leu-Thr~...~-Leu-Ser-Tyr-Ile-Tyr-Gln-Val-Tyr-Phe-Ile-NH₂
4 HOOC-Tyr~...~-Ile-Ser-Tyr-Met-Phe-Tyr~...~-Leu-Glu-Cys-Val-Ile-Tyr~...~-Ile-Ser-Tyr-Met-Phe-Tyr-NH₂
   5'-ATA TCA TAT ACT GTA CAT GAA ATA TCA AAG TTC ACA AAG TTC ACA AAC TAT GTA CAT GTA CAT AAA ATA TC
      133::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::
   3'-TAT AGT ATA TGA CAT GTA CTT TAT AGT TTC AAG TGT TTG ATA TAT AAT ATA TGA CAT GTA TTT TAT AG
1 H₂N-Ile-Ser-Tyr-Thr-Val-His-Glu-Ile-Ser-Lys-Phe-Thr-Asn-Tyr-Ile-Leu-Tyr-Ile-Leu-Ile-COOH
2 H₂N-Tyr-His-Ile-Leu-Tyr-Met-Lys-Tyr-Gln-Ser-Ser-Gln-Thr-Tyr-Ile-Leu-Tyr-Ile-Lys-Lys-Ile-COOH
3 H₂N-Ile-Ile-Tyr-Cys-Thr~...~-Asn-Ile-Lys-Val-His-Lys-Tyr-Ile-Leu-Tyr-Ile-Cys-Thr~...~-Asn-Ile-COOH

6 HOOC-Phe-Tyr-Gly-Phe-...-Val-Asn-Tyr-Val-Thr-Cys-Leu-Leu-His-Phe-...-Met-Leu-Ile-Cys-Ile-Asn-...-NH$_2$

5 HOOC-Leu-Phe-Gly-Ser-Tyr-Met-Ile-Tyr-Gln-Val-Tyr-Phe-Ile-Phe-Asn-Cys-...-Val-Tyr-Ile-Ile-Lys-NH$_2$

4 HOOC-...-Leu-Val-Trp-Val-Ile-Cys-...-Ile-Ser-Tyr-Met-Phe-Ser-Ser-Ile-Val-Asp-Phe-Met-Tyr-Lys-NH$_2$

5'-TCA AAG TAC CCA AAC TTA TAT ACA TTA TAT ACT GTA CAT AAA AGA TGA CAT TAC ATC AAA CAT ATA TTT TA

199::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::

3'-AGT TTC ATG GGT TTG ATA TGT AAT ATA TGA CAT GTA TTT TCT ACT GTA ATG TAG TTT GTA TAT AAA AT

1 H$_2$N-Ser-Lys-Tyr-Pro-Asn-Tyr-Thr-Leu-Tyr-Thr-Val-His-Lys-Arg-...-Asn-Tyr-Ile-Lys-His-Ile-Phe-COOH

2 H$_2$N-Gln-Ser-Thr-Gln-Thr-Ile-His-Tyr-Ile-Leu-Tyr-Ile-Leu-Lys-Lys-Asp-Glu-Ile-Tyr-Ser-Asn-Ile-Tyr-Phe-COOH

3 H$_2$N-Lys-Val-Pro-Lys-Leu-Tyr-Ile-Tyr-Ile-Tyr-Cys-Thr-...-Lys-Met-Lys-Leu-His-Gln-Thr-Tyr-Ile-Leu-COOH

6 HOOC-...-Thr-Arg-Leu-Ile-His-Phe-Tyr-Met-Leu-Tyr-Leu-Asn-Tyr-Arg-Tyr-Met-Phe-Tyr-Leu-Thr-Gly-NH$_2$

5 HOOC-Asn-Pro-Val-Tyr-Phe-Ile-Phe-Thr-Phe-Thr-Cys-...-Ile-Tyr-Ile-Ile-Asp-Thr-Cys-Leu-Ile-Cys-Leu-Val-NH$_2$

4 HOOC-Ile-Leu-Tyr-Thr-Phe-Tyr-Ser-Leu-Val-Asp-Phe-Ile-Ser-...-Ile-Gln-Val-Tyr-Phe-Val-Phe-Tyr-NH$_2$

5'-TAT TAG GTA CGT AAA ATA TGA AAG TAC ATC AAA TAT AGA TTA TAT CTG TAC ATA AAA TAC AAA GTA CC

265::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: :::

31-ATA ATC CAT GCA TTT TAT ACT TTC ATG TAG TTT ATA TCT AAT ATA GTC ATG TAT TTT ATG TTT GAT GG

1 H$_2$N-Tyr-...-Val-Arg-Lys-Ile-...-Lys-Tyr-Ile-Lys-Tyr-Arg-Leu-Tyr-Ile-Lys-Tyr-Lys-Tyr-Val-COOH

2 H$_2$N-Ile-Arg-Tyr-Val-Lys-Tyr-Glu-Ser-Thr-Ser-Asn-Ile-Asp-Ile-Cys-Thr-...-Asn-Thr-Lys-Tyr-COOH

3 H$_2$N-Leu-Gly-Thr-...-Asn-Met-Lys-Val-His-Gln-Ile-...-Ile-Ile-Ser-Val-His-Lys-Ile-Gln-Ser-Thr-COOH

6 HOOC-Leu-Tyr-Ile-Asn-...-Val-Arg-Ser-Phe-Tyr-...-Thr-Arg-Cys-...-Ile-Asn-Tyr-Val-Ser-Val-Tyr-NH$_2$

5 HOOC-Trp-Ile-Tyr-Ile-Lys-Tyr-Ile-Glu-Val-Ser-Ile-Asp-Leu-Val-Ser-Tyr-Ile-Tyr-Gln-Cys-Met-NH$_2$

4 HOOC-Gly-Phe-Ile-Tyr-Lys-Ile-Ser-...-Gln-Phe-Ile-Leu-Asn-...-Val-Ile-Tyr-...-Ile-Ser-Val-Cys-NH$_2$

```
5'-CCC AAA TTT ATA TTT TAT ACT CTA CTG AAA TTT CAA GTT CTA CAC TAT ATA TTA TAT ACT GAC ACA TA
332::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::: ::
3'-GGG TTT ATA TAT AAA ATA TGA GAT GAC TTT AAA GTT CAA GAT GTG ATA TAT AAT ATA TGA CTG TGT AT
```

1 H$_2$N-Pro-Lys-Tyr-Ile-Phe-Thr-Leu-Phe-Tyr-Ile-Leu-Lys-Tyr-Gln-Val-Leu-His-Tyr-Ile-Leu-Tyr-Thr-Asp-Thr-COOH

2 H$_2$N-Pro-Asn-Ile-Tyr-Phe-Ile-Tyr-...-Asn-Ile-Lys-Phe-Tyr-Thr-Ile-Tyr-Tyr-Ile-Leu-Tyr-Thr-His-COOH

3 H$_2$N-Gln-Ile-Tyr-Ile-Leu-Tyr-...-Thr-Ile-Leu-Tyr-Ser-Thr-Ser-Thr-Ile-Ser-Val-Tyr-Ile-Tyr-...-His-COOH

6 HOOC-Ile-Asp-Leu-Val-Leu-Tyr-Ile-Asn-NH$_2$

5 HOOC-Phe-Ile-Leu-Tyr-Trp-Thr-Tyr-Ile-...-NH$_2$

4 HOOC-Leu-Tyr-...-Thr-Gly-Pro-Ile-Tyr-Glu-NH$_2$

```
5'-TAA ATA TCA AGT ACC AGG TAT ATA TTC A
398::: ::: ::: ::: ::: :
3'-ATT TAT AGT TCA TGG TCC ATA TAT AAG T
```

1 H$_2$N-...-Ile-Ser-Ser-Thr-Arg-Tyr-Ile-Phe-COOH

2 H$_2$N-Lys-Tyr-Gln-Val-Pro-Gly-Ile-Tyr-Ser-COOH

3 H$_2$N-Asn-Ile-Lys-Tyr-Gln-Val-Tyr-Ile-COOH

TABLE 6

Open Reading Frames Identified in N6.1 Insert DNA

Reading Frame	Number of Codons	Position From	To	Position of First Met	Val
1	8	1	24	—	—
	13	28	66	31	—
	7	70	90	73	—
	5	94	108	—	—
	43	112	240	—	—
	8	244	267	—	—
	4	271	282	—	—
	37	286	396	—	—
	8	400	423	—	—
2	10	2	31	—	—
	8	35	58	—	—
	4	62	73	—	—
	80[a]	77	316	149	—
	12	320	355	—	—
	22	359	424	—	—
3	49	3	149	—	—
	13	153	191	—	—
	13	195	233	—	—
	13	237	275	240	—
	7	279	299	282	—
	29	303	389	—	—
	10	393	422	—	—
4	6	423	406	—	—
	6	402	385	—	—
	3	381	373	—	375
	5	369	355	—	—
	15	351	307	—	—
	33	300	202	258	—
	6	198	181	192	—
	6	177	160	—	—

TABLE 6 (Cont'd)

Open Reading Frames Identified in N6.1 Insert DNA

Reading Frame	Number of Codons	Position From	Position To	Position of First Met	Position of First Val
4	6	156	139	150	-
	25	135	61	84	-
	6	57	40	42	-
	10	36	7	-	-
5	42	421	296	397	-
	13	292	254	-	-
	28	256	167	220	-
	15	163	119	139	-
	6	115	98	109	-
	13	94	56	67	-
	13	52	14	25	-
	3	9	1	-	-
6	15	422	378	-	395
	3	374	366	-	-
	5	362	348	-	-
	25	344	270	317	-
	5	263	249	251	-
	10	245	216	-	-
	13	212	174	-	-
	28	170	87	-	167
	13	83	45	-	-
	13	40	2	11	-

Reading frames 1, 2 and 3 represent codons starting at nucleotide positions 1, 2 or 3 from 5' end of the DNA.

Reading frames 4, 5 and 6 represent codons on the complementary strand starting at nucleotide positions 1, 2 or 3 from its 5' end.

[a]Indicates the longest ORF observed.

CHAPTER 4

Discussion

Human DNA digested with the restriction enzyme HindIII,
electrophoretically fractionated and stained with EBr,
reveals a series of bands upon exposure to UV light. The
periodicity of these bands is approximately 170 bp. A
similar, but more intense pattern can be identified in gel
electrophoresis of AGM DNA treated with HindIII, indicative
of the tandemly repetitive alpha DNA which comprises 20-25%
of the genome (Maio 1971). Hybridization of AGM alpha
probes to human DNAs showed that constituents of the probe
are evolutionarily conserved (Musich et al. 1980).

By constructing a plasmid library of human HindIII DNAs
and screening with an AGM genomic alpha probe, seven clones
containing evolutionarily conserved inserts were identified.
The seven recombinant clones, PuHu7, 10, 21, 25, 26, 31 and
32, were further screened with a cloned AGM alpha probe but
no homology was detected. Hybridization of the PuHu clones
to human DNA digested with HindIII, EcoRI, AluI and XbaI
demonstrated that the HindIII recombinant inserts belong to
repetitive DNA families. The autoradiographic patterns
showed that for four of the clones, PuHu 10, 25, 31 and 32,
hydridization represented interspersed repetitive sequences.
The other cloned HindIII segments studied (PuHu7 and
PuHu26) produced multimeric hybridization patterns

indicative of repetitive DNA families with tandem genomic organizations.

PuHu7: An XbaI Family Variant

The insert contained within PuHu7 is approximately 340 bp long. A comparison of the genomic hybridization patterns of PuHu7 and a cloned XbaI family descriptor, pE1 (Gray et al 1985), showed similar but non-identical patterns. The differences in the hybridization patterns show that the XbaI family is composed of at least two variants and the nucleotide sequence between these variants were sufficiently different to identify distinct genomic organizations in the HindIII digests. Examination of the genomic/HindIII hybridizations shows PuHu7 homologous sequences at 0.17, 0.34 kb and high molecular weight regions. While the pE1 probe identifies these segments, there was also intense hybridization with segments of 1.0 and 1.6 kb. Two possible organizations of PuHu7 sequences can explain the differences between PuHu7 and pE1 genomic/HindIII hybridizations. The first organizational scheme requires fewer HindIII restriction sites within PuHu7 variants, hence fewer hybridizable segments in the 1.0 and 1.6 kb range. An alternate explanation requires a high density of adjacent HindIII restriction sites which, upon cleavage, would produce fragments of 0.17 and 0.34 kb too small to be bound to the hybridization matrix and, hence, non-detectable by

hybridization.   Further experimentation is required to
determine which of the possibilities is most likely.

Characterization of the PuHu26 Repetitive DNA Family

The PuHu26 family of repetitive DNAs has not been
previously described. The descriptor of this family is
PuHu26, which contains an insert of approximately 600 bp.
The tandem organization of this family was revealed upon
hybridization of human genomic DNA digested with the
restriction enzymes EcoRI and HindIII. The homologous
segments identified in these genomic digests reflected a
restriction site periodicity and repeat length of
approximately 170 bp. Spacings longer than the unit length
reflected the absence of internal restriction sites in
adjacent repeats. The intensity of hybridization to the
EcoRI segments was much less than to the HindIII segments.
This suggests that a subpopulation of the PuHu26 family
contains a restriction site periodicity for EcoRI. The size
reduction of PuHu26 homologs by AluI digestion showed that
the PuHu26 family contains a high density of AluI
restriction sites. However, a general hybridization
background revealed that these sites are irregularly spaced
in the PuHu26 arrays. Comparisons of these hybridization
patterns with that of cloned representatives of the EcoRI
dimer and the XbaI families distinguished the PuHu26 family
from either of these.

Classification of the PuHu26 Family

In the literature, the XbaI and EcoRI dimer families are reported as human alphoid DNAs. The term "alphoid" is used to indicate a sequence relationship to the archetypal AGM alpha repetitive DNA family, even though sequence analysis has shown a homology of only 65% for the EcoRI dimer (Manuelidis 1978)and 68% for the XbaI clone pEl (Gray et al. 1985). Slot-blot hybridizations with a PuHu26 probe were used to test its homology to cloned EcoRI dimer, XbaI and AGM alpha repetitive DNAs. No detectable homology was found.

A number of tandemly repetitive sequences have been reported (Table 7). In order to discriminate among these tandemly repetive DNA families and the PuHu26 family, one of the following two criteria were used: 1) Differences in the reported genomic HindIII or EcoRI hybridization patterns under highly stringent conditions (requiring at least an 80% homology) or 2) a high degree of sequence homology to the EcoRI alphoid family (see Table 7). With these criteria it was possible to exclude the PuHu26 DNA from each of the families studied. By varying the temperature of hybridizations from 65°C, 60°C and 55°C, Musich et al. (1980) showed that, the human alphoid DNAs consist of multiple families. Reducing the temperature of PuHu26 hybridizations may reveal homologies between the PuHu26 family and other tandemly repetitive DNAs.

Possible Roles for Tandemly Repetitive DNAs

A variety of functions for tandemly repeated DNAs have been postulated (Bostock 1983). They could serve as structural elements required for chromatin condensation or as possible transcription regulation elements. The available data are most consistent with the idea that some centromeric repetitive DNAs provide recognition sites during chromosome pairings. Lee and Singer (1982) identified distinct subsets of alpha repeats organized in specific domains on an AGM chromosome. Recently, specific subpopulations of human alphoid DNAs have been found on chromosomes 7, 13, 17, 21, 22, X and Y (McDermid et al. 1986; Jorgensen et al. 1986, 1987; Wayne et al. 1987; Willard et al. 1986). Since the centromere governs meiotic segregation of chromosomes, it is conceivable that uniquely organized alphoid blocks may form the basis for recognition between individual chromosome homologues during chromosome pairings, potentially affecting recombination between the homologues (Bostock 1983).

Interspersed Repetitive Cloned Sequences

Recombinant DNAs in PuHu10, 25, 31 and 32 were designated as interspersed repetitive sequences as a result of their genomic hybridization patterns. Additional studies of these cloned sequences were not performed, since a more

Table 7

Homology Relationships Between Primate

Repetitive DNAs and PuHu26 DNA

Segment	Difference from PuHu26	Reference
Alphoid DNAs		
AGM	No homology	Musich et al. 1980
Baboon[*]	"	Donehower et al. 1980
Rhesus[*]	"	Pike et al. 1986
Human		
p17H8	Different genomic EcoRI pattern	Willard et al. 1986
DXZ1	" " EcoRI "	Willard 1985
pMGB7	" " HindIII "	Wayne et al. 1987
p22/1:0.73	" " HindIII "	McDermid et al. 1986
p82H	" " HindIII "	Mitchell et al. 1985
Sau3A family	" " EcoRI "	Kiyama et al. 1986
XbaI family	" " HindIII "	Gray et al. 1985
EcoRI family	" " HindIII "	Manuelidis 1976
SINE families		
L1 family	Different genomic organization	Singer et al. 1982b
L2Hs family	" " "	Musich and Dykes 1986
Hind 1.9	Not tandem	Manuelidis and Biro 1984
SINE family		
Alu family	Different genomic organization	Houck et al. 1979
Mini satellite		
Spc-HeLa	Heterologous DNAs	Jones and Potter 1985

* Due to homologies of greater than 80% between AGM alpha and the cloned alphoids of the baboon and rhesus, no homology would be detected between these DNAs and PuHu26 with the hybridizations conditions used.

interesting repetitive sequence, the L2Hs DNA, had been isolated.

The L2Hs Family

In an analysis of sequence variation within the primate KpnI families, Musich and Dykes (1986) identified a distinct and previously undescribed KpnI LINE family, the L2Hs. Recombinant clones N6.1, N6.3 and N6.4 were identified as sequence variants of this family and were used to study the genomic organization of L2Hs family members.

Hybridization analysis of restriction digests showed that one or possibly both inserts contained in N6.3, the single insert in N6.1 and one of the three inserts of N6.4 were homologous. These data also showed that the homologous inserts were each 0.6 kb long and contained a large number of RsaI restriction sites. Serial digests with HinfI, HaeIII, AluI, MspI and MboII revealed that genomic L2Hs DNAs contain very few restriction sites for these enzymes which frequently cleave human DNA. Some L2Hs homologous DNA in these digests was reduced to segments less than 2.0 kb, but the majority remained high in molecular weight. These results corroborate the restriction profiles of the N6 cloned segments which were not cleaved by these enzymes, suggesting that the primary sequences of L2Hs genomic DNAs are similar to those of the N6 clones. Data from the serial digests argue against a completely random, single copy interspersion of the L2Hs DNAs throughout the genome. If

L2Hs sequences were interspersed in such a fashion, the serial digests and subsequent hybridizations would have recorded the progressive release of L2Hs DNAs to smaller fragments due to the cleavage of adjacent non-L2Hs sequences. Furthermore, the hybridization data show that the pattern of insensitivity to these restriction enzymes is conserved between three individual human genomes.

L2Hs Organization in Namalwa DNA. Partial KpnI digests of Namalwa DNA describe at least two populations of genomic L2Hs sequences, one KpnI-sensitive, another KpnI-resistant. The data suggests two possible organizations for the KpnI sensitive population, as illustrated in the diagram below. Diagram A shows N6 segments (----) arranged as a cluster, each containing five tandemly arranged 0.6 kb KpnI sequences. Alternatively, diagram B shows several L2Hs sequences interspersed with non-L2Hs DNA (xxxx) which is also cleaved at a 0.6 kb spacing by KpnI to produce intermediate segments noted in the partial digests (see Figure 8, pg 56).

A   /----/----/----/----/----/

B   /----/xxxx/---/xxxx/----/

The conversion of high molecular weight DNA to long segments in AccI+KpnI double digests suggests that the KpnI-resistant population can be cleaved into resolvable segments by AccI digestion. The KpnI-resistant population, as described by AccI and AccI+KpnI digests, does not present a simple

polymeric pattern and, therefore, suggest a possible organization similar to the KpnI-sensitive population.

L2Hs Polymorphisms. Hybridizations of lymphocyte DNAs from different individuals with an N6.1 probe identified both qualitative and quantitative polymorphisms. A qualitative polymorphism was identified with KpnI and AccI restriction enzyme digests. Among individuals, each of the restriction enzymes produced an unique hybridization pattern suggesting different genomic organizations of N6.1-homologous sequences. When double digests of KpnI+AccI were performed, each individual again produced a unique N6.1 hybridization pattern. No two individuals shared identical hybridization patterns, although there were some bands in common. Varying hybridization intensities of L2Hs-homologous sequences in AccI+KpnI digests also identify quantitative polymorphisms. This polymorphism may result from quantititive differences in L2Hs family subsets. Unlike restriction fragment length polymorphisms (RFLPs) which are predominantly dimorphic and show simple hybridization patterns, L2Hs probes identify a large number of bands. This suggests that L2Hs polymorphisms are a result of a mechanism more complex than single base substitutions or deletions. Several repetitive sequences have been reported which possess an internal repeat motiff (IRM) and display quantitative polymorphic features (Gelderman 1975; Jefferys et al. 1985; Vassart et al. 1987;

Willard et al. 1986). These polymorphic DNAs display hybridization patterns as complex as the L2Hs patterns. Jefferys et al. (1985) have proposed a recombination mechanism by which quantitative and qualitative polymorphisms can be generated.

Sequence Analysis of N6.1. Sequence analysis of N6.1 indicates that the probability of a functional translational product from an N6.1 transcript is not high, even though RNAs homologous to L2Hs DNAs have been identified (P. R. Musich, personal communication). The N6.1 sequence contains stop and nonsense codons in all reading frames (see Figure 12, page 66). Translation of such mRNAs would result in oligopeptides. Only one of the readings frames analyzed was capable of producing a coding region longer than 40 bases. Valine and methionine have been identified as initiation codons in some eukaryotic proteins. A search of all reading frames failed to identify potential transcripts with valine or methionine as the first codon in the ORF (see Figure 13 page 70 and Table 6 page 71). This indicated that if transcripts of N6.1 have a function, it would probably be in some capacity other than for translation.

Although no identifiable protein is indicated by the N6.1 sequence, several intriguing sequence properties have been identified which may affect the three dimensional structure of this molecule, thereby elucidating potential functions.

The most familiar model of DNA is the "B" form (Zimmerman 1982). It is presented as a right-handed, double-stranded helix containing major and minor grooves. Calladine (1982) has suggested that steric hinderance of nearest neighbor purines on opposite strands determines the conformational change of DNA from the ideal "B" form and that alternating purine-pyrimidine tracts result in minor groove distortions. N6.1 contains 14 sites of alternating purines-pyrimides which also contain the RsaI sites sequence 5'-GTAC-3'. Crystallography of DNA molecules which consist of alternating purines and pyrimidines has shown that these sequences can assume a left-handed conformation and have been termed "Z" DNA (Arnott et al. 1980). A DNA crystal of a hexamer sequence containing the GTAC sequence site has been identified in a Z DNA geometry (Wang et al. 1984). N6.1 contains 11 regions of at least 8 bp which have an alternating purine-pyrimidine sequence (see Table 5, page 69). Seven of these regions contain RsaI restriction recognition sites. The large number of alternating purine-pyrimidine tracts suggests that the N6.1 insert is capable of assuming an altered geometry, which may be important for cellular functions other than coding for a protein.

The three-dimensional conformation of non-B form DNA is believed to have a significant role in recombinational events. The transition region where B-DNA becomes Z-DNA (B-Z junction) is partially unwound and sensitive to S1

nuclease (Kohwi-Shigematsu et al. 1987). This has prompted speculation that single-stranded B-Z junctions may be substrates for recombination enzymes; these enzymes generate free ends which may invade homologous duplexes (Nordheim et al. 1983, Treco and Arnheim 1986). The reported conformational changes in Z-DNA regions would provide a possible explanation for the high degree of L2Hs polymophism. As N6.1 contains 11 potential Z-DNA regions (see Table 5 page 69) and, if other L2Hs DNAs contain a similar number of such regions, there exists an increased probability for altered DNA structures suitable for recombination.

A 13 base internal repeat sequence (IRS) (5'-GTACATAAAATAT-3') is represented four times in N6.1. The presence of a repetitive sequence can promote exchange by increasing the likelihood that a pairing partner will encounter a homologous sequence. The four IRSs constitute 12% of N6.1. Four additional sequences which differ by one nucleotide from the IRS have been identified (Figure 14). If these variants are included in calculating the proportion of internal repeats composing N6.1, the value rises to 25%. Five of the eight IRSs are contiguous with potential Z-DNA regions and all IRSs contain an A-rich 3' region. The juxtaposition of the IRS to potential Z-DNA regions and the decreased B-DNA stability of A-T rich regions suggests that these sequences would be susceptible to strand separation,

Figure 14. Distribution of the 13 Base IRS and Four Variants within N6.1. Analysis of the partial N6.1 sequence identified four 13 base IRSs, 5'-GTACATAAAATAT-3'. The four variants of this sequence, each differing by one nucleotide, are identified by a lower case letter at the position(s) where they differ. The IRS and its variants are underlined. The four repeat sequences and the four variant repeats represent approximately 25% of the partial N6.1 sequence. The potential Z-DNA regions of the partial sequence are identified by dots over the sequence. Numbers above the nucleotide sequence mark the position of every 10th base.

```
        10              20      .....  30             40             50
AAACT ATACA TTATA TACTG TACAT AAAAT ATGAA ATTAC ATCAA ATATA


        60              70             80             90            100
TATTA TATTA GGTAC ATAAA ATATG AAAGT ACATC AAATA TAGAT TATAT


   .. ..110            120            130            140            150
ACTGT ACATA AAATA TCAAA GTACC CAAAT ATATA TCATA TACTG TACAT


       160            170            180         ..190 ...       200
GAAAT ATCAA AGTTC ACAAA CTATA TATTA TATAC TGTAC ATAAA ATATC


       210            220            230 .....  240            250
AAAGT ACCCA AACTA TACAT TATAT ACTGT ACATA AAAGA TGAAA TTACA


    ..260 ..          270            280            290            300
TCAAA CATAT ATTTT ATTAG GTACG TAAAA TATGA AAGTA CATCA AATAT


       310    ..... ..320            330         ..340 ...       350
AGATT ATATG TGTAC ATAAA ATACA AAGTA CCCAA ATATA TATTT TATAC


       360            370            380            390            400
TCTAC TGAAA TATCA AGTTC TACAC TATAT ATTAT ATACT GACAC ATAAA


       410            420
TATCA AGTAC CAGGT ATATA TTCA
```

allowing for a pairing between two non-contiguous single-stranded DNAs. Treco and Arnheim (1986) concluded that the number of B-Z junctions, rather than the total length of the potential Z-DNA forming sequences, may be more important in promoting recombinational exchange. Thus, the large number of B-Z junctions in the N6.1 sequence indicates that this molecule may have an increased susceptibility to recombinational mechanisms.

As Z-DNA may represent unique recognition signals for specific recombinational proteins, it may also participate in the organization of chromatin and the control of gene expression. The DNA of eukaryotes is organized into chromatin domains so that every 50-100 kb is attached to the scaffold protein of the nuclear matrix (reviewed in Chimera 1984). Hamada et al. (1982) identified a sequence of 50 bp which is represented every 50-100 kb in eukaryotes. This sequence was later shown to form Z-DNA structures (Nordheim and Rich 1980). The Z-DNA forming segment is long enough so that a B-DNA to Z-DNA conversion could change the superhelical structure of the entire domain (Rich et al. 1984). The change in domain superhelicity may then lead to enhancer-regulated transcription.

Sequences other than those which can assume Z-DNA conformations have been identified in chromatin organization. Cockerill et al. (1987) have shown that an AT-rich sequence in the immunoglobulin gene is a putative

binding site for nuclear matrix proteins. Avila et al. (1983) identified the sequences AAAA and TTTT as murine microtubule-associated protein (MAP) binding sites. N6.1 contains eight AAAA or TTTT sequences which may serve as potential binding sites for human MAPs (Figure 15). Of the nine sequences identified, seven are contiguous to regions of alternating purines and pyrimidines. Although a direct relationship between the AAAA sequences and potential Z DNA regions is not known, it is possible that a B-Z transition could regulate the binding of human MAPs.

Conclusions and Prospectus

Several families of repetitve DNAs were identified and characterized. The organization of a novel repetitive DNA family, PuHu26, was studied and it was shown to be organized in tandem arrays. Genomic hybridizations with a PuHu26 probe, identified HindIII and EcoRI restriction site periodicities and a repeat length of about 170 bp within the PuHu26 family. Comparative genomic hybridizations between PuHu26 and cloned representatives of the human EcoRI and XbaI families showed that the PuHu26 family is organized differently and, therefore, is distinct from either of these families. Additional genomic hybridizations with cloned representatives of these families under less stringent hybridization conditions may identify organizational features common to all three families. Such studies would

Figure 15. Positions of Potential Binding Sites for Microtubule-Associated Proteins. The sequences for putative human MAP are underlined. Z-DNA regions contiguous to the MAP binding sites have dots above the sequences. Numbers above the nucleotide sequence mark the position of every 10th base.

```
      10          20            30          40          50
AAACT ATACA TTATA TACTG TACAT AAAAT ATGAA ATTAC ATCAA ATATA

      60          70          80          90         100
TATTA TATTA GGTAC ATAAA ATATG AAAGT ACATC AAATA TAGAT TATAT

     110         120         130         140         150
ACTGT ACATA AAATA TCAAA GTACC CAAAT ATATA TCATA TACTG TACAT

     160         170         180           190         200
GAAAT ATCAA AGTTC ACAAA CTATA TATTA TATAC TGTAC ATAAA ATATC

     210         220         230           240         250
AAAGT ACCCA AACTA TACAT TATAT ACTGT ACATA AAAG TGAAA TTACA

       260        270         280         290         300
TCAAA CATAT ATTTT ATTAG GTACG TAAAA TATGA AAGTA CATCA AATAT

     310           320         330           340           350
AGATT ATATG TGTAC ATAAA ATACA AAGTA CCCAA ATATA TATTT TATAC

     360         370         380         390         400
TCTAC TGAAA TATCA AGTTC TACAC TATAT ATTAT ATACT GACAC ATAAA

     410         420
TATCA AGTAC CAGGT ATATA TTCA
```

allow a more complete understanding of the organization and possible interrelations among these tandemly repetitive DNAs.

The organization of several variants of a newly discovered LINE family, L2Hs, were also studied. These variants exhibit a clustered interspersed organization and are polymorphic. Sequence analysis of one of these variants, N6.1, identified several nucleotide regions which may assume a Z-DNA conformation. If an equivalent number of potential Z-DNA regions occur in other L2Hs DNAs, it may explain the highly polymorphic nature of this DNA family. In addition it may provide insights into the potential role of the L2Hs sequences in the regulation of transcription.

Currently, there exists a paucity of information detailing the effect of atypical DNA structures on eukaryotic gene expression. Burke et al. (1987) have reported a method which permits cloning of very long segments of DNA (100 kb or greater). The size of these cloned inserts is approximately that of eukaryotic transcriptional domains. The construction of such a gene bank would allow experimentation and determination of the effects sequences with potential Z-DNA regions, such as N6.1, have on gene expression. However, given the high densities of potential Z-DNA regions and IRSs, and the overall high AT content of the L2Hs DNAs, it may prove difficult to clone large segments of L2Hs-containing DNA.

BIBLIOGRAPHY

Antonarkis, S. E.; Boehm, C. D.; Giardin, P. J. V.; Karazian, H. H. 1982. Nonrandom association of polymorphic restriction sites in the beta-globin gene cluster. Proc. Natl. Acad. Sci. USA 79:137-141.

Arnott, S.; Chandrasekaran, R.; Birdsall, D. L.; Leslie, A. G. W.; Ratliff, R. L. 1980. Left-handed DNA helices. Nature 283:743-745.

Avila, J.; Montejo de Garcini, E.; Wandosell, F.; Villasante A.; Sogo, J. M.; Villanueva, N. 1983. Microtubule associated protein MAP$_2$ preferentially binds to dA/dT sequence present in moûse satellite DNA. EMBO J. 2:1229-1234.

Banner, D. B. 1982. Recovery of DNA fragments from gels by transfer to DEAE-Paper in an electrophoresis chamber. Anal. Biochem. 125:139-142.

Beridze, T. 1986. Satellite DNA p.77-91; Springer-Verlag, New York.

Bober ,E.; Singh, S.; Agarwal, D. P.; Goedde, H. W. 1986. Isolation of repetitive clones from human muscle cDNA library. Biosci. Report 6:633-638.

Bostock, C. 1983. A function of satellite DNA. In: DNA Makes RNA Makes Protein. T. Hunt, S. Prentis, J. Tooze (Eds.) p.111-117.

Botstein, D.; White, R. L.; Skolnick, M.; Davis, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Genet. 32:314-331.

Brunk, C. F.; Jones, K. C.; James, T. W. 1979. Assay for nanogram quantities of DNA in cellular homogenates. Anal. Biochem. 92:497-500.

Bullock, P.; Miller, J.; Botchan, M. 1986. Effects of poly[d(pGpT)·d(pApC)] and poly[d(pCpG)·d(pCpG)] repeats on homologous recombination in somatic cells. Molec. Cell. Biol. 6:3948-3953.

Burke, D. T.; Carle, G. F.; Olson, M. V. 1987. Cloning of large segments of exogeneous DNA into yeast by means of artificial chromosome vectors. Scence 236:806-812.

Calabretta, B.; Robberson, H.; Maizel, A. L.; Saunders, G. F. 1981. Messenger RNA in human cells contains sequences complementry to the Alu family repeated DNA. Proc. Natl. Acad. Sci. USA 78:6003-6007.

Callabretta, B.; Robberson, D. L.; Berrera-Saldana, H. A.
Lambrou, P. T.; Saunders, G. F. 1982. Genome instability
in a region of human DNA enriched in Alu repeat sequences.
Nature 295:219-225.

Calladine, C. R. 1982. Mechanics of sequence-dependent
stacking of bases in B-DNA. J. Mol. Biol. 19:122-133.

Chimera, J. 1984. Ph.D. Dissertation: Characterization of
KpnI Interspersed Repetitive DNA Sequences and Their
Association with the Nuclear Matrix, East Tennessee State
University, Johnson City, TN.

Cockerill, P. N.; Yuen, M.-H.; Garrard, W. T. 1987. The
enhancer of the immunoglobin heavy chain locus is flanked by
presumptive chromosomal loop anchorage elements. J. Mol.
Biol. 262:5394-5397.

Cooke, H. J.; Hindley, J. 1979. Cloning of human satellite
III DNA: Different components are on different chromosomes.
Nuc. Acids Res. 6:3177-3197.

Corneo, G.; Ginelli, E.; Polli, E. 1967. A satellite DNA
from human tissue. J. Mol. Biol. 23:619-622.

Corneo, G.; Ginelli, E.; Polli, E. 1970. Repeated
sequences in human DNA. J. Mol. Biol. 48:319-327.

Deininger, P. L.; Jolly, D. J.; Rubin, C.M.; Friesman, T.;
Schmid, C. W. 1981. Base sequence studies of 300
nucleotide renatured repeated human DNA clones. J. Mol.
Biol. 151:17-33.

Denhardt, D. T. 1966. A membrane filter technique for the
detection of complementry DNA. Biochem. Biophys. Res.
Commun. 23:641-646.

Donehower, L.; Furlong, C.; Gillespie, D.; Kurnit, D. 1980
DNA sequence of baboon highly repeated DNA: Evidence for
evolution by nonrandom unequal crossovers. Proc. Natl. Acad
Sci. USA 77:2129-2133.

Duggleby, R. G.; Kinns, H.; Road, J. 1981. A computer
program for determining the size of DNA restriction
fragments. Anal. Biochem. 110:49-55.

Feinberg, A.P.; Vogelstein, B. 1984. A technique for
radiolabeling DNA restriction endonuclease fragments to high
specific activity. Anal. Biochem. 137:266-269.

Furlong, N. B.; Marien, K.; Flook, B.; White, J. 1986. Characteristics of site variation among clones of the 340-base pair, tandemly repeated EcoRI family of human DNA. Biochem. Genet. 24:71-78.

Gelderman, H. 1975. Investigation on inheritance of quantitative characters in animals by gene markers. Meth. Theor. Appl. Genet. 46:300-319.

Goldfurb, M.; Shimizu, K.; Perucho, M.; Wigler, M. 1982. Isolation and characterization of a human transforming gene from T24 bladder carcinoma. Nature 296:404-409.

Gosden, J. R.; Mitchell, A. R.; Buckland, R. A.; Clayton, R. P.; Evans, H. J. 1975. The location of four human satellite DNAs on human chromosomes. Exp. Cell Res. 92:148-158.

Gray, K. M.; White, J. W.; Costanzi, C.; Gillespie, D.; Schroeder, W. T.; Calabretta, B.; Saunders, G. F. 1985. Recent amplification of an alpha satellite DNA in humans. Nucl. Acids Res. 13:521-535.

Hamada, H.; Petrino, M. G.; Kakunaga, T. 1982. A novel repeated element with Z-DNA forming potential is widely found in evolutionary diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA 79:6465-6469.

Hattori, M.; Hidaka, S.; Sakaki, Y. 1985. Sequence analysis of a KpnI family member near the 3'end of the beta globin gene. Nucl. Acids Res. 13:7813-7827.

Higgs, D. R.; Wainscoat, J. S.; Flint, J.; Hill, A. V. S.; Thein, S. L.; Nichols, R. D.; Teal, H.; Ayyub, H.; Peto, T. E. A.; Falusi, A. G.; Jarman, A. P.; Clegg, J. B.; Weatherall, D. J. 1986. Analysis of the human alpha globin gene cluster reveals a highly informative genetic locus. Proc. Natl. Acad. Sci. USA 83:516-5169.

Houck, C. M.; Rhinehart, F. P.; Schmid, C. W. 1979. Ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. 132:289-306.

Jefferys, A. J.; Wilson, V.; Thein, S. W. 1985. Hypervariable 'minisatellite' regions in human DNA. Nature 314:67-73.

Jelinek, W. R.; Schmid, C. W. 1892. Repetitive sequences in eukaryotic DNA and their expression. Ann. Rev. Biochem. 51:813-844.

Johnson, D. 1984. Improved technique utilizing non-fat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene Anal. Techn. 1:3-8.

Jones, R. S.; Potter, S. S. 1985. Characterization of cloned human alphoid satellite with an unusual monomeric construction: Evidence for enrichment in HeLa small polydispersed circular DNA. Nucl. Acids Res. 13:1027-1041.

Jorgensen, A. L.; Bostock, C. J.; Bak, A.L. 1986. Chromosome-specific subfamilies within human alphoid repetitive DNA. J. Mol. Biol. 187:185-196.

Jorgensen, A. L.; Bostock, C. J.; Bak, A. L. 1987. Homologous subfamilies of human alphoid repetitive DNA on different nucleous organizing chromosomes. Proc. Natl. Acad. Sci. USA 84:1075-1079.

Kiyama, R.; Hideki, M.; Oiahi, M. 1986. A repetitive DNA family (Sau3A family) in human chromosomes: Extrachromosomal DNA and DNA polymorphism. Proc. Natl. Acad. Sci. USA 83:4665-4669.

Kohwi-Shigematsu, T.; Manes, T.; Kohwi, Y. 1987. Unusual conformational effects exerted by Z-DNA upon its neighboring sequences. Proc. Natl. Acad. Sci. USA 84:2223-2227.

Kole, L. B.; Haynes, S. R.; Jelnick, W. R. 1983. Discrete and heterogeneous high molecular weight RNAs complementary to a long dispersed repeated family (a possible transposon) of human DNA. J. Mol. Biol. 165:257-286.

Lee, T. N. H.; Singer, M. F. 1982. Structural organization of alpha-satellite in a single monkey chromosome. J. Mol. Biol. 161:323-342.

Lowe, J. R. 1986. Transportable microcomputer programs for DNA and protein analysis. Fed. Proc. 45:1852a.

Maio, J. J. 1971. DNA strand reassociation and polyribonucleotide binding in the African greeen monkey, Cercopithecus althrops. J. Molec. Biol. 56:579-595.

Maniatis, A. M.; Fritsch, E. F.;Sambrook, J. 1982. Molecular Cloning. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Manuelidis, L. C. 1976. Repeating restriction fragments of human DNA. Nucl. Acids Res. 3:3036-3076.

Manuelidis, L. 1978. Chromosomal localization of complex and simple repeated human DNAs. Chromosoma (Berl.) 66:23-32.

Manuelidis, L.; Wu, J. C. 1978. Homology between human and simian repeated DNA. Nature 276:92-94.

Manuelidis, L.; Biro, P.A. 1982. Genomic representation of the HindIII 1.9 kb repeat DNA. Nucl. Acids Res. 10:3221-3239.

Manuelidis, L. 1984. Repeated DNA sequences and nuclear structure. In: Genome Evolution G. A. Dover and R. B. Flavel, (Eds), p. 263-285.

McDermid, H. E.; Duncan, A. M. V.; Higgins, M. J.; Hammerton, J. L. Rector, E.; Brasch, K. R.; White, B. N. 1986. Isolation and characterization of an alpha satellite repeat sequence from human chromosome 22. Chromosoma (Berl.) 94:228-234.

Messing, J.; Vieira, J. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19:269-276.

Mitchell, A. R.; Beauchamps, R. S.; Bostock, C. J. 1979. A study of sequence homologies in four satellite DNAs of man. J. Mol. Biol. 135:127-149.

Mitchell, A. R.; Gosden, J. R.; Miller, D. A. 1985. A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma (Berl.) 92:369-377.

Musich, P. R.; Brown, F. L.; Maio, J. J 1980. Highly repetitive component alpha and related alphoid DNAs in man and monkeys. Chromosoma (Berl.) 80:331-348.

Musich, P. R.; Dykes, R. J. 1986. A long interspersed (LINE) DNA exhibiting polymorphic patterns in human genomes. Proc. Natl. Acad. Sci. USA 83:4854-4858.

Nakamura, Y.; Leppert, M.; O'Connell, P.; Wolff, R.; Holm, T.; Culver, M.; Martin, C.; Fujimoto, E.; Hoff, M.; Kumlin, E.; White, R. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science 235:1616-1622.

Nordheim, A.; Rich, A. 1980. The sequence (dC-dA)·(dG-dT) forms left handed Z-DNA in negatively super-coiled plasmids Proc. Natl. Acad. Sci. USA 80:1821-25.

Nordheim, A.; Hao, W. M.; Wogan, G. N.; Rich, A. 1983. Salt-induced conversion of B-DNA to Z-DNA inhibited by aflotoxin-B1. Science 219:1434-36.

O'Connell, P.; Lathrop, G. M.; Law, M.; Leppert, M.; Nakamura, Y.; Hoff, M.; Kumlin, E.; Thomas, W.; Ballard, L.; Goodman, P.; Azen, E.; Sadler, J. E.; Cai, G. Y.; Lalouel, J.-M.; White, R. 1987. A primary genetic linkage map for human chromosome 12. Genomics 1:93-102.

Pike, L.M.; Carlisle, A.; Newell, C.; Hong, S. B.; Musich, P.R. 1986. Sequence and evolution of rhesus monkey alphoid DNA. J.Mol. Evol. 23:127-137.

Pritchett, R.; Pederson, M.; Keiff, E. 1976. Complexity of EBV homologous DNA in continuous lymphoblastoid cell lines. Virology 74:227-231.

Prosser, J; Frommer, M.; Paul, C.; Vincent, P. C. 1986. Sequence relationships of three human satellite DNAs. J. Mol. Biol. 187:145-155.

Rich, A.; Nordheim, A.; Wang, A. H. J. 1984. The chemistry of left-handed Z-DNA. Ann. Rev. Biochem. 53:791-846

Rotwein, P.; Yokoyama, S.; Didier, K.; Chirgwin, J. M. 1986. Genetic analysis of the hypervariable region flanking the human insulin gene. Am. J. Genet. 39:291-299.

Sakaki, L.; Kurata, N.; Miyake, T. and Saigo, K. 1983. Two-dimensional gel electrophoretic analysis of the HindIII 1.8 kb repetitive sequence family in the human genome. Gene 24:179-190.

Sanger, F.; Nicklen, S.; Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

Schleif, R. F.; Wensink, P. C. 1981. Practical Methods in Molecular Biology. Springer-Verlag, New York, p. 97.

Shafit-Zagardo, B.; Maio, J. J.; Brown, F. 1982. KpnI families of long interspersed sequences in mammalian genomes. Nucl. Acids Res. 10:3135:3193.

Shapiro, J. A. 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposible elements. Proc. Natl. Acad. Sci. USA 76:1933-1937.

Singer, M. F. 1982a. SINEs and LINEs: Highly repeated short and long interspersed sequences in mammalian genomes. Cell 28:433-434.

Singer, M. F.   1982b.   Highly repeated sequences in
mamalian genomes.   Inter. Rev. Cytol. 76:67-112.

Singer, M. F.; Skowronski, J.   1985.   Making sense out of
LINES: Long interspersed repeat sequences in mammalian
genomes.   Trends in Biochem. Sci. 3:119-121.

Skowronski, J.; Singer, M.F. 1985.   Expression if a
cytoplasmic LINE-1 transcript is regulated in a human
teratocarcinoma cell line. Proc. Natl. Acad. Sci. USA
82:6050-6054.

Southern, E. M.   1975.   Detection of specific sequences
among DNA fragments separated by gel electrophoresis.   J.
Mol. Biol. 79:503-517.

Taub, F.; Thompson, B. 1982.   An improved method for
preparing large arrays of bacterial colonies containing
plasmids for hybridization: In situ purification and stable
binding of DNA on filters.   Anal. Biochem. 126:222-230.

Tautz, D.; Renz, M.   1983.   An optimized freeze-squeeze
method for the recovery of DNA fragments from agarose gels.
Anal. Biochem. 132:14-19.

Treco, D.; Arnheim, N.   1986.   The evolutionarily conserved
repetitive sequence d(TG·AC) promotes reciprocal exchange
and generates unusual recombinant tetrads during yeast
meiosis.   Molec. Cell. Biol. 6:3934-3947.

Tracey, S.   1981.   Improved rapid methodology for the
isolation of nucleic acids from agarose gels.   Prep.
Biochem. 11:251-261.

Ullu, E.   1980.   The human Alu family of repeated DNA
sequences.   Trends in Biochem. Sci. 5:119-124.

Vassart, G.; Georges, M.; Monsieur, R.; Brocas, H.;
LeQuarre,S.; Christophe, D.   1987.   A sequence in M13 phage
detects hypervariable minisatellites in human and animal
DNA.   Science 235:683-684.

Volkin, E.; Cohen, W. E.   1954.   In: Methods of Biochemical
Analysis, Vol I. D. Glick, ed., Interscience, New York p.
283-305.

Wang, A. H.-J.; Hakoshima, T.; van der Marel, G.; van Boom,
J. H.; Rich, A.   1984.   AT base pairs are less stable than
GC base pairs in Z-DNA: The crystal structure of
d(m^{5}CGTAm^{5}CG).   Cell 37:321-331.

Wayne, J. S.; England, S. B.; Willard, H. F. 1987. Genomic organization of alpha satellite DNA on human chromosome 7: Evidence for two distinct alphoid domains on a single chromosome. Molec. Cell. Biol. 7:349-356.

Willard, H. F. 1985. Chromosome-specific organization of human alpha satellite DNA. Am. J. Hum. Genet. 37:524-532.

Willard, H. F.; Wayne, J. S.; Skolnick, M. H.; Schwartz, C. E.; Powers, V. E.; England, S. B. 1986. Detection of restriction fragment length polymorphism at the centromeres of human chromosomes by using chromosome-specific alpha satellite DNA probes: Implications for development of centromere-based genetic linkage maps. Proc. Natl. Acad. Sci. USA 83:5611-5615.

Willard, H. F.; Waye, J. S. 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. Trends in Genet. 3:192-198.

Wyman, A. R.; White, R. 1980. A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA 77:6754-6758.

Zimmerman, S. B. 1982. The three-dimensional structure of DNA. Ann. Rev. Bioch. 51:395-427.

VITA

ROBERT M. ROUDABUSH

Personal Data: Date of Birth: January 16, 1955
                  Place of Birth: Okinawa, Japan
                  Marital Status: Married

Education: Langley High School, McLean, VA
                  Virginia Polytechnic Institute; Biology
                      B. Sc., 1977
                  East Tennessee State University, Johnson
                      City, Tennessee; Microbiology, M.S. 1979
                  East Tennessee State University, Johnson
                      City, Tennesee; Biomedical Sciences,
                      Ph.D. 1988

Professional   Research Biochemist, Molecular Biology
Experience: Sigma Chemical Co., St. Louis MO. 1987-
                      present
                  Graduate Research Assistant, Dept. of
                      Biochemestry, East Tennessee State
                      University, 1982-1987
                  Research Assistant, Dept. of
                      Microbiology, East Tennessee State
                      University, 1980-1982

Publications: Herald, W.; Roudabush, R.; Pike, L. M.
                      1979 Pyrimidine Nucleotide Synthesis in
                      Activated Bovine Lymph Node T-cells,
                      Southeastern Regional Meeting American
                      Chemical Society
                  Roudabush, R.; Musich, P.R.M. 1986.
                      Variants of a Novel Human Line Sequence
                      Family, L2Hs. J. Cell Biol. 103(5):309a