



National Library of Canada
Collections Development Branch

Canadian Theses on
Microfiche Service

Bibliothèque nationale du Canada
Direction du développement des collections

Service des thèses canadiennes
sur microfiche

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

**THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED**

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

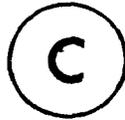
Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, R.S.C. 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

**LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE**

Repetitive Sequences in the Halobacterial Genome

by



Carmen Sapienza

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
at Dalhousie University

Department of Biochemistry
Dalhousie University
Halifax, Nova Scotia

Table of Contents

	<u>Page</u>
Abstract	i
Acknowledgements	ii
<u>Halobacterium halobium</u> Colour Plate.	vi
I. <u>Introduction</u>	1
II. <u>Materials and Methods</u>	
A. Sources of Strains.	4
B. Growth of strains	4
C. Purification of DNA	
1. for agarose gel-screening of recombinant plasmids	5
2. for nick-translation.	9
D. Restriction endonuclease digestions	9
E. Agarose gel electrophoresis and transfer of DNA to nitrocellulose	10
F. DNA-DNA hybridization	13
G. <u>In vitro</u> formation of recombinant molecules and transformation of <u>E. coli</u> JF1754.	14
H. <u>In vitro</u> ³² P-labelling of DNA	14
I. 5'-end labelling of restriction fragments	15
III. <u>Results and Discussion</u>	
A. <u>H. halobium</u> plasmid DNAs.	16
B. Identification of repeated sequences on the plasmid DNA of strain R1.	18
C. <u>H. halobium</u> contains more repeated sequences than <u>E. coli</u>	28
D. Repeated sequences in an <u>H. halobium</u> R1 genomic library	30
E. Repeated sequences in an <u>EcoRI</u> library of <u>H. halobium</u> NRC-1 genomic clones.	35
F. Repeated sequences in the genome of <u>H. volcanii</u>	41
G. Rearrangements affecting <u>H. halobium</u> repetitive sequences	43
H. Nucleotide sequence of an <u>H. volcanii</u> repeated element.	47
I. Unselected rearrangements affecting the position of repeated elements in <u>H. halobium</u>	53
J. Implications of genomic rearrangement at high frequency	68
K. Evolutionary conservation of repeated sequences among the Halobacteria.	82
L. Other archaeobacterial genomes	95
M. Organisms: How low do they go?	97
References	103

Abstract

The extremely halophilic Archaeobacterium, Halobacterium halobium, is shown to exhibit an unusual degree of genetic plasticity. The genome of H. halobium contains many repeated-sequence families which are arranged in both clustered and dispersed fashion on both plasmid and chromosomal DNAs. These repeated-sequence families are highly mobile within the genome and appear capable of both increase and decrease in copy number. Wild-type isolates of H. halobium derived from a single colony show polymorphisms in the position and number of members of individual repeated-sequence families. Genome rearrangements affecting the position and number of repeated sequences occur at the rate of $0.004 \cdot \text{repeated-sequence family}^{-1} \cdot \text{cell generation}^{-1}$ in such isolates.

The genomes of H. volcanii, H. salinarum, H. trapanicum, H. valis-mortis and H. saccharovorum also contain repeated-sequence families, some of which are shared by H. halobium. These repeated-sequence families have been subjected to expansion, contraction and occasional loss during the divergence of the halobacteria, but are more highly conserved than are unique-sequence DNAs, suggesting they may have profited from stochastic or selective intragenomic processes which have maintained them in sequence over long evolutionary periods.

. . . For Linda, Francis and all of the gorillas who perished in the attempt to swim Block Island Sound.

Acknowledgments

My mother never thought I was cut out to be a scientist. She may yet be proven correct. Certainly, there are other careers which would suit my temperament as well. If I could command an exorbitant salary for drinking beer and reading novels, I might choose to do so. But I am not known for being rational, so (therefore?) I might not.

Partly in deference to my mother's (not unfounded) fears and partly because of my attraction to things fantastic, I have gravitated towards those areas of science which address questions that are either ultimately unknowable (Evolutionary Biology) or so hopelessly complex that they give the illusion of being ultimately unknowable (Biological Oceanography). In my quest to know the unknowable (I also have a tendency to be overly dramatic), it has been my experience that only a very few scientists really love science. I am privileged to call two of them my friends and teachers. The first is Ian Morris, and I credit him with (but do not hold him responsible for) initiating both my intellectual development and moral decay.

Continuation of a task so ably begun required the services of no ordinary man. Ford Doolittle has admirably risen to the occasion on both fronts. Professionally, he has been of more help than any student has a right to expect. As a friend, he has always "been there", which is the best thing one can say about a friend. I have taken great delight in all of our discussions, be they drunken and personal at 2:00 a.m., drunken and philosophical at midnight, scientific at any and all hours, or one of our endless series of literary disagreements (I leave the sentence beginning with "But . . ." in the first paragraph as a parting gift). I shall miss those times more than any others.

I have spent several hours (depriving my wife of her shower in the process) and wasted many sheets of paper trying to figure out a way to thank those many others who have helped me along the way without saying, "I would like to thank A, B and C for X, Y and Z. (cha-cha-cha)". You may (then again, you may not) derive some comfort from the knowledge that I tried. Unfortunately, I did not succeed. Thanks to (in no particular order - you may all go on believing you deserve more credit than you receive): The Food Chain Research Group (I'm not a failure, am I, Uncle Mike?), The Bigelow Laboratory, The Thistle Inn, The Fats and Greasers, Mike Gray, Chris Helleiner, Cathy Lazier, Colin Stuttard, Annalee Cohen, Frank Sasinek, Kris Calhoun, The Late Tim Mague, Penny "missed my chance" Chisholm, Linda "too embarrassed to breathe" Bonen, Murray "Scumbucket" Schnare, David "just when you thought it couldn't get any worse . . ." Spencer, Jimmy the Kid; Buckwheat, All of The Laps, Ex You, Glen "The Buzzard" Cota, R a w w n ' M u h K a a y, Large Alex and Jason "better living through chemistry" Hofman.

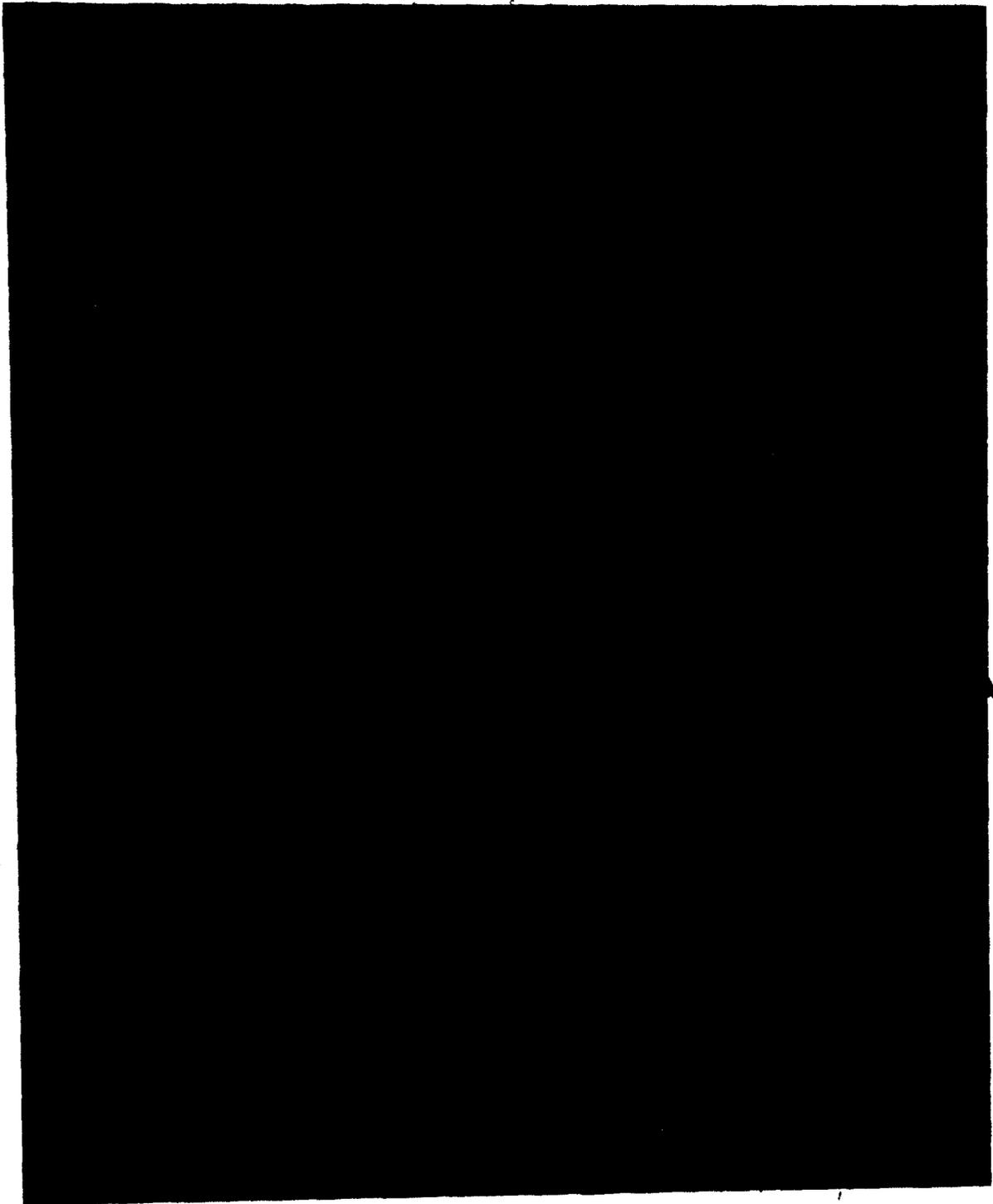
Secure in the knowledge that I may do so without being corny, I thank my mother and father and my mother-in-law and father-in-law for their continued emotional, and too often financial, support through a period when most normal men in their late 20's had "real" jobs. My wife, Linda, deserves more thanks than I am capable of putting on paper, but then, she probably already knows this.

Carmen Sapienza
17 March 1982
Halifax, Nova Scotia

-v-

My grandfather ... used to talk for hours without ever linking one subject to another. He was absolutely incoherent. But then, he was Zoroaster, the prophet of Truth; and just as the One God that he served is obliged to entertain, simultaneously, every aspect of all creation, so did His prophet Zoroaster. The result was inspiring if you could ever make sense of what he was saying.

Gore Vidal,
Creation, 1981



Phenotyp^{ic} exhibited by Halobacterium halobium colonies on solid medium. Wild-type colonies are light pink, gas vacuole-deficient colonies are brown and colonies altered in pigment production are dark pink or whitish.

I. Introduction

In 1977, Woese and Fox proposed a dramatic reclassification of the biological world. Based on molecular sequence data of 16S and 18S ribosomal RNAs derived from a large number of organisms, they distinguish three "primary kingdoms": the Eubacteria, the Urkaryotes and the Archaeobacteria. Most of the better known prokaryotes, including the Cyanobacteria (blue-green algae), constitute the Eubacteria, while the Urkaryotes comprise the nuclear-cytoplasmic components of all eukaryotic cells so far examined. Contained within the third kingdom, the Archaeobacteria, are a collection of prokaryotes diverse in metabolic capabilities and ecological habitat which include the methanogens, the extreme halophiles and three thermoacidophiles (Fox *et al.*, 1980; Zillig *et al.*, 1981) (this list continues to grow). These three groups appear to have diverged at a very early stage in cellular evolution (>3.5 billion years ago) and, as Woese and Fox note, the Archaeobacteria are "no more related to typical bacteria than they are to eukaryotes".

The Archaeobacteria exhibit, in fundamental aspects of their physiology, biochemistry and molecular biology, some traits which are eubacterial, some which are eukaryotic and others which appear unique (Woese, 1981): (1) Archaeobacterial cell walls do not contain muramic acid; (2) their cell membranes contain branched chain ether-linked lipids (rather than ester-linked) and in at least one case, the cell membrane may not be a bi-layer in the classical sense; (3) transfer RNA modification patterns are distinctive and the initiating tRNA in protein synthesis is non-formylated; (4) the RNA polymerase subunits are distinctive and result in antibiotic sensitivities which are unusual; (5) some members of the Archaeobacteria (methanogens) contain cytochromes not found outside this group.

The extreme halophiles are the Archaeobacteria most readily cultured in the laboratory. This attribute, together with the existence of visually identifiable mutants in pigment and gas vacuole production, make the halobacteria the logical choice for genetic and molecular biological investigations.

Extremely halophilic bacteria (species of the genera Halobacterium and Halococcus) require salt concentrations of at least 15% for structural integrity and grow best at salt concentrations (as NaCl) of 20-30%. Their natural habitats are salterns and hypersaline lakes and lagoons, but halobacteria are also capable of growing on salted fish and hides (Kushner, 1978).

While much data accumulate on many aspects of the biology of Archaeobacteria, very little is known about the genomes of Archaeobacteria except their genetic complexities (as measured by renaturation kinetics). The genetic complexity of Methanobacterium thermoautotrophicum is ca. 1.1×10^9 daltons (Mitchell et al., 1979), that of Thermoplasma acidophilum ca. 0.8×10^9 daltons (Searcy and Doyle, 1975) and those of several halobacteria ca. 2.5×10^9 daltons (Moore and McCarthy, 1969a,b).

The DNA of Halobacterium halobium and related species can be separated into components of 66-68 and 57-60 mol percent G+C by CsCl density gradient equilibrium centrifugation (Weidinger et al., 1979). Much but not all of the latter fraction is the DNA of a 150 kbp (kilobase pair) plasmid present in some 4 to 5 copies in H. halobium. Changes in plasmid DNA restriction endonuclease digestion patterns are associated with mutations in gas vacuole or pigment production (Pfeifer et al., 1981a,b). These changes may be interpreted as resulting from complex and multiple insertions, deletions and rearrangements of DNA. Such

alterations occur with astonishing frequency (Pfeifer et al., 1981b), suggesting that the halobacterial genome may contain many transposable elements or regions of sequence homology promoting recombination within and between chromosomal and plasmid DNAs.

Renaturation kinetic analyses show no substantial rapidly reannealing fraction which might represent the DNAs of a few, high copy number, repeat sequence families (Mboore and McCarthy, 1969b). Data presented here, however, show that: (1) the genomes of Halobacteria harbor many different families of repeated sequences; (2) the repeated sequence families are small (2-20 ? members) and the members of these families are arranged in both dispersed and clustered fashion; (3) repeated sequences are found on both plasmid and chromosome in H. halobium; (4) some repeated DNAs are more highly conserved in sequence between distantly related halobacterial species than are unique sequence DNAs; (5) genomic rearrangements affecting the location of repetitive sequences are very frequent and not obligately associated with detectable phenotypic alterations.

II. Materials and Methods

A. Sources of strains: Halobacterium halobium NRC-1 and R1 and H. salinarum were obtained from R.D. Simon (University of Rochester). H. halobium R1 is a spontaneous, gas vacuole-deficient mutant of H. halobium NRC-1 originally isolated in 1969 by W. Stoeckenius (University of California, San Francisco). H. volcanii was obtained from C.R. Woese (University of Illinois, Urbana). H. trapanicum, H. vallismortis and H. saccharovororum were obtained from G.E. Fox (University of Houston). Escherichia coli JF1754 (leuB hisB metB r_K⁻ m_K⁻ lac gal) was obtained from J. Friesen (University of Toronto).

B. Growth of Strains: H. halobium NRC-1 and R1 and H. salinarum were grown in a medium containing (per L): 250 g NaCl, 20 g MgSO₄·7H₂O, 3 g Na₃ citrate, 2 g KCl, 0.2 g CaCl₂·2H₂O, 3 g yeast extract (Difco) and 5 g tryptone (Difco) (Gochnauer and Kushner, 1969). The salt mixture (900 ml) and the yeast extract/tryptone (100 ml) were autoclaved separately (20 p.s.i. for 20 minutes) and combined after cooling to less than 50°C. H. volcanii were grown in a medium containing (per L): 125 g NaCl, 45 g MgCl₂·6H₂O, 10 g MgSO₄·7H₂O, 10 g KCl, 1.34 g CaCl₂·2H₂O, 3 g yeast extract and 5 g tryptone. Salts were autoclaved separately from the nutrient solution. H. trapanicum, H. vallismortis and H. saccharovororum were grown as described by Bayley (1971). Solid media were made by the addition of 1.8% agar (Difco) to the salt solution before autoclaving. Liquid cultures were grown with illumination at 37°C on a New Brunswick Scientific shaking platform (250 r.p.m.). Halobacteria grown on solid medium (in 100 mm x 15 mm plastic petri plates, Fisher Scientific) were kept at 37°C in a dark incubator or illuminated in a 37°C constant temperature room. E. coli JF1754 were

grown in either LB medium or M9 medium as described by Miller (1972) with the exception that 10 ml of 10% casamino acids (Difco) solution was added per liter of M9 medium in lieu of 40 $\mu\text{g}\cdot\text{ml}^{-1}$ of D,L leucine, histidine and methionine. Uridine ($1\text{ mg}\cdot\text{ml}^{-1}$) was also added to M9 medium when plasmid DNA was to be isolated from the culture (Norgard et al., 1979). Solid media contained 1.5% agar. Selective media contained either 80 μg penicillin G $\cdot\text{ml}^{-1}$ (approximately 130 U $\cdot\text{ml}^{-1}$) (Sigma Chemical Company) or 12.5 μg tetracycline $\cdot\text{ml}^{-1}$ (Sigma Chemical Company).

C. Purification of DNA: Total halobacterial DNA was prepared by the method of Hofman et al. (1979). H. halobium NRC-1 and R1 plasmid DNA extracts were prepared by the method of Weidinger et al. (1979). CsCl/ethidium bromide density gradient equilibrium centrifugation of H. halobium plasmid DNA extracts was done according to the procedure of Lau and Doolittle (1978) with the exception that gradients were centrifuged for 20 hours (20°C) at 48000 r.p.m. in a Beckman VT150 rotor. Plasmid and chromosomal DNA bands were located in the gradient by their fluorescence under ultraviolet light (a hand-held UV source). Lower band DNA was collected by side puncture of the gradient tube with a syringe. Ethidium bromide was removed from the sample by at least three extractions with an equal volume of either iso-amyl alcohol or iso-propanol. Samples were desalted on PD-10 columns (Pharmacia Fine Chemicals) (used as described by the manufacturer) and DNA precipitated by the addition of 1/10 volume of 3M Na acetate and 2.5 volumes of 95% ethanol and storage at -20°C for at least 3 hours or -70°C for 30 minutes. Precipitated DNA was collected, by centrifugation, pellets washed with 95% ethanol, dried under vacuum, redissolved in 10 mM Tris Cl pH 7.5 and the quantity

of DNA determined spectrophotometrically. The DNA was then reprecipitated and dissolved to the desired concentration in 10 mM Tris Cl pH 7.5.

Plasmid pBR322 DNA was isolated from pBR322 transformed E. coli MM294 (obtained from K. Talmadge, Harvard University) by the procedure of Clewell and Helinski (1969) after amplification by addition of 100 µg chloramphenicol·ml⁻¹ (Sigma Chemicals) and purified by two cycles of CsCl/ethidium bromide density gradient equilibrium centrifugation.

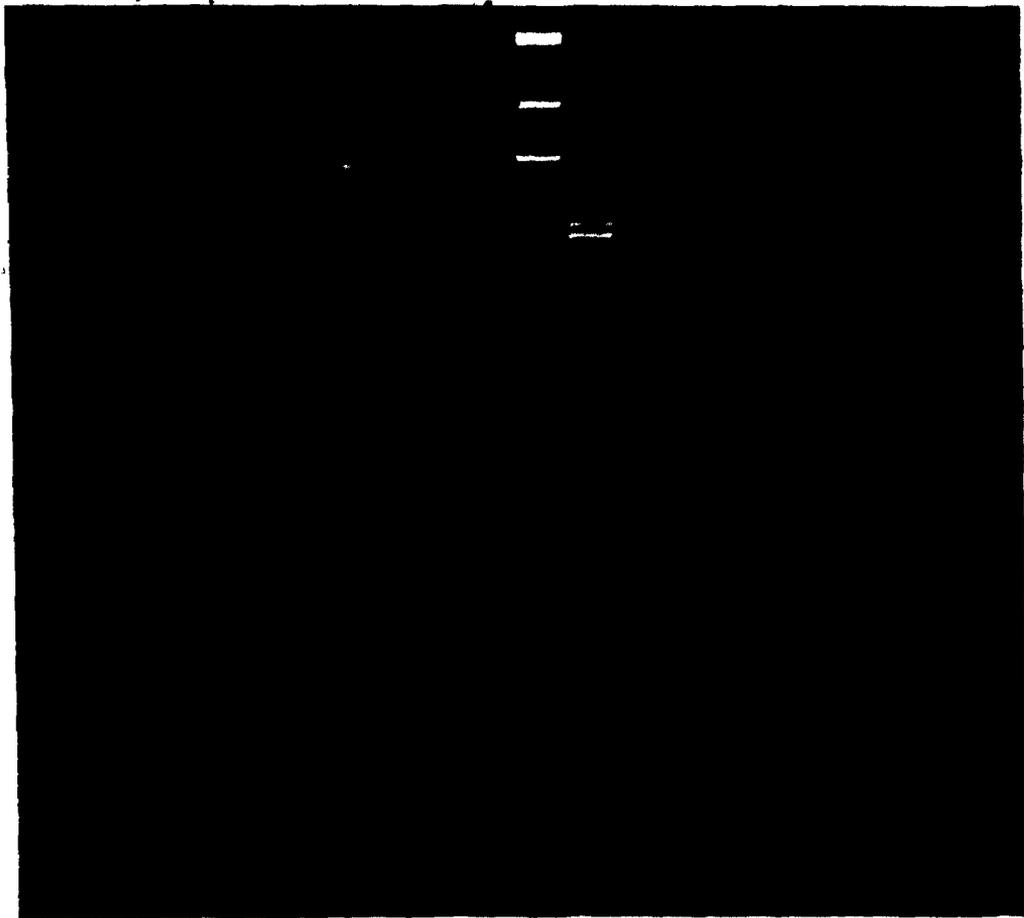
Large-scale isolation of Halobacterium DNA-pBR322 recombinant plasmids was also done in this manner with the exception that only a single cycle of CsCl/ethidium bromide density gradient equilibrium centrifugation was used, and samples were desalted by first adding 1.25 volumes of distilled H₂O and precipitating the DNA with two volumes of 95% ethanol at -20°C for 3 hours. If only small amounts of recombinant plasmid were needed, the following procedures were used.

(1) For agarose gel-screening of recombinant plasmids: this procedure was adapted from Holmes and Quigley (1981). A large colony (approximately 3 mm or a similar number of cells scraped from a plate) was transferred with a toothpick and suspended in 50 µl of 50 mM Tris Cl pH 8.0/50 mM EDTA/5% Triton X-100/8% sucrose/1 mg lysozyme·ml⁻¹ (in a 0.5 ml Eppendorf centrifuge tube). The tube was placed in a boiling water bath for 60-70 seconds and then centrifuged for 10 minutes in an Eppendorf microcentrifuge. The gelatinous pellet was skewered on the end of a clean, plastic micropipette tip and removed. The resulting supernatant may be run directly on an agarose gel (Fig. 1), or, after precipitating and redissolving the nucleic acids in an appropriate buffer, the DNA was digested with 2 units of a restriction endonuclease

Fig. 1. 0.8% agarose gel stained with ethidium bromide showing plasmid DNAs prepared from individual tet^r E. coli JF1754 colonies by the method of Holmes and Quigley (1981).



Fig. 2. 1.2% agarose gel stained with ethidium bromide showing EcoRI cleaved recombinant plasmid DNAs prepared from single colonies by the method of Holmes and Quigley (1981).



(Fig. 2) (1 μ l of pre-boiled 10 mg RNase A \cdot ml⁻¹ [Sigma Chemicals] was also added) for 15 minutes (longer digestion times resulted in non-specific degradation of DNA). All recombinant plasmids used as hybridization probes were screened by these procedures.

(2) For "nick-translation": This procedure is adapted from Birnboim and Doly (1979). A small colony (approximately 1 mm or a similar number of cells from a plate) was transferred to 1 ml of M9 medium (containing 1 mg of uridine (Sigma) and the selective antibiotic) in a sterile 1.5 ml Eppendorf centrifuge tube. The cells were grown at 37°C on a shaking platform to A₅₅₀ of 0.5-1.0. At this time 100 μ l of 1 mg chloramphenicol \cdot ml⁻¹ was added to the tube and incubation continued overnight. Cells were harvested by centrifugation (3 minutes) in an Eppendorf microcentrifuge. The cells were resuspended in 100 μ l of freshly prepared 25 mM Tris.Cl pH 8.0/10 mM EDTA/50 mM glucose/2 mg lysozyme \cdot ml⁻¹ (Sigma) and incubated at room temperature for 5-15 minutes. 200 μ l of 0.2 N NaOH (freshly prepared) was then added and the contents of the tube well-mixed. After incubation at room temperature for 15 minutes, 150 μ l of cold 3 M Na acetate (pH 4.8) was added and the tube incubated at 0°C for 60 minutes. After centrifugation for 10 minutes (Eppendorf microcentrifuge), the supernatant was extracted with an equal volume of phenol (equilibrated against 0.1 M Tris Cl pH 8.0). The nucleic acids in the aqueous phase were ethanol precipitated at least twice before further manipulation. There was usually enough plasmid DNA in such preparations to do 2 restriction endonuclease digestions and 1 "nick-translation".

D. Restriction endonuclease digestions: DNA to be digested with restriction endonuclease EcoRI, BamHI, PstI, HindIII, AvaI, AluI, HaeIII, HpaII, MspI, CfoI, TaqI or Sau3A (restriction endonucleases were

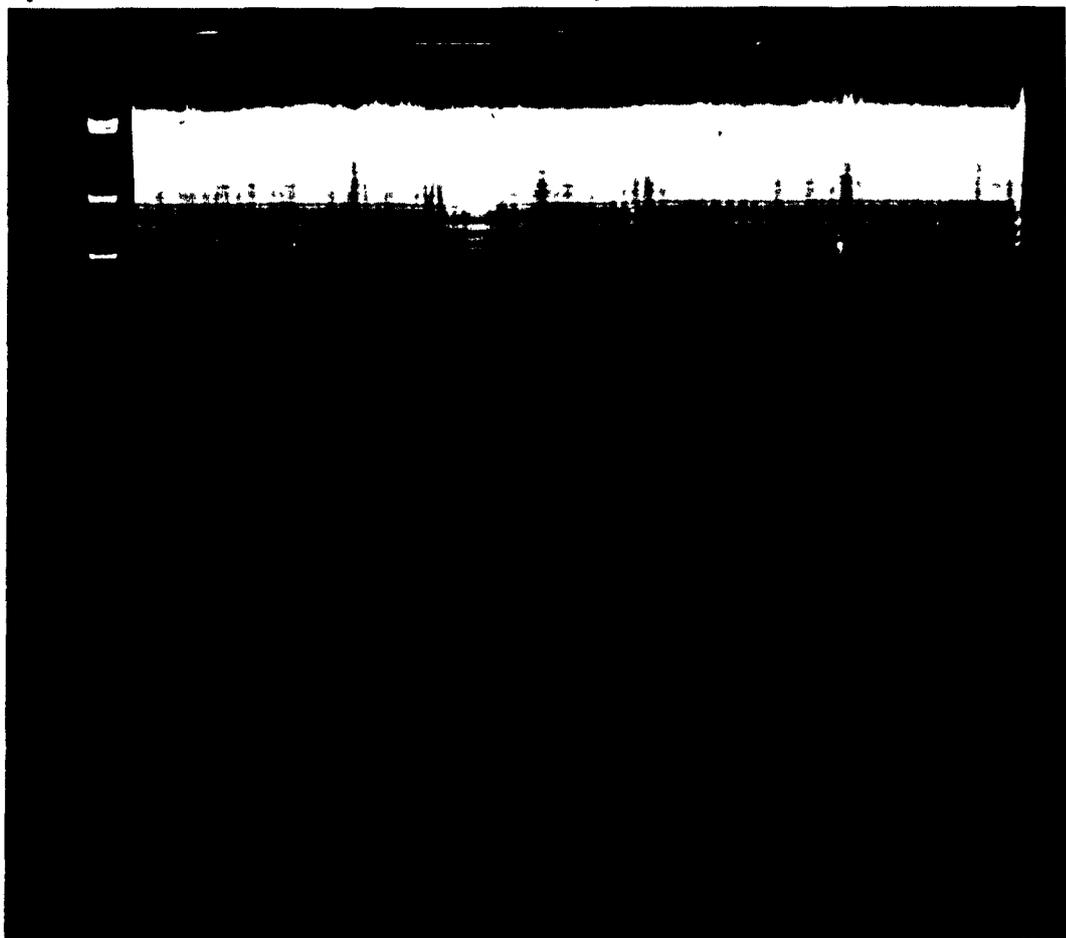
obtained from Boehringer-Mannheim, BRL or New England Biolabs) was dissolved in 50 mM Tris.Cl pH 7.5/10 mM MgCl₂/10 mM DTT (dithiothreitol)/50 mM NaCl/100 µg·ml⁻¹ autoclaved gelatin. DNA to be digested with Sall was dissolved in 50 mM Tris.Cl pH 7.5/10 mM MgCl₂/10 mM DTT/150 mM NaCl/100 µg·ml⁻¹ autoclaved gelatin. At least 2 units (1 unit = that amount of enzyme necessary to completely digest 1 µg of λDNA in 60 minutes at 37°C) of the desired enzyme were added per µg of DNA and time or enzyme concentration was adjusted to achieve at least 4-fold overdigestion. All restriction endonuclease digestions were carried out at 37°C with the exception of TaqI digestions which were incubated at 65°C. The reactions were terminated by the addition of 1/10 volume 100 mM EDTA pH 8.0/10% ficoll (MW 400000, Sigma)/0.075% bromophenol blue (as a dye marker for electrophoresis), by heating at 65°C for 15 minutes, or by extraction with an equal volume of phenol (equilibrated against 0.1 M Tris.Cl pH 8.0) if DNA was to be used in subsequent T4 DNA ligase reactions.

E. Agarose gel electrophoresis and transfer of DNA to nitrocellulose:

Molecular weight markers were λCI857 Sam7 DNA (BRL) digested with HindIII, φX174 DNA (BRL) digested with HaeIII, pBR322 or φX174 digested with AluI. Agarose gel electrophoresis was done essentially as described by Lau and Doolittle (1978). Transfer of DNA fragments to nitrocellulose was essentially as described by Southern (1975). Isolation of DNA restriction fragments from Low Melting Point Agarose (BRL) was carried out as described by the manufacturer. For preparation of a large number of "Southern blots" of the same restriction endonuclease digested DNA, DNA was loaded into a long sample well (6.7 cm x 0.1 cm x 0.2 cm or 11.6 cm x 0.1 cm x 0.2 cm or 16.2 cm x 0.1 cm x 0.2 cm depending on the

number of "Southern blots" required) at a concentration of 5-10 $\mu\text{g DNA}\cdot\text{cm}^{-1}$. After electrophoresis, the gels were stained with ethidium bromide ($1\ \mu\text{g}\cdot\text{ml}^{-1}$) for 20 minutes and destained (to eliminate background fluorescence) in distilled water for 20 minutes. The gels were transilluminated with short wave UV light and photographed with either a Polaroid MP-4 camera (Polaroid type 667 black and white film, red filter with λ cut-off = 600 nm) or a 35 mm camera (Kodak "plus X" black and white film) (Fig. 3). To enable large DNA fragments to be Southern transferred, the gels were exposed to UV light as described by Lau et al. (1980). The gels were then removed from the gel frame, placed in a glass or enamel baking dish and covered with 0.2 N NaOH/0.6 M NaCl. The gels were soaked in this mixture for 40 minutes with occasional gentle agitation. The denaturing solution was then aspirated away and the gel rinsed briefly with distilled water. 1.0 M Tris Cl pH 7.5/0.6 M NaCl was then added to the baking dish and the gel allowed to neutralize for 40 minutes. After rinsing with distilled water the gel was placed on top of a stack (approximately 3.5 cm high) of 3 MM filter paper pre-soaked with 20X SSC (1X SSC is 0.15 M NaCl/0.015 M Na₃ citrate) in the bottom of a large area plastic tray filled to just below the level of the 3 MM filter paper stack (with 20X SSC). Nitrocellulose filter paper (Millipore Corporation) was pre-cut to the desired size; pre-wet in distilled water, rinsed with 20X SSC and placed over the area of the gel containing the DNA fragments to be transferred, taking care to eliminate bubbles between the gel and the nitocellulose. A small stack of 3 MM filter paper (1-2 cm) cut to 0.5 cm smaller than the nitrocellulose paper in both dimensions was then placed on top of the nitrocellulose and sufficient paper towel added (approximately 8 cm) to allow transfer

Fig. 3. EcoRI cleaved H. halobium strain NRC-1 DNA. 80 μ g of this DNA was loaded into the 16.2 cm x 0.1 cm x 0.2 cm well of a 1.2% agarose gel and subjected to electrophoresis at 40v (80mA) for 14 hours. DNA fragments were transferred to a 16.5 cm x 14.5 cm nitrocellulose sheet. The sheet was then cut into 0.5 cm x 14.5 cm strips for hybridization.



to proceed for at least 8 hours without wetting of the entire stack. After transfer the nitrocellulose was briefly dipped into a tray of distilled water, blotted on paper towel and laid on a piece of clean cardboard. Using a piece of ruled graph paper (10 mm to the cm) laid over the nitrocellulose and a metal ruler as a guide, 0.5 cm strips were cut from the nitrocellulose with a single edged razor blade. The strips were numbered sequentially with a BIC pen (blue ink is preferable to black ink because it does not bleed under hybridization conditions) and baked at 80°C in vacuo for 2 hours. Strips were stored in baked glass test tubes at room temperature.

F. DNA-DNA hybridization: Nitrocellulose filter-bound DNA was hybridized with in vitro ^{32}P -labelled probes essentially as described by Jeffreys and Flavell (1977). Nitrocellulose filter-bound DNA was pre-incubated in 5X SSC/0.1 M Na phosphate pH 7.0/0.02% polyvinyl pyrrolidone (MW 360,000, Sigma)/0.02% ficoll/0.02% bovine serum albumin (Sigma)/50 μg sheared salmon sperm DNA $\cdot\text{ml}^{-1}$ (Sigma)/10 mM EDTA/0.5% sarcosyl (approximately 4 ml per 0.5 cm x 14.5 cm nitrocellulose strip) at 65°C for at least 8 hours in zip-lock plastic bags. The plastic bags containing the strips and hybridization mix were suspended vertically in a 65°C oven to ensure the nitrocellulose strips and all liquid remained at the bottom of the bag. Denatured ^{32}P -labelled probes were added directly to the bags after pre-hybridization and hybridization continued for 18-36 hours. Filter strips were then removed from the bags and placed in a baking dish containing approximately 500 ml of room temperature 4X SSC. One more rinse in room temperature 4X SSC was done to remove most of the unhybridized ^{32}P -label. All subsequent washes were carried out at 65°C in a shaking water bath (approximately 40 r.p.m.).

The washing procedure involved 3 washes (20 minutes each) in 4X SSC (500 ml each wash), 1 wash in 2X SSC (60 minutes), 1 wash each in 1X SSC (20 minutes), 0.3X SSC (20 minutes) and 0.1X SSC (20 minutes). The filter strips were blotted dry on paper towels, mounted on cardboard, covered with plastic wrap and exposed to X-ray film (Kodak X-OmatAR "XAR-5" or 3M) at -70°C with Dupont "Lightning-Plus" intensifying screens.

G. In vitro formation of recombinant molecules and transformation of *E. coli* JF1754: Ligation of restriction endonuclease digested Halobacterium DNA to similarly cleaved plasmid pBR322 (Sutcliffe, 1978) was done as suggested by the manufacturer of the T4 DNA ligase (Boehringer-Mannheim) as modified from Graf (1979). Transformation of *E. coli* JF1754 was done as described by Norgard et al. (1978).

H. In vitro ^{32}P -labelling of DNA: Plasmid DNAs isolated by the modified procedure of Birnboim and Doly (1978) were in vitro ^{32}P -labelled by "nick translation" (Rigby et al., 1977). DNAs to be nick-translated (1-2 μg or an unknown amount from "rapid plasmid" preparations) were dissolved in 45 μl of 10 mM Tris.Cl pH 7.5 and 50 μl of 100 mM Tris.Cl pH 7.5/10 mM MgCl_2 /2 mM DTT/2 μM dGTP (Sigma)/2 μM dTTP (Sigma) was added. 5 pmoles (approximately 1.5 μl) of [α - ^{32}P] dATP (New England Nuclear, approximately 3000 Ci. \cdot mmole $^{-1}$), 5 pmoles (approximately 1.5 μl) [α - ^{32}P] dCTP (NEN, 3000 Ci. \cdot mmole $^{-1}$), 1 μl DNase I (500 ng. \cdot ml $^{-1}$) and 1 μl (4 U) DNA polymerase I (NEN) was then added to each tube. The contents of the tube were gently mixed and the tubes incubated at 37°C for 30 minutes. Reaction was terminated and the ^{32}P -labelled DNA denatured by heating in a boiling water bath for 5 minutes, followed immediately by chilling in a 0°C water bath. Specific activities of approximately $1-5 \times 10^7$ cpm. \cdot μg DNA $^{-1}$ were regularly achieved by this method. This mixture

was added directly to plastic bags containing pre-hybridized nitrocellulose strips.

I. 5'-end labelling of restriction fragments: DNA restriction fragments were occasionally 5'-end labelled with ^{32}P (Maxam and Gilbert, 1980) to facilitate mapping of cloned fragments or to generate radioactive molecular weight markers. DNAs to be end-labelled were dissolved in 10 mM Tris Cl pH 7.5 (1-2 μg in 50 μl). 50 μl of 200 mM imidazole (pH 6.6)/36 mM MgCl_2 /20 μM ATP/600 μM ADP/10 mM DTT/200 μg gelatin $\cdot\text{ml}^{-1}$ was added to 40 pmoles [γ - ^{32}P] ATP (which had been evaporated to dryness in the bottom of a 0.5 ml Eppendorf centrifuge tube). The DNA was introduced to this mixture and 1 μl (5 U) polynucleotide kinase (Boehringer-Mannheim Biochemicals) added. The reaction was incubated at 37°C for 20 minutes and stopped by extraction with an equal volume of phenol (equilibrated against 0.1 M Tris.Cl pH 8.0). [γ - ^{32}P] ATP (4000 Ci $\cdot\text{mmole}^{-1}$) was prepared according to the procedure of Walseth and Johnson (1979) and was a gift from Dr. Jason Hofman.

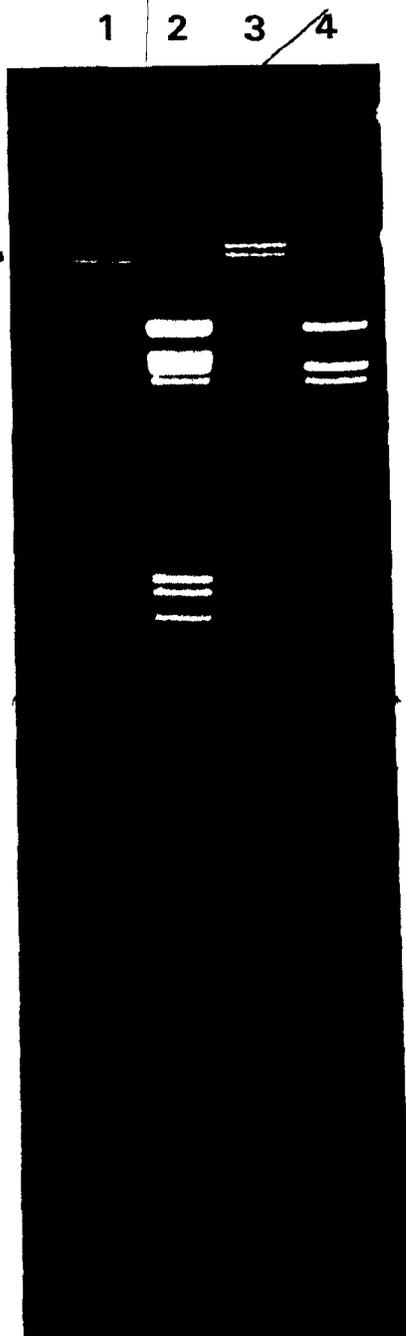
III. Results and Discussion

A. H. halobium plasmid DNAs

Weidinger et al. (1979) and Pfeifer et al. (1981) observed the restriction endonuclease digestion patterns of plasmid DNA from wild-type and gas vacuole-deficient isolates of H. halobium to differ. Weidinger et al. (1979) suggested that the genes involved in gas vacuole formation are encoded on the plasmid DNA of H. halobium. Plasmid DNAs from wild-type H. halobium NRC-1 and three gas vacuole deficient mutants (strain R1 of Stoeckenius and two independent spontaneously occurring mutants isolated in this laboratory) showed many differences in restriction endonuclease digestion pattern (Fig. 4). Strain R1 plasmid DNA appears to have, at most, two common BamHI fragments with the parent NRC-1 plasmid. The lack of similarity may be the result of successive rearrangements which occurred gradually over the approximately 11 years these two strains have been separated. However, one of the other mutants (M9) has a BamHI digestion pattern similar to that of strain R1, indicating major changes in plasmid DNA restriction endonuclease digestion patterns may occur over a much shorter period. The BamHI digestion pattern of mutant M6 plasmid DNA is most similar to that of wild-type plasmid DNA when the restriction patterns of all four are considered.

The molecular weight of the wild-type plasmid determined by summing the molecular weights of the BamHI restriction fragments is 92.3×10^6 (approximately 140 kilobase pairs). Allowing for the error involved in determining the molecular weight of the largest restriction fragments, this estimate is in reasonable agreement with the estimate of 100×10^6 made by Weidinger et al. (1979). The molecular weights of the mutant

Fig. 4. BamHI cleaved H. halobium plasmid DNAs resolved on a 1.1% agarose gel. Track 1: strain NRC-1 plasmid DNA (vac^+). Track 2: strain RI plasmid DNA (vac^-). Tracks 3 and 4: M6 and M9 plasmid DNAs, respectively. These two strains are independent vac^- isolates of strain NRC-1.



plasmid DNAs are 37×10^6 (strain R1), 76×10^6 (M6) and 30×10^6 (M9). All of the mutant plasmid DNAs have apparently undergone extensive deletions. Pfeifer *et al.* (1981) also observed deletions in plasmid DNAs isolated from different mutants, but in most cases the deletions were not so extensive. The mutant plasmids must also have undergone insertions and rearrangements (or multiple deletions) because a single large deletion could not give rise to the many restriction pattern differences observed. It is impossible to determine from these data which, if any, wild-type restriction fragment encodes the gas vacuole protein.

Pfeifer *et al.* (1981) observed insertions of 0.35×10^6 or 1.2×10^6 daltons in a particular restriction fragment (E6) of all gas vacuole deficient mutants they examined. However, several revertants to wild-type (three of the 9 examined) maintained these insertions. The assertion that plasmid genes control gas vacuole formation seems therefore unjustified. Pfeifer *et al.* (1981) also note that introduction of one insertion into the plasmid seems to trigger multiple insertions, deletions or rearrangements. Such multiple events make it difficult to identify relationships between genotypic and phenotypic alterations.

B. Identification of repeated sequences on the plasmid DNA of strain R1

When the approximately 56 kbp (kilobase pair) plasmid from strain R1 is digested with HindIII, three fragments of 7.1 kbp, 8.0 kbp and approximately 40 kbp are generated. 50 μ g of this plasmid were digested with HindIII and the fragments separated by electrophoresis on 1.0 % low melting point agarose. Each DNA fragment was isolated from the gel (Fig. 5) and separately cloned into the HindIII site of pBR322. During restriction mapping of the 7.1 kbp and 8.0 kbp HindIII fragments, it became apparent that the two fragments were related. If an insertion of

Fig. 5. . HindIII fragments of H. halobium strain R1 plasmid purified by electrophoresis and isolated from a low-melting point agarose gel. Samples of each fragment were separately ligated into HindIII cleaved pBR322.



some 900 bp of DNA were made in the 7.1 kbp clone, the restriction maps would coincide (Fig. 6). This insertion did not occur during cloning because samples of the originally isolated DNA fragments also showed similarities in restriction endonuclease digestion patterns (Fig. 7).

All three plasmid DNA fragments cross-hybridize to some extent under stringent conditions (Fig. 8). Clone II probes all BamHI and EcoRI fragments of clone IV and also weakly hybridizes to the largest BamHI and EcoRI fragments of clone X. Clone IV probes all BamHI and EcoRI fragments of clone II and probes the largest BamHI and EcoRI fragments of clone X to a greater degree than clone II. Clone IV also probes a 4 kbp EcoRI fragment of clone X and weakly hybridizes to a 900 bp EcoRI fragment. Clone X hybridizes strongly to the largest BamHI fragment of clone IV and hybridizes to the largest BamHI fragment of clone II to a lesser degree. Clone X also hybridizes to the smallest EcoRI fragment of clone II and to all three EcoRI fragments of clone IV, although the largest EcoRI fragment is probed only weakly. These data indicate that sequences homologous to the shaded area of clones II and IV in Fig. 6 are also present on the 40 kbp HindIII fragment.

The 2.4 kbp HindIII-EcoRI fragment from the right side of clone II (see Fig. 6) was used as a hybridization probe against EcoRI digested total DNA from strain NRC-1, strain R1 and 16 independently isolated, spontaneous gas vacuole deficient or pigment deficient mutants. Fig. 9 shows that this probe hybridizes to two EcoRI fragments in all strains examined. A 4.2 kbp EcoRI fragment is commonly probed in all strains and a larger fragment of either 9.1, 11.7, 12.3, 13.3, 14.4 or 15.2 kbp is also probed. In the case of strain R1, both probed bands are plasmid derived sequences. The 4.2 kbp band must represent at least two plasmid

Fig. 6. Restriction endonuclease cleavage maps of strain R1 plasmid HindIII fragments.

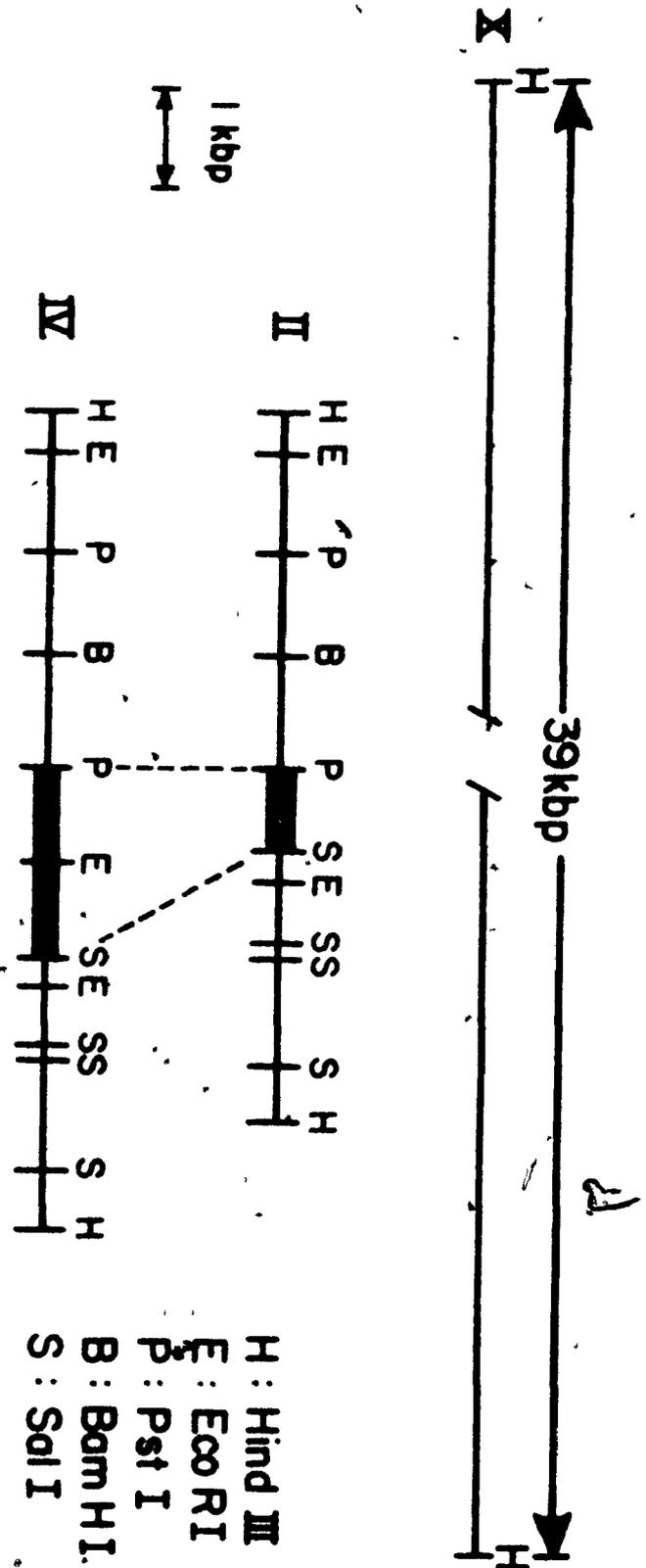


Fig. 7. Electrophoretically purified 8.0 kbp (tracks 1, 3 and 5) or 7.1 kbp (tracks 2, 4 and 6) HindIII fragments of strain R1 plasmid DNA cleaved with BamHI (tracks 1 and 2), EcoRI (tracks 3 and 4) or PstI (tracks 5 and 6).

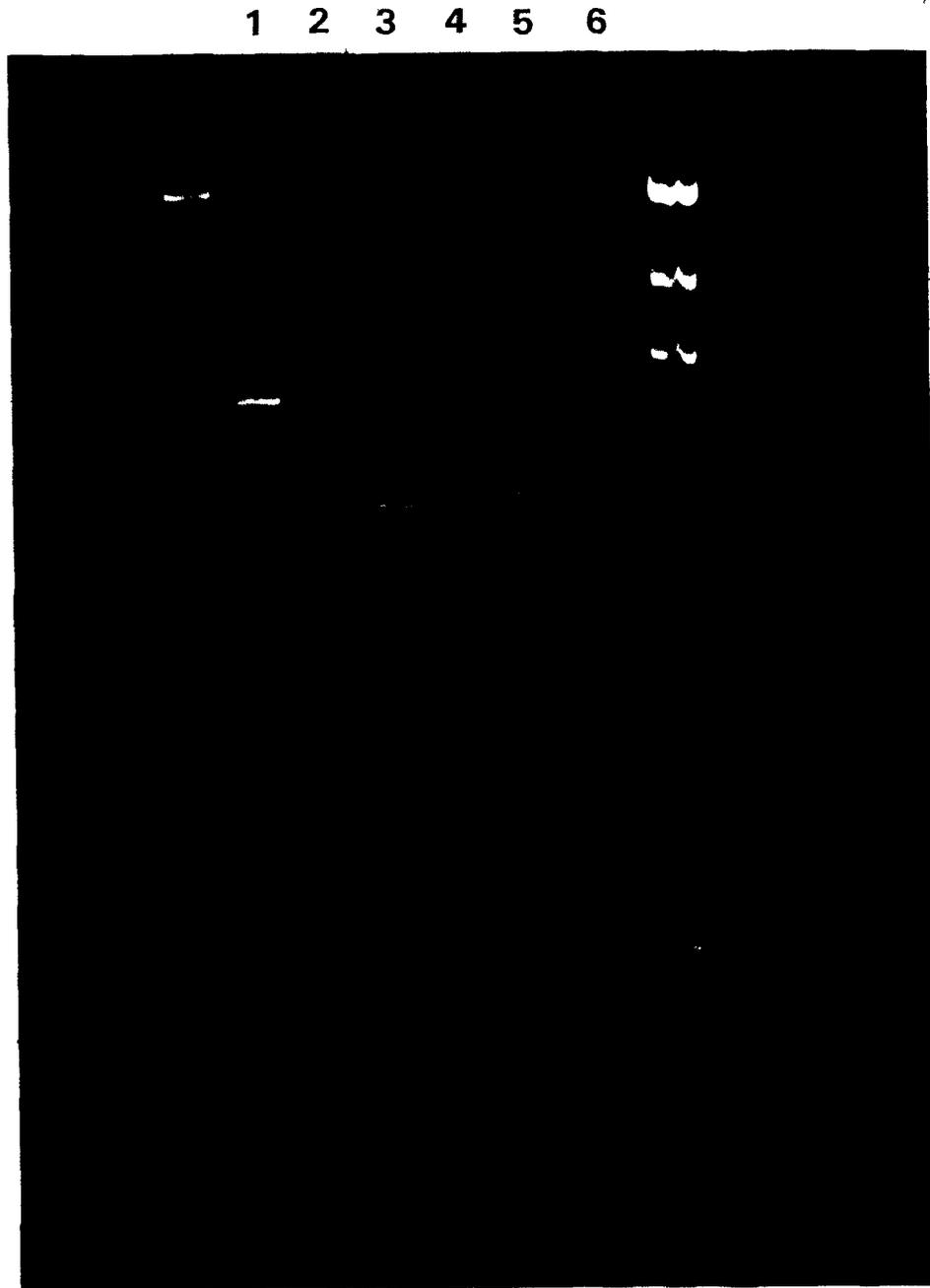


Fig. 8.A. One of three identical sections of an agarose gel transferred and hybridized with ^{32}P -labelled cloned strain R1 plasmid ^{32}P , HindIII fragments. Tracks A, B and C: HindIII-BamHI cleaved clone II, clone IV and clone X DNA respectively. Tracks D, E and F: HindIII-EcoRI cleaved clone II, clone IV and clone X DNA respectively. Track I: HindIII cleaved λ DNA and HaeIII cleaved ϕX174 DNA.

B. Hybridization of ^{32}P -labelled clone II, clone IV or clone X DNA to nitrocellulose filter bound DNAs described in A. Tracks G, H: HindIII-BamHI and HindIII-EcoRI cleaved pBR322 respectively.

17. 30

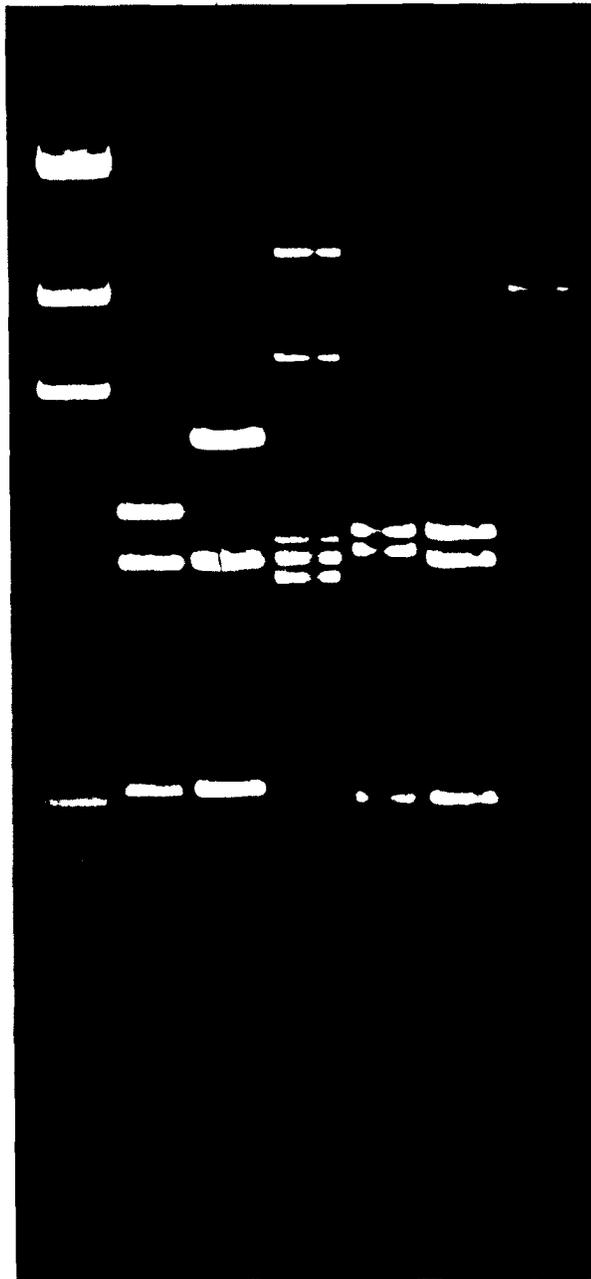


Fig. 8B

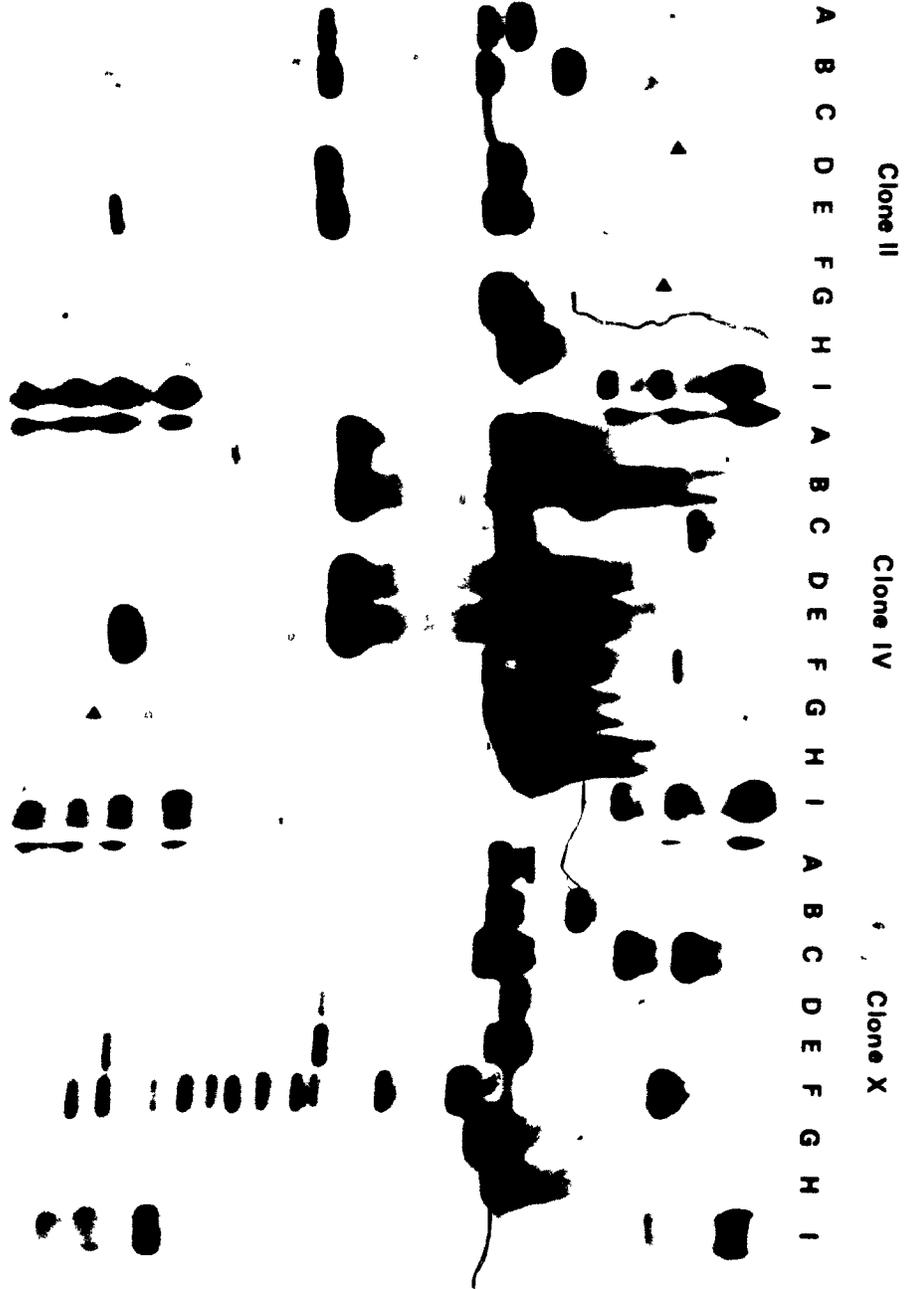


Fig. 9. Hybridization of ^{32}P -labelled 2.4 kbp HindIII-EcoRI fragment derived from the right hand side of clone II (see Fig. 6) to EcoRI cleaved total DNAs from strain MRC-1 (lane A), strain R1 (lane B) and 16 independently isolated gas vacuole or pigment deficient mutants (lanes C-R). Track X: molecular weight markers.



copies of this sequence (those derived from the right hand ends of clones 2 and 4) while the 9.1 kbp band must represent at least 1 plasmid copy of this sequence (derived from an internal EcoRI fragment of clone X). The possibility that chromosomal copies of this sequence are present may not be eliminated, although it seems unlikely that chromosomal EcoRI fragments containing this sequence should be of the same mobility as plasmid EcoRI fragments. It seems more likely that all fragments containing this sequence are plasmid derived, and that differences in hybridization pattern between strains reflect rearrangements in the plasmid DNA of these strains. There are, however, strain R1 plasmid DNA sequences which are present on the chromosome of strain NRC-1. Fig. 10 shows the hybridization pattern obtained when HindIII-EcoRI double-digested plasmid and chromosomal DNAs from strain NRC-1 are probed with clone II. At least three non-plasmid bands are probed in strain NRC-1. The largest fragment in the NRC-1 plasmid track is probably also of chromosomal origin, because there was slight contamination of NRC-1 plasmid DNA with chromosomal DNA. Schnabel et al. (1982) also observed that some plasmid sequences are present on the chromosome of H. halobium.

Because the HindIII-BamHI fragment from the left hand side of clone II does not probe chromosomal bands (data not shown), the chromosomal bands probed in strain NRC-1 must contain sequences homologous to the cross-hatched region of clone II. It is in this region that clone IV differs from clone II by an insertion of some 900 bp of DNA. The area where the insertion occurred was further localized by restriction mapping and the region containing the insertion in clone IV was sub-cloned into pBR322 (as PstI-EcoRI and EcoRI-SalI fragments) and used as a

Fig. 10. Hybridization of ^{32}P -labelled clone II (see Fig. 6) to HindIII-EcoRI cleaved plasmid. (lane A) or chromosomal (lane B) DNA from strain NRC-1. Closed triangles denote the positions of non-plasmid fragments. Open triangles denotes the positions of probable non-plasmid fragment.



hybridization probe against BamHI cut strain R1 plasmid and total DNA. Fig 11 shows that sequences homologous to this insertion are also present on the chromosome of strain R1.

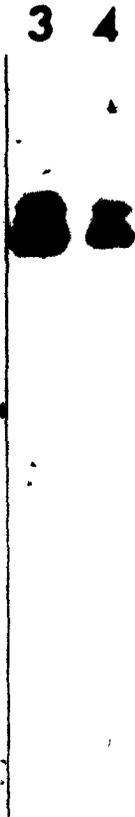
The presence of repeated sequences in H. halobium and the high frequency of plasmid DNA rearrangements observed suggests that these repeated sequences may be transposable elements like those observed in E. coli (Kleckner, 1981). Alternatively, recombination within and between plasmid and chromosomal DNAs mediated by these homologous regions may generate the DNA rearrangements observed.

C. H. halobium contains more repeated sequences than E. coli

Eukaryotic nuclear genomes differ from eubacterial genomes in the quality and quantity of repetitive sequences they contain. Most eukaryotic nuclear genomes contain dispersed repeated and tandemly repeated sequence families which often comprise substantial fractions of the total genome (but see Timberlake, 1978; Krumlauf and Marzluf, 1980). In contrast, eubacterial genomes are not known to contain any substantial fraction of tandemly repeated sequences and only about 1% of the E. coli genome is composed of dispersed repeated sequences (Kopecko, 1980). Given the antiquity of the divergence of the Archaeobacterial, Eubacterial and Urkaryotic lineages, it is possible that Archaeobacterial genomes differ from eubacterial genomes in the number and kind of repeated sequences present.

Restriction endonuclease fragments of a given size resolved on an agarose gel of completely digested total DNA should not form hybrids with fragments of greater or lesser size resolved on the same gel, unless they share common sequences. BamHI digested total DNA from H.

Fig. 11: Hybridization of ^{32}P -labelled PstI-SalI fragment from clone IV (shaded area in Fig. 6) to BamHI cleaved plasmid (lane 3) and total (lane 4) DNAs from strain R1.



halobium NRC-1 was resolved as 11 cm x 15 cm bands on a 1% low-melting-point agarose gel. The gel was cut in half vertically (in the direction of migration). DNA fragments from one-half of the gel were transferred to a nitrocellulose filter, which was then cut vertically into ten strips. Ten horizontal sections containing fragments of defined ranges of molecular weight were excised from the other half of the gel. DNAs were recovered from each excised section, labelled by nick translation and used separately to probe one of the ten nitrocellulose filter strips. Hybridization signals appearing below or above the intense signal (which correspond to hybridization with fragments of the same molecular weight range) indicate the presence of repeated sequences in fragments of different molecular weights. There are at least 14 such "extra" bands in the autoradiograph shown in Fig. 12. A similar experiment performed with DNA from E. coli MM294 showed one or at most two "extra" bands (Fig. 13). This simple technique is relatively insensitive and probably detects only repeated sequences of reasonably high copy number. It is nevertheless clear that the genome of H. halobium is not like that of E. coli in terms of quantity of repeated sequences.

D. Repeated sequences in an H. halobium R1 genomic library

The detection of repeated sequences in the H. halobium genome by the rather insensitive experiment described in the previous section suggests the presence of a large number of repeated sequences. In order to estimate the number of repeated sequences present in the H. halobium genome, randomly cloned fragments of H. halobium R1 DNA were screened for the presence of repeated sequences.

Total strain R1 DNA doubly-digested with BamHI and EcoRI was ligated to similarly cleaved pBR322 and used to transform E. coli

Fig. 12. Hybridization of ^{32}P -labelled *H. halobium* strain NRC-1 DNA fragments of defined size (see text) to BamHI cleaved strain NRC-1 DNA. Arrow heads denote the positions of fragments which hybridize to probes of larger or smaller size class.



Fig. 13. Hybridization of ^{32}P -labelled E. coli MM294 DNA fragments of defined size (see text) to BamHI cleaved E. coli MM294 DNA. Arrow heads denote the positions of fragments which hybridize to probes of larger or smaller size class.



JF1754. Thirty-five $\text{amp}^r \text{tet}^s$ transformants were randomly selected, and their recombinant DNAs ^{32}P -labelled by nick-translation and used to probe EcoRI digested H. halobium NRC-1 or R1 DNAs bound to nitrocellulose filter strips. Thirty-one of the labelled DNAs probed more than a single EcoRI fragment of strain NRC-1 or strain R1 DNA, indicating that each has one or more sequences present more than once in these genomes. Results for 25 of these cloned probes are shown in Fig. 14, with faint bands detectable on the original autoradiograph indicated by closed triangles. All but clone 9 probe EcoRI fragments in addition to those corresponding to the fragment cloned. Some 60-70 such "extra fragments" are probed in strain R1 DNA (tracks b); slightly fewer strain NRC-1 DNA fragments (tracks a) are probed, and many of these are of mobilities different from strain R1 fragments. Clones 2 and 24 each probe single and different fragments (indicated by large closed circles) which are similar in mobility to EcoRI fragments of strain R1 plasmid DNA (data not shown, but positions of plasmid bands indicated by small closed circles in track X); most of the remaining probed fragments are of chromosomal origin. Although no clones seem to show identical probing patterns (when patterns obtained with strain NRC-1 and strain R1 DNA are both considered), all but clones 4, 21, 22, 23 and possibly 24 probe, with varying intensities, a common ca. 5/kbp chromosomal fragment in strain R1 DNA (open circles) which is not probed in strain NRC-1 DNA. For four of those clones which do probe this fragment (clones 6, 11, 14 and 18), it is the only fragment in addition to that cloned which is probed. Of the remaining 15 probing this common 5 kbp fragment, all probe fragments in addition to this and the fragment cloned (except clone 9, which may contain a different 5 kbp fragment). The

Fig. 14. Autoradiogram showing hybridization of individual ^{32}P -labelled recombinant plasmid DNAs from a genomic library of H. halobium R-1 to total DNA from H. halobium NRC-1 (tracks labelled a) and H. halobium R-1 (tracks labelled b). Closed triangles denote the position of faint bands of hybridization. Open circles denote the position of an approximately 5 kbp band probed by many of the cloned DNAs. Large closed circles in tracks 2b and 24b indicate hybridization to plasmid DNA fragments. Small closed circles in track X indicate the positions of EcoRI fragments of strain R-1 plasmid DNA.

approximately 5 kbp common fragment may thus contain copies of many (10) different elements which are present in one or two locations elsewhere in the genome. Alternatively, this 5 kbp fragment may be the result of a genomic rearrangement which occurred later in the culture of H. halobium R1. This interpretation is favoured by data to be presented later.

The 25 clones used for Fig. 14 and the remaining 10 (not shown), contain in total some 250 kbp of H. halobium DNA or 6% of the genome and bear, at a minimum, 31 different elements present elsewhere in the genome. If they represent a random sample of total DNA, there must be at least 500 repeated elements comprising many families in the strain R1 genome. Some elements must be clustered on cloned DNA, or present in copy numbers greater than 10 in the genome, since clones 21 and 24 alone probe more than one-third of the "extra" fragments. Variation in the intensity of hybridization signals (most obvious with clones 17, 21 and 24) could mean that the fragment probed contains varying numbers of copies of such elements, or that the probes themselves contain several different elements. The less intensely probed fragments could also result from genomic rearrangements which occurred during the growth of the culture, or divergent copies of elements present on the probe.

E. Repeated sequences in an EcoRI library of H. halobium NRC-1 genomic clones

H. halobium R1 is a "mutant", and it is possible that the unknown event which created it resulted in the dispersal of repeated sequences within the genome. A library of wild-type (H. halobium NRC-1) total DNA fragments was created by ligating EcoRI digested total DNA into EcoRI digested pBR322. Plasmid DNAs from 28 of the resulting transformants (pre-screened to ensure recombinant molecules were present) were

labelled in vitro by nick-translation and used to probe nitrocellulose filter strips of strain NRC-1 or R1 DNA, cleaved with EcoRI (Fig. 15). One of these (clone 62) probed only a single fragment of NRC-1 and R1 DNA. The rest all probed multiple fragments of either NRC-1 or R1 DNA. Because the NRC-1 plasmid is large, not easily isolated, and gives a complex EcoRI digestion pattern, it is difficult to tell how many probed fragments are of plasmid origin. The majority of the fragments probed by at least one of these clones (clone 37) are not of plasmid origin, as judged by comparison of mobilities of probed EcoRI, HindIII and Sall cleaved total DNA or plasmid DNA produced by these enzymes (Fig. 16).

All cloned fragments produced different Southern hybridization patterns with DNAs of strain NRC-1 and strain R1 (Fig. 15, tracks a and b respectively), and all (except for clones 37 and 53, which may be identical) produced different, probe-specific, patterns with each of these DNAs. There are however a number of fragments which are jointly probed in different combinations by different cloned fragments, and can be used to distinguish at least four of these shared "sub-patterns" (indicated by open and closed circles and squares in Fig. 15), which are probed in various combinations by clones 36, 37, 40, 43, 50, 53, 68, 73, 74, 77, 83 and 85. Arrowheads are used to identify some fragments which distinguish clones 74 and 85 from each other and from clone 83, which otherwise probe common fragments. These "sub-patterns" are also apparent when the same clones are used to probe PstI digested strain NRC-1 DNA; those which probe various sets of EcoRI fragments in different characteristic combinations also probe different sets of PstI fragments in different characteristic combinations (data not shown). The remaining clones produce hybridization patterns with EcoRI-digested strain NRC-1

Fig. 15. Autoradiogram showing hybridization of individual ^{32}P -labelled recombinant plasmid DNAs from a genomic library of H. halobium NRC-1 to EcoRI digested total DNA from H. halobium NRC-1 (tracks labelled a) and H. halobium R-1 (tracks labelled b).

Open and closed circles and squares indicate shared sub-patterns probed by some of the cloned fragments. Small and large arrow heads indicate fragments which distinguish clones 74 and 85 from each other and from clone 83. Closed triangles adjacent to clone 85 (track a) indicate EcoRI fragments of H. halobium NRC-1 DNA which are also probed by several H. volcanii cloned DNAs (see Fig. 17B).

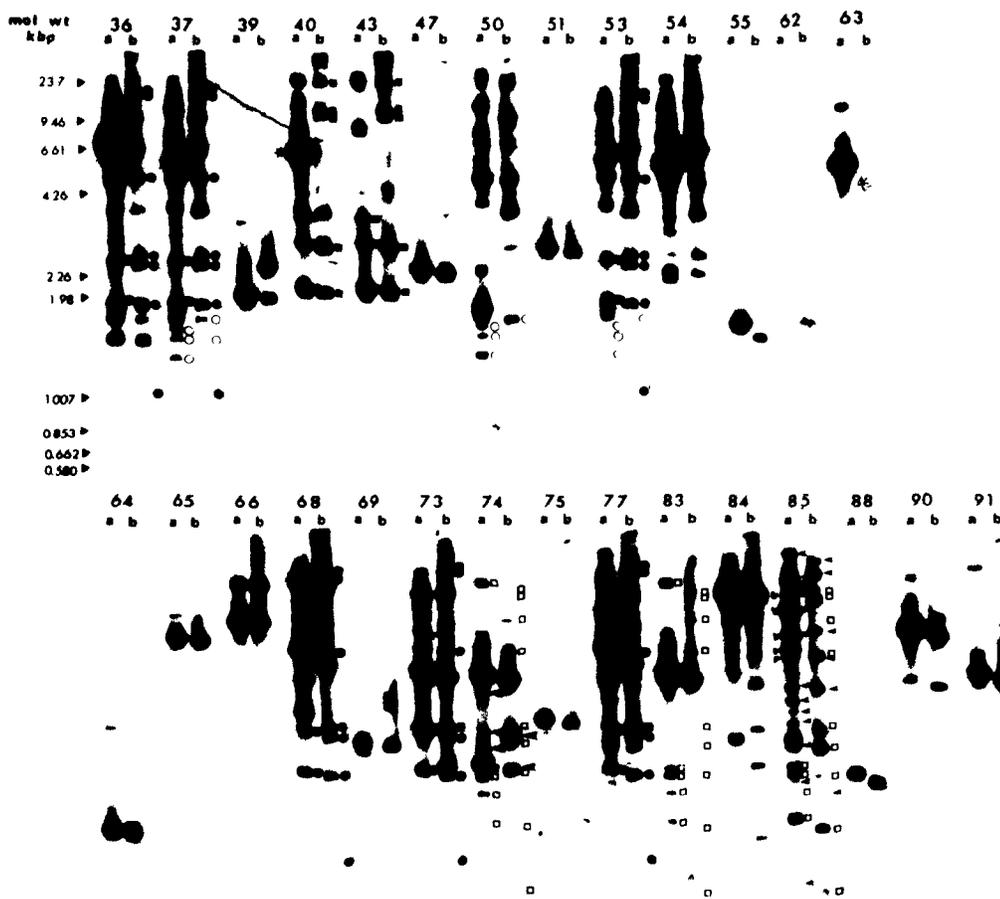
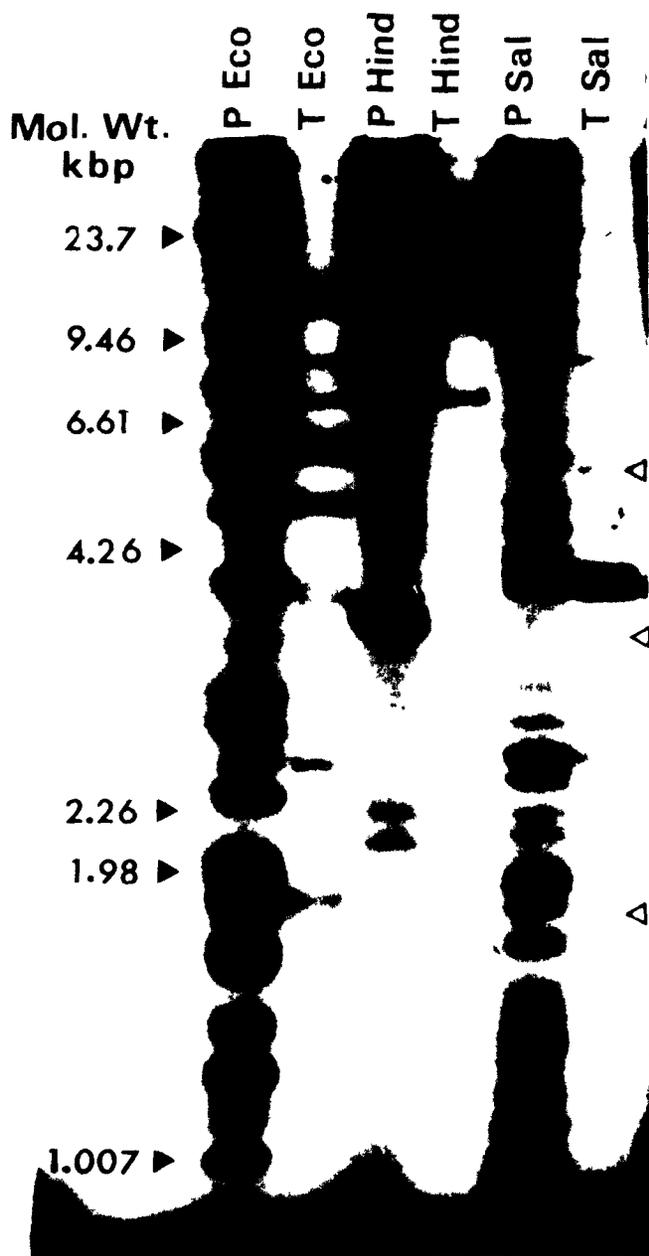


Fig. 16. Comparison of electrophoretic mobility of restriction fragments of strain NRC-1 plasmid DNA with those fragments of strain NRC-1 total DNA hybridized by clone 37 plasmid DNA (see Fig. 15). Strain NRC-1 plasmid (tracks labelled P) was digested with each restriction endonuclease, 5'-³²P-labelled and run adjacent to similarly cleaved non-radioactive total DNA (tracks labelled T). After transfer, the nitrocellulose filter was probed with ³²P-labelled clone 37 plasmid DNA.



DNA which appear unique (and also produce unique patterns with PstI-digested DNA). Intensity variations suggest that these patterns are also composed of "sub-patterns". There are, in any case, at least 21 different repetitive sequences present among these 28 clones (some several times), and most or all of these are different (because they produce different probing patterns) from the repetitive sequences present in the BamHI-EcoRI strain R1 genomic library. It should be noted that clones 40 and 43 probe common fragments and thus must contain a common element(s), and yet the NRC-1 DNA insert in clone 40 (indicated by an asterisk) is not itself detectably probed by clone 43. It should also be noted that some identical fragments probed by clones 50 and 53 (or 83 and 85) produce hybridization signals of different intensities.

Many of the cloned fragments in the EcoRI strain NRC-1 library bear copies of different repetitive elements. Some of these elements must lie rather close to each other in the genome. Cloned fragments which unquestionably bear two different elements (clones 36, 37, 40, 50, 53, 74, 77 and 85) have an average length of only 3.3 kbp (range 1.75-4.6 kbp). It is likely that some differences in the intensity with which individual fragments are probed by different cloned DNAs reflect the presence of different numbers of copies, on both probed and cloned fragments, of shared sequences. Other intensity differences probably reflect genome rearrangements which occur in only a proportion of the population during culture growth (discussed in more detail later).

There should be (on the basis of G+C content) some 300 EcoRI fragments in the H. halobium genome. Elements indicated by symbols in Fig. 15 appear to be present on ten or fewer EcoRI fragments. Among the 28 clones, one of these elements is present five or six times and two of

the elements are present two or three times. There are two possible explanations for this apparent statistical improbability. (1) Such elements are present in many copies on the few EcoRI fragments which are detected as being probed, and also present on a much larger number of EcoRI fragments, but in too few copies to give detectable hybridization signals. Clones bearing EcoRI fragments of the latter sort would then only probe strongly fragments of the former sort. The results with clones 40 and 43 noted above might be consistent with this interpretation. (2) The repetitive sequences represented in Fig. 15 are preferentially located in A+T rich regions of the genome which are more frequently cleaved by EcoRI. Indeed, most of the cloned fragments are smaller (average length 3 kbp) than the expected average size of EcoRI fragments (12 kbp). A small portion of the genome may have been sampled repeatedly.

To test the latter possibility, a small library of PstI fragments of NRC-1 DNA inserted into PstI-digested pBR322 was constructed. The PstI recognition site is similar in G+C content to the H. halobium genome. Thirteen amp^r tet^r transformants were chosen randomly, recombinant plasmid DNA isolated, ³²P-labelled by nick-translation and used to probe PstI digests of total DNA of strain NRC-1 (data not shown). None probed fragments other than that corresponding in mobility to the fragment cloned, and yet these 13 clones contained in total some 48.7 kbp of H. halobium NRC-1 DNA, or 51% as much as the 94.7 kbp represented by EcoRI clones shown in Fig. 15, in which there are at least 30 repetitive elements of at least 21 different families. Thus, repetitive sequences may be clustered in comparatively A+T rich regions of the genome.

F. Repeated sequences in the genome of H. volcanii

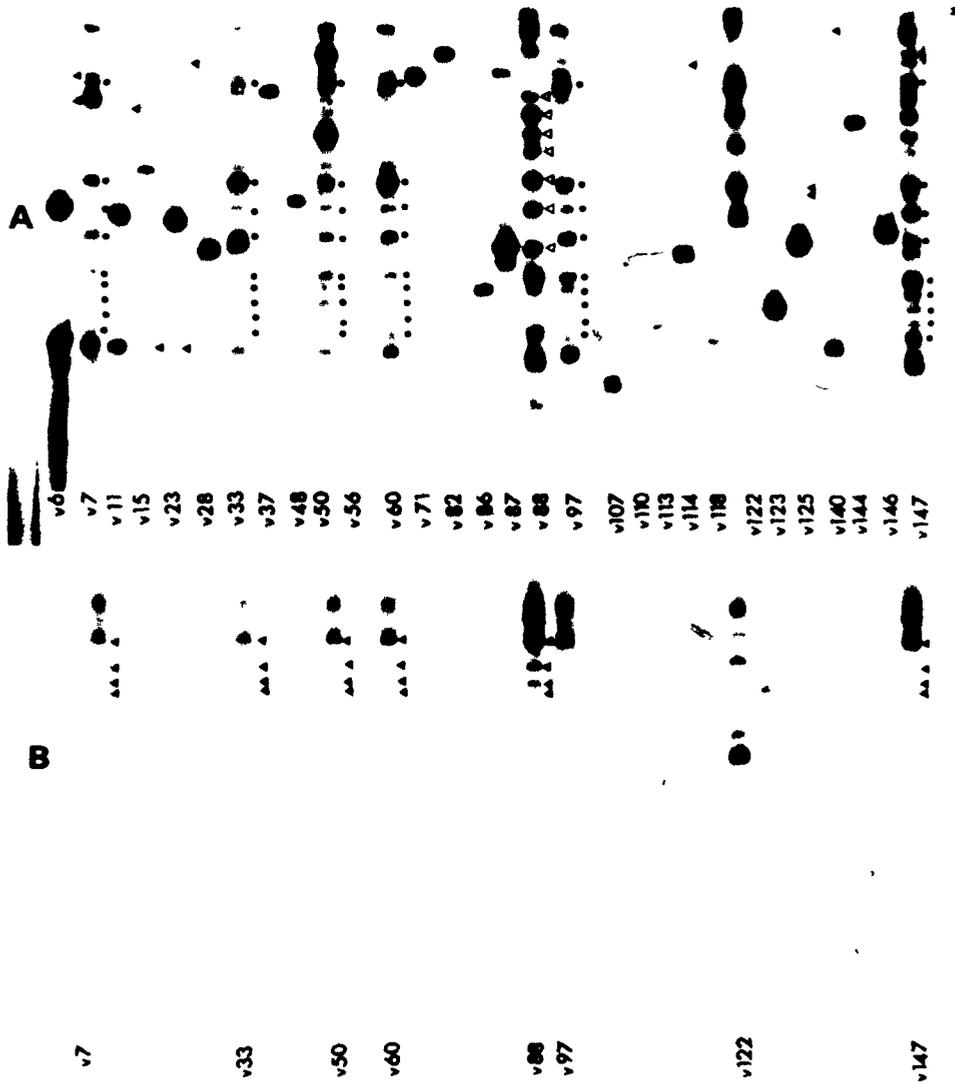
16S rRNA T1 oligonucleotide catalogue analyses show H. volcanii to be as remote, phylogenetically, from H. halobium as Pasteurella or Aeromonas species are from E. coli (Fox et al., 1980). To determine whether it also contains families of repeated elements and whether it shares any of these with H. halobium, an EcoRI library of H. volcanii total DNA was constructed and individually cloned fragments used to probe EcoRI-digested H. volcanii or H. halobium NRC-1 DNAs (Figs. 17A and B). Of 30 randomly-selected clones, ten (V71, V82, V86, V107, V110, V113, V118, V123, V144 and V146) contained unique sequence DNA and probed only fragments identical in mobility to the fragments cloned. None of these probed any fragment of strain NRC-1 DNA. The remaining 20 cloned DNAs probed multiple fragments of H. volcanii DNA (the faintest of which are marked by closed triangles in Fig. 17A). Five different cloned DNAs (V7, V33, V50, V60 and V97) probed an identical large set of H. volcanii fragments (some of which are indicated by closed circles). Clone V147 probed this same set, plus several additional fragments. The remaining multiply-probing cloned fragments produced hybridization patterns which appear unique, although only two of these (V88 and V122) probed more than 3 or 4 fragments in addition to that representing the cloned insert. Only those H. volcanii clones probing more than three or four H. volcanii DNA fragments probed H. halobium NRC-1 DNA but each of these probed multiple fragments (Fig. 17B). Thus some repeated sequences are more highly conserved than are unique sequence DNAs.

Those clones probing H. volcanii fragments indicated by closed circles in Fig. 17A, probed multiple and similar fragments of EcoRI-digested H. halobium NRC-1 DNA. Clone V147, which must contain two

Fig. 17. Autoradiogram showing hybridization of individual ^{32}P -labelled recombinant plasmid DNAs of a genomic library of H. volcanii to EcoRI digested total DNA from H. volcanii (panel A) or H. halobium NRC-1 (panel B). The library was constructed by ligating EcoRI digested total H. volcanii DNA to similarly digested pBR322.

Closed triangles in panel A indicate faint bands detected on the original autoradiogram. Small closed circles indicate EcoRI fragments of H. volcanii DNA hybridized by several H. volcanii cloned DNAs and also by two H. halobium cloned DNAs (see Fig. 18). Open triangles adjacent to clone v88 indicate EcoRI fragments of H. volcanii DNA also hybridized by three H. halobium cloned DNAs (see Fig. 18).

Closed triangles in panel B indicate EcoRI fragments of H. halobium DNA probed by several H. volcanii cloned DNAs and also probed by H. halobium clone 85 (see Fig. 15).



different repetitive elements, probed these same fragments and one or two others. Clone V122 probed a unique set of H. halobium strain NRC-1 fragments. More unexpectedly, clone V88 probed the same set of H. halobium NRC-1 fragments as did clone V7, V33, V50, V60, V97 and V147, although it does not probe the same set of H. volcanii fragments. Thus, repeated elements which are usually separated in H. volcanii are usually associated in H. halobium.

In a reciprocal experiment (Fig. 18), selected EcoRI clones of H. halobium NRC-1 DNA which probed multiple EcoRI fragments of that DNA were used to probe EcoRI-digested H. volcanii DNA. Clones 54 and 85 (Fig. 15) probed identical sets of H. volcanii DNA fragments, which appear to include all those fragments probed by H. volcanii clones V88 and V97 (or others like it), plus some additional high molecular weight fragments. This provides reciprocal confirmation of the conclusion just reached. These clones (clones 54 and 85) contain inserts of only 3.3 and 1.75 kbp, respectively. However, H. halobium NRC-1 clone 84 probes only those fragments probed by H. volcanii clone V88, so this clustering is not obligatory in H. halobium. Of the remaining seven H. halobium NRC-1 clones tested, four (37, 40, 51 and 91) probed unique sets of H. volcanii Eco RI fragments and three probed no H. volcanii fragments.

G. Rearrangements affecting H. halobium repetitive sequences

Fig. 19 shows results obtained by probing EcoRI-digested (panel A) or Sall digested (panel B) DNAs from strain NRC-1 (track 1), two NRC-1 single colony isolates picked from a single plate as phenotypically indistinguishable from NRC-1 (tracks 2 and 3), two spontaneously arising gas vacuole-deficient NRC-1 variants (tracks 4 and 5) and strain R1

Fig. 18. Autoradiogram showing hybridization of H. halobium NRC-1 recombinant plasmid DNAs (see Fig. 15) to EcoRI digested H. volcanii DNA. Closed circles and open triangles denote fragments also hybridized by H. volcanii cloned DNAs (see Fig. 17).

36 37 40 51 55 54 83 84 85 91

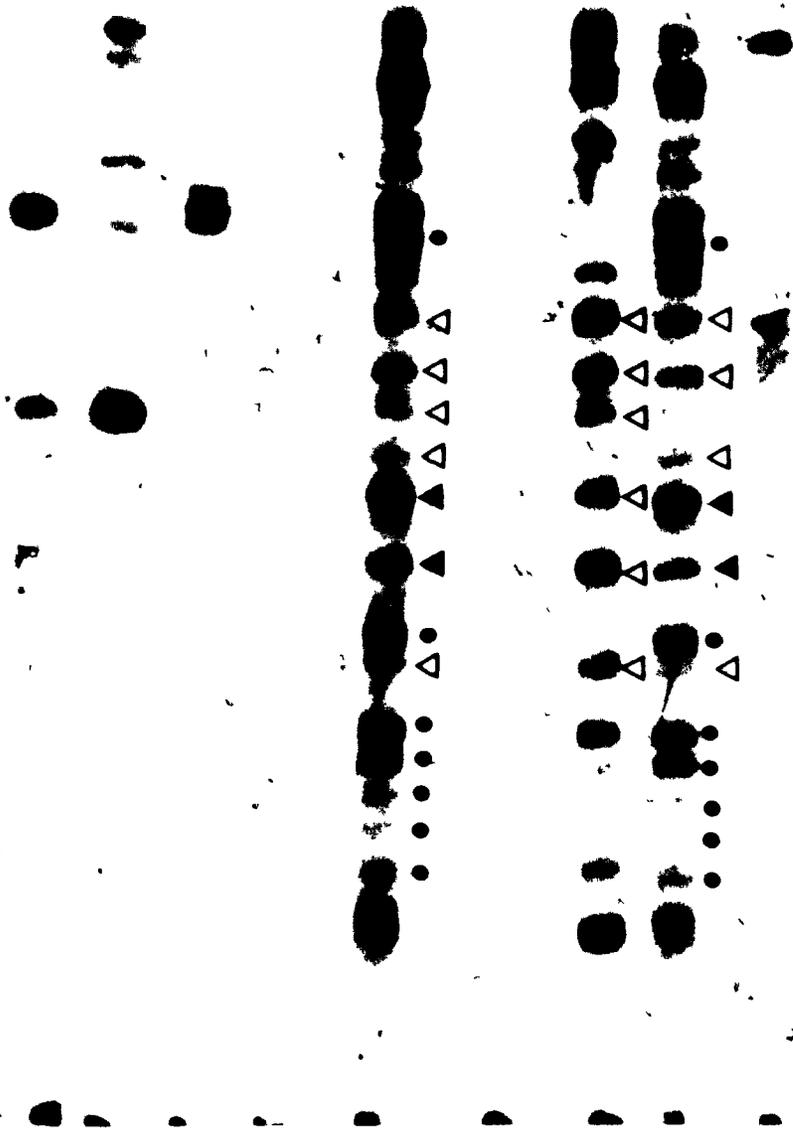


Fig. 19.A. Autoradiogram showing hybridization of a ^{32}P -labelled 330 bp AluI fragment of H. halobium NRC-1 DNA to EcoRI cleaved total DNA from H. halobium NRC-1 (lane 1), two apparently wild-type colonies of strain NRC-1 selected from 20 such isolates (see text) (Lanes 2 and 3), two gas vacuole-deficient mutants of strain NRC-1 (such mutants arise at a frequency of $1-3 \times 10^{-2}$) (lanes 4 and 5), and H. halobium strain R-1 (lane 6). Lane M contains ^{32}P -labelled λCI857 Sam7 DNA cleaved with HindIII. The 330 bp AluI fragment used as a probe appears to comprise most of a ca. 500 bp repetitive element present in one copy on each of two HindIII-EcoRI cloned fragments of strain NRC-1 DNA which do not otherwise cross-hybridize (Fig. 20).

B. Hybridization of a ^{32}P -labelled 900 bp PstI-SalI strain R-1 cloned DNA fragment to SalI cleaved DNAs from different H. halobium strains. Lane designations as in A.

A

M 1 2 3 4 5 6



B

M 1 2 3 4 5 6



Fig. 20. Autoradiogram showing hybridization of clone 7 DNA to AluI cleaved clones 7 and 14 DNAs. Plasmid DNAs from clones 7 and 14 were isolated on CsCl gradients containing ethidium bromide, cleaved with AluI, subjected to electrophoresis in a 2.2% agarose gel and transferred to a nitrocellulose filter. Hybridization probe was a 3.5 kbp EcoRI halobacterial DNA insert in clone 7, isolated preparatively by electrophoresis in 1% agarose (low melting point agarose, Bethesda Research Laboratories), labelled with ^{32}P and hybridized as described in Materials and Methods. Lane 1: clone 14 DNA; lane 2: clone 7 DNA. Autoradiogram was overexposed to show hybridization to 160 bp fragment.



(track 6) with two DNAs bearing different and single repetitive elements. The first (panel A) is a 330 bp AluI fragment which appears to comprise most of a ca. 500 bp repetitive element present in one copy on each of two cloned fragments of NRC-1 DNA which do not otherwise cross-hybridize (Fig. 20). The second (panel B) is a ca. 900 bp fragment containing an element present on one of two otherwise identical HindIII fragments of the H. halobium R1 plasmid (Fig. 6). (This element hybridizes to at least a dozen fragments in the genome of H. volcanii and several H. salinarum DNA fragments [Fig.21].) None of the six DNAs probed produces identical hybridization patterns with either of these two probes, even though the cell populations used to prepare these DNAs derive from phenotypically identical colonies ("wild-type" for tracks 1, 2 and 3; gas vacuole deficient for tracks 4, 5 and 6) and were, except for strain R1, separated from each other by only that number of generations (ca. 30) required to produce enough cells to prepare DNA. Mobility of repeated elements (even randomly selected repeated elements) is remarkably high.

H. Nucleotide sequence of an H. volcanii repeated element

Eubacterial transposable elements have characteristic structures consisting of a central region containing one or more structural genes, regulatory information and transposition determinants, bounded by inverted or direct repeats of varying length and sequence homology (Kleckner, 1981). In order to determine whether halobacterial repeated elements have such characteristic structures, DNA sequence studies were done on two cloned H. volcanii EcoRI fragments bearing members of the same repeated sequence family. Figure 17A shows that clone V97 hybridizes to a subset of H. volcanii EcoRI fragments probed by clone V147.

Fig. 21. Hybridization of ^{32}P -labelled PstI-SalI fragment from clone IV (shaded area in Fig. 6) to SalI cleaved total DNA from H. salinarum (lane 1) and H. volcarii (lane 2).



Both clones also hybridize to several H. halobium EcoRI fragments. The 1200 bp EcoRI fragment in clone V147 was cleaved with Sau 3A and sub-cloned into the BamHI site of the M13mp7 phage vector (Messing et al., 1980). DNAs from phage plaques were fixed to a nitrocellulose filter and probed with ³²P-labelled clone V97 plasmid DNA. One of the phage DNAs which hybridized, designated V147 7S9, was selected for Southern blot analysis and nucleotide sequencing by the Sanger dideoxy interrupted synthesis method (Messing et al., 1980). The 800 bp EcoRI fragment from clone V97 was re-cloned into the EcoRI site of M13mp7 to facilitate nucleotide sequencing. (All nucleotide sequences were kindly provided by Dr. Jason Hofman).

Figure 22 shows the results of hybridizing clones V97 and V147 7S9 (both ³²P-labelled by uninterrupted synthesis) to DNAs from H. halobium, H. salinarum, H. volcanii, H. trapanicum, H. vallismortis and H. saccharovorum. Clone V147 7S9 probes all fragments hybridized by clone V97 in all species, indicating the repeated sequence present on clones V97 and V147 is contained within the 300 nucleotides represented in clone V147 7S9. No other Sau 3A fragments of clone V147 are probed by clone V97 (data not shown). This sequence is also affected by rearrangements in the H. halobium genome (Figure 23). Such rearrangements are discussed in more detail in the following section.

Comparison of the nucleotide sequences so far obtained from clones V97 and V147 7S9 shows a region of approximately 88% homology extending over some 60 bases (Figure 24). The region of homology may be extended over 100 bases if numerous insertions and deletions are allowed; however, these extended regions are only 64% homologous. If the 60 bases which are 88% homologous are considered to be the repeated element, the

Fig. 22. Hybridization of ^{32}P -labelled clones V97 or V147 7S9 to EcoRI cleaved total DNA from H. halobium strain NRC-1 (lanes A), H. salinarum (lanes B), H. volcanii (lanes C), H. trapanicum (lanes D), H. vallismortis (lanes E) and H. saccharovorum (lanes F).

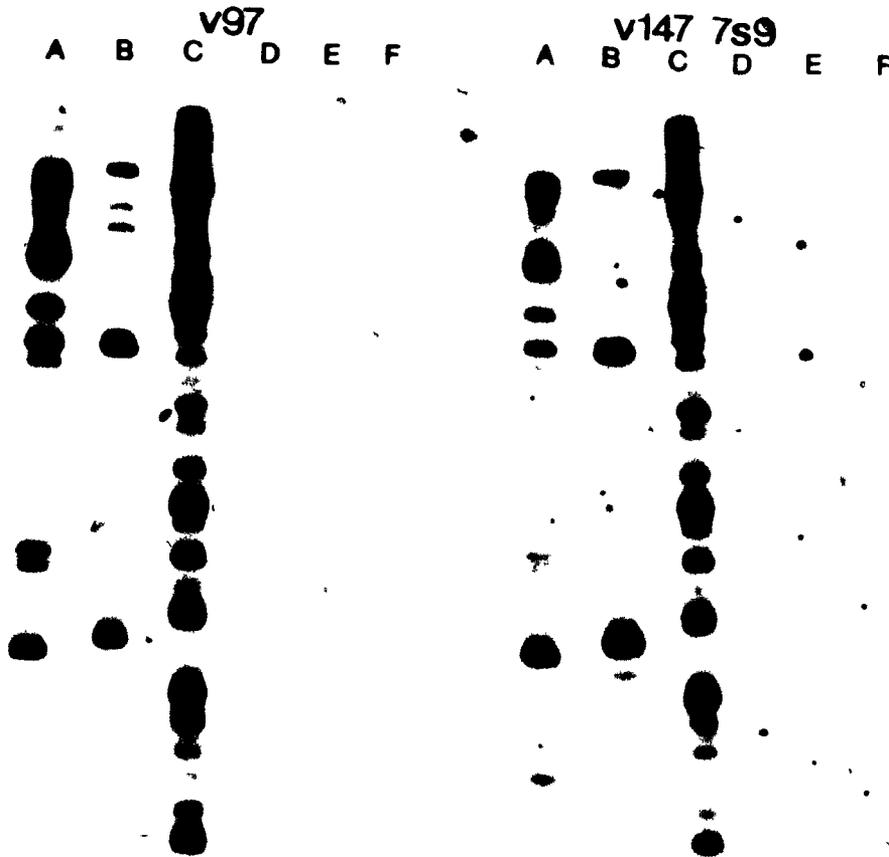
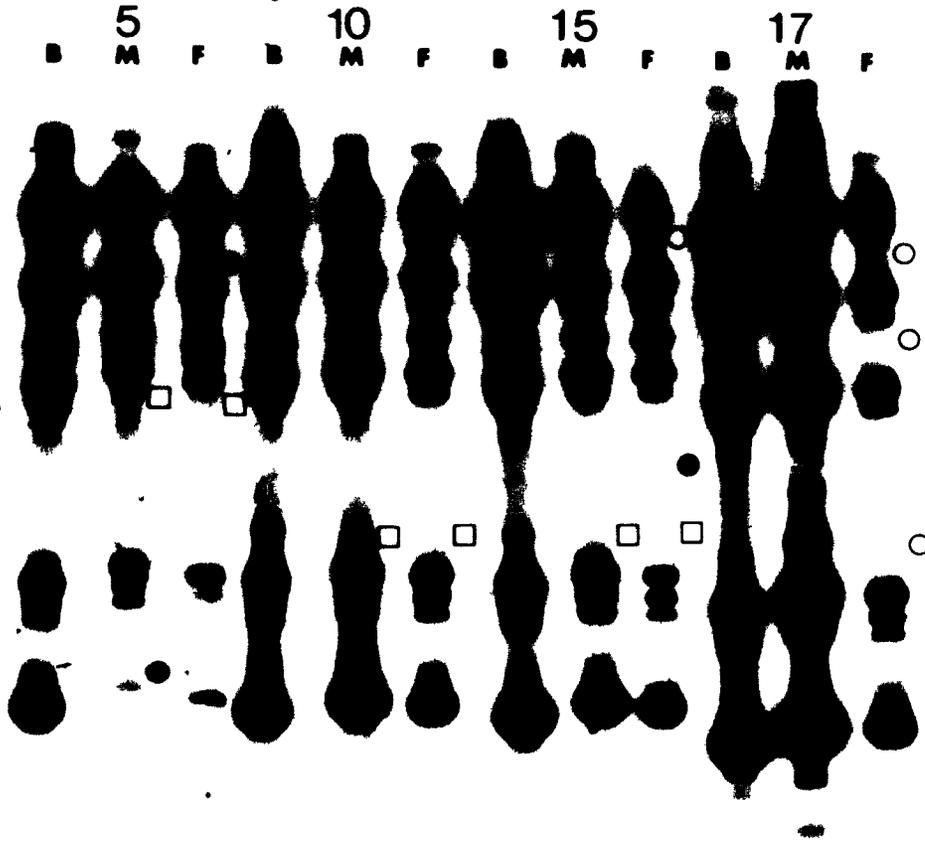


Fig. 23. Hybridization of ^{32}P -labelled clone V147 7S9 to EcoRI cleaved DNA from four independent wild-type isolates of strain NRC-1 (5, 10, 15 and 17). M and F denote DNA samples prepared from each isolate after several additional single colony isolations from B. Filled squares denote hybridizing fragments in M and F not present in B. Open squares denote fragments hybridized in B and not present in M or F. Closed circles denote fragments hybridized in only M or F. Open circles denote fragments missing from F but present in B and M.

v147 7s9



■ ■ ■ ■

Fig. 24. Partial nucleotide sequences of clones V97 and V147 7S9. Bold-face type indicates regions of homology, allowing insertions and deletions. Thick lines over sequences indicate the presence of the hexanucleotide, CCTCGT (or extended versions), thin lines indicate the presence of the complement dry sequence, GGAGCA (or extended versions).

v 97 --- AGCGC **CGACGAGGACCTCGTAG** ATG **TCTT** 30
v147 7s9 GCGC A GAGG CCTCGTAG **TATG** CTT

v 97 CAGC AC **CGATTTTCAGCGTTTTTCAGCGAGA** 60
v147 7s9 CAGC **TACGAGTTTCAGCGTTTTTCAGCGAG**

v 97 gTGAGTGCA **ACTTCC** TGTC **AAGGGAGTTG** 90
v147 7s9 TGAGTGCA TTxxT GT AGGG TG

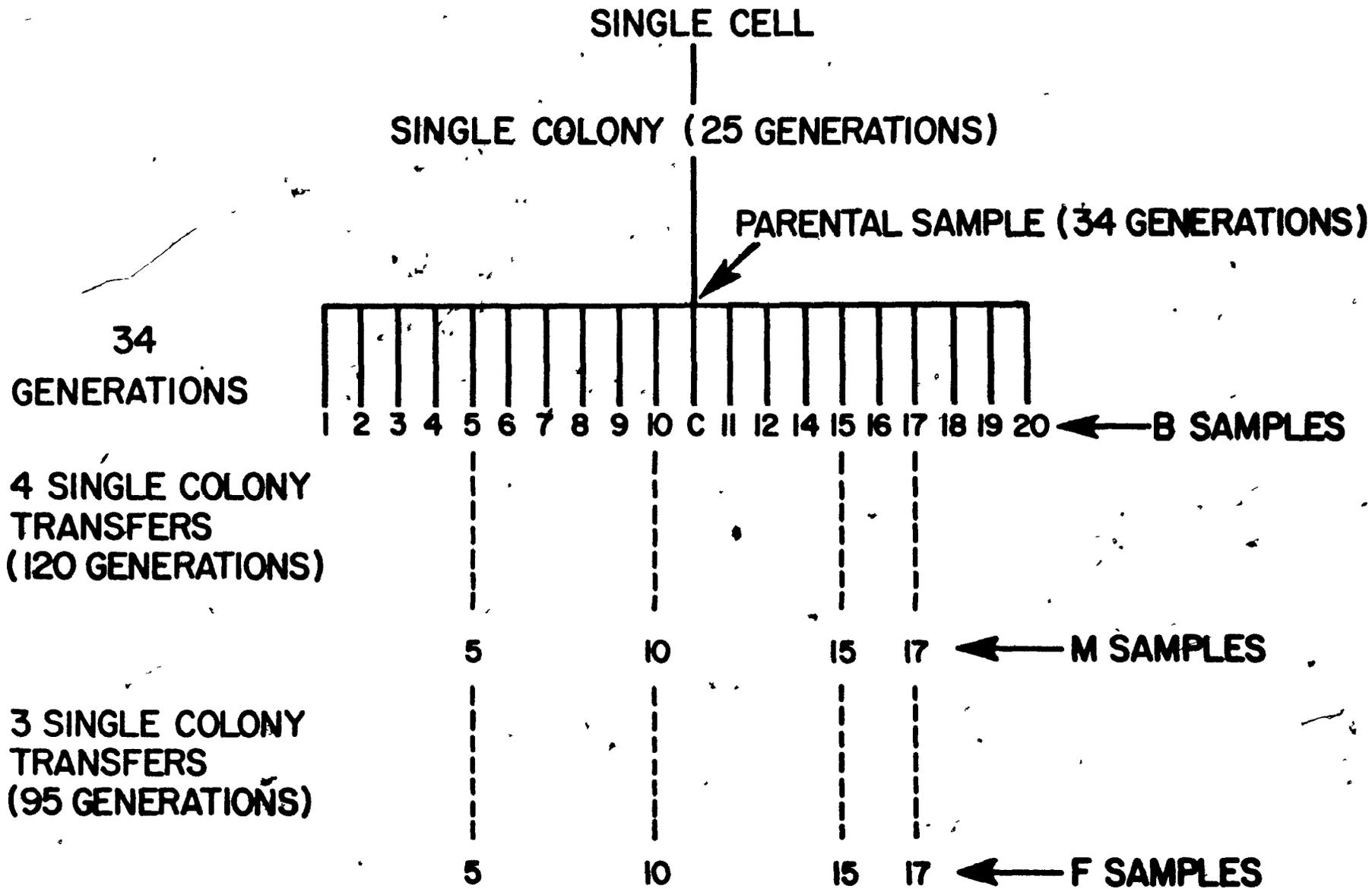
v 97 AcGAGGAAGTTAAGGAGCTGGTCCTCGTGG 120
v147 7s9 A GAG A TT AGG ---

element is unusually small to be independently transposable. It may represent only one end of an element flanked by these 60 bases. Further sequencing should elucidate this possibility. Within the 120 bases of clone V97 shown in Figure 24, the hexanucleotide CCTCGT is repeated four times and its complement, GGAGCA, three times. Several other repeats with these hexanucleotides at their core are also present within these 120 bases. Numerous complex secondary structures may be drawn by base pairing these regions but the significance of such structures is unknown. Neither of the sequences shown in Fig. 24, nor their complements, appear to be able to adopt a transfer RNA-like secondary structure.

I. Unselected rearrangements affecting the position of repeated elements in *H. halobium*

The experiment outlined in Figure 25 was designed to measure rates of spontaneous genetic rearrangements involving repeated elements, in cellular lineages diverged from a single common ancestor for a minimum number of generations. That minimum is set at about 34 by requirements for single colony formation and enough subsequent generations to prepare sufficient DNA for Southern blot analysis. A culture containing approximately 2^{34} cells, was established from a single phenotypically wild-type colony formed by a single *H. halobium* strain NRC-1 cell. Most of this culture was used to prepare DNA taken to be representative of the genome of the initial "parental" cell (tracks labelled P in Figures 26-34). The other portion of the culture was diluted to approximately 10^3 cells \cdot ml $^{-1}$ and plated on solid medium to give single colonies. Some of these showed visible phenotypic alterations (Vac $^{-}$, Pum $^{-}$ or Rub $^{-}$). Nineteen colonies which did not were randomly selected and cultured in

Fig. 25. Flow diagram for experiment designed to measure unselected rates of rearrangement affecting the position of repeated sequences (see text). Nineteen wild-type isolates were established from a population derived from a single cell after that cell had undergone 34 generations. Four of these isolates were carried through seven additional single colony isolations.



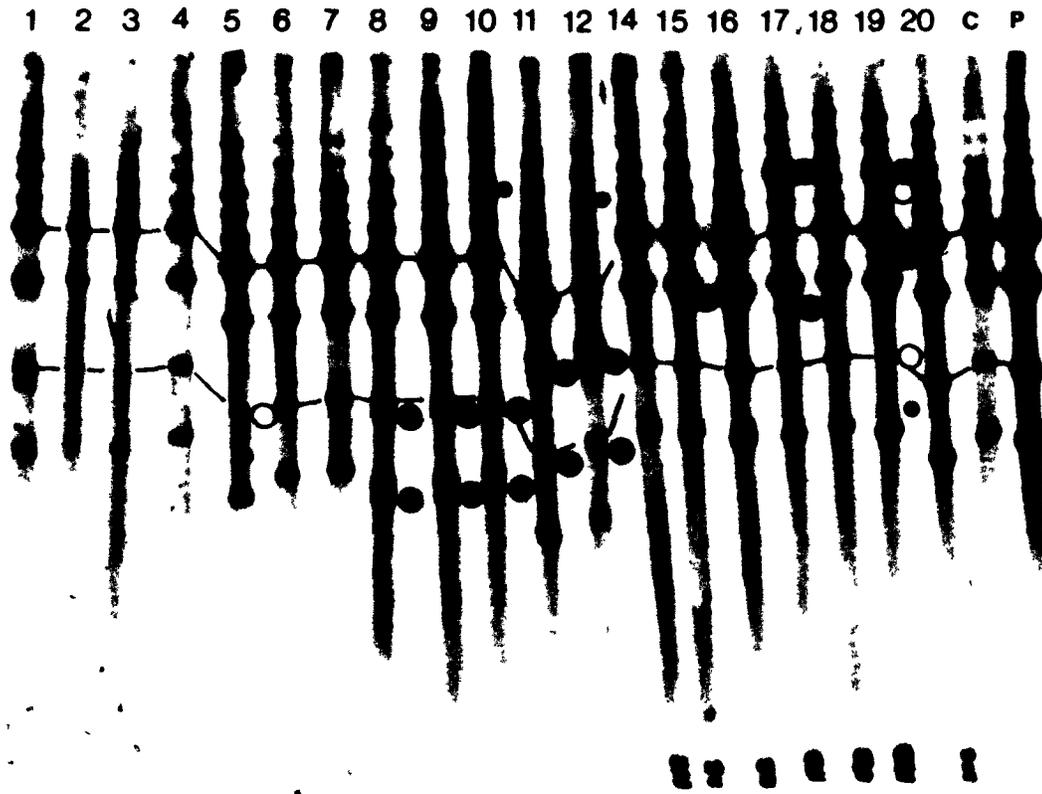
10 ml of liquid medium to approximately 2^{34} cells. DNA was prepared from these cultures and used for the Southern blot analyses in Figures 26-34.

Figures 26-33 show results of hybridizing ^{32}P -labelled H. halobium strain NRC-1 DNA probes (cloned as EcoRI fragments into pBR322), bearing members of 8 distinct repeat sequence families, to EcoRI-digested DNAs from each of these nineteen isolates. Many changes in both number and position of hybridized fragments are apparent. Closed circles denote the position of fragments probed in addition to those present in the parental sample. Open circles denote the absence of fragments probed in the parental sample. Some of the hybridization probes share common sequences ("shared subpatterns", as noted in Figure 15) and therefore detect some of the same changes in hybridization pattern, but each probe also hybridizes to fragments unique to that probe. For example, both probes 7 (Figure 26) and 37 (Figure 27) detect the same additional fragments in isolates 15, 17 and 19, but probe 7 detects a fragment missing from isolate 19 whose absence is not detected by probe 37.

The molecular processes giving rise to the observed changes in hybridization pattern are unknown. However, the appearance of a new fragment unaccompanied by the disappearance of an old fragment is formally analogous to duplicative transposition (isolate 20 in Figure 30; isolate 8 in Figure 31 for examples). The disappearance of a fragment unaccompanied by the appearance of a new fragment is formally analogous to deletion of a repeated element (isolate 5 in Figure 26; isolate 19 in Figure 29 for examples). The disappearance of a fragment accompanied by the appearance of a new fragment (isolate 9 in Figure 28; isolate 9 in

Figs. 26-34. Nitrocellulose filter-bound EcoRI cleaved DNAs from 19 different wild-type, strain NRC-1 isolates (lanes 1-12, 14-20) diverged from the parental isolate (lane P) by 34 generations, probed with ^{32}P -labelled cloned repeated (Figs. 26-33) or unique-sequence (Fig. 34) DNAs from strain NRC-1. Lane C contains DNA derived from the parental cell culture after an additional 34 generations (see Fig. 25). All probes are EcoRI fragments except for unique sequence (Fig. 34) DNA probe which contains two PstI fragments. Lines between lanes join fragments of the same molecular weight.

Fig. 26. Probe: clone 7.



⑦

Fig. 27. Probe: clone 37.

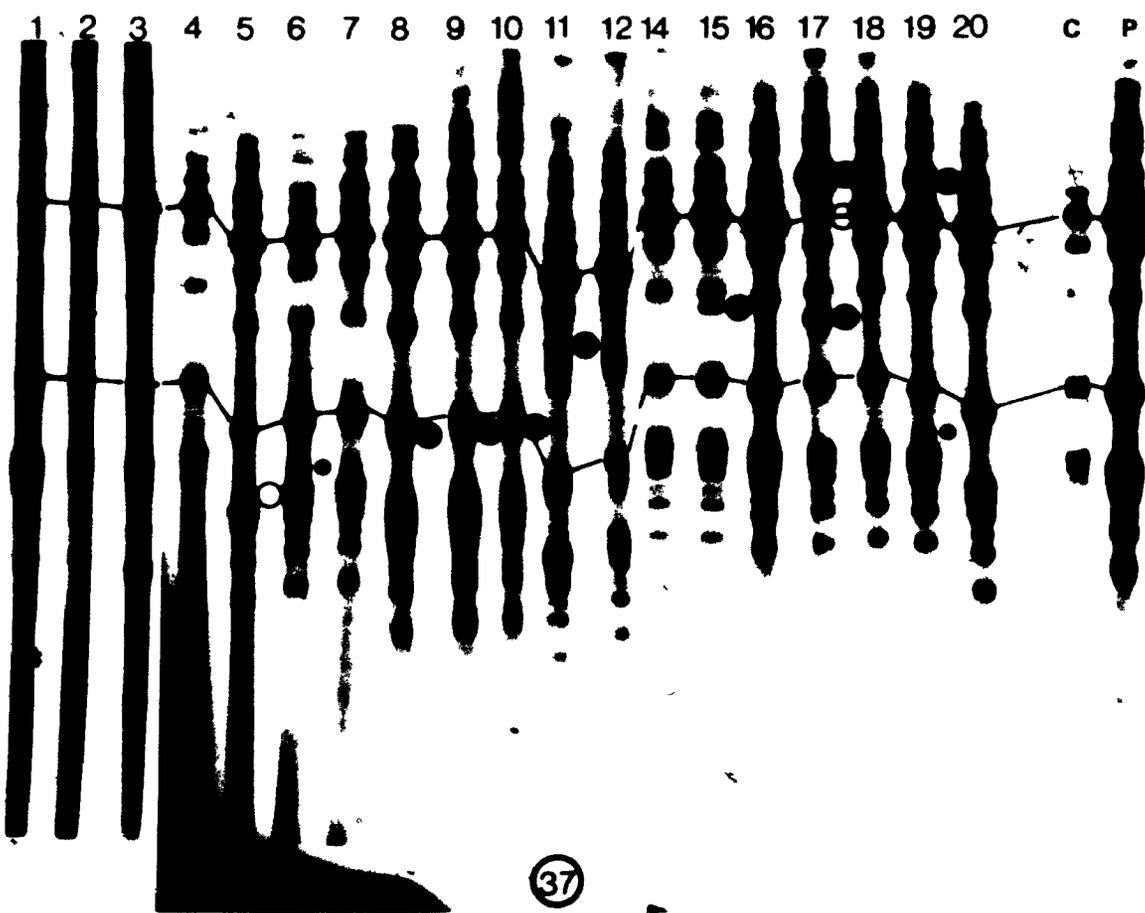
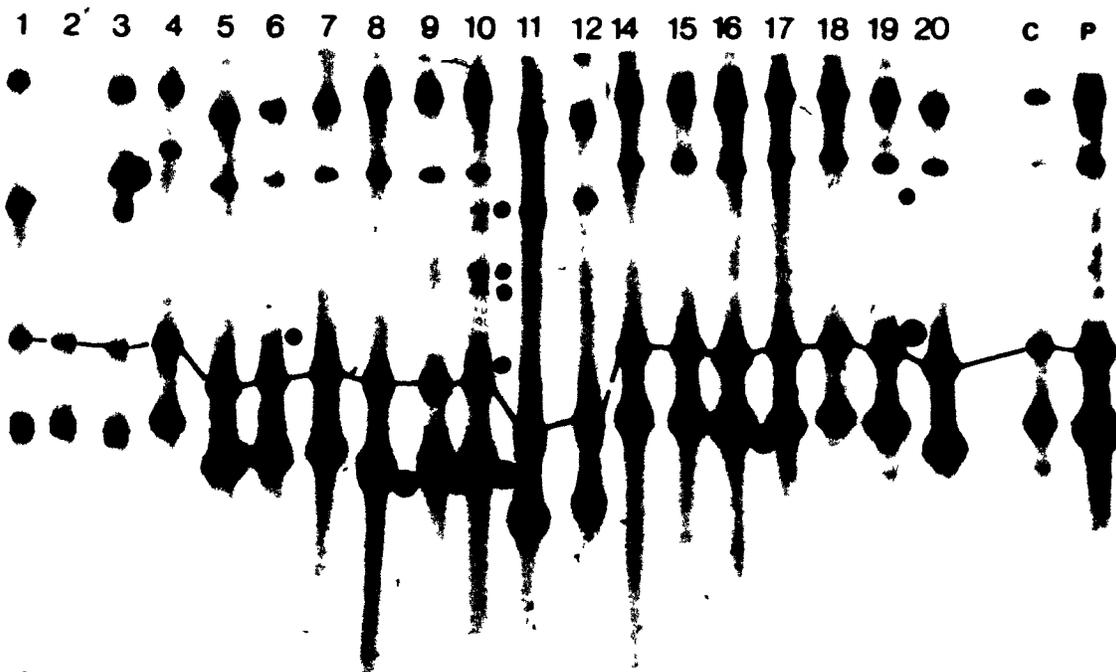


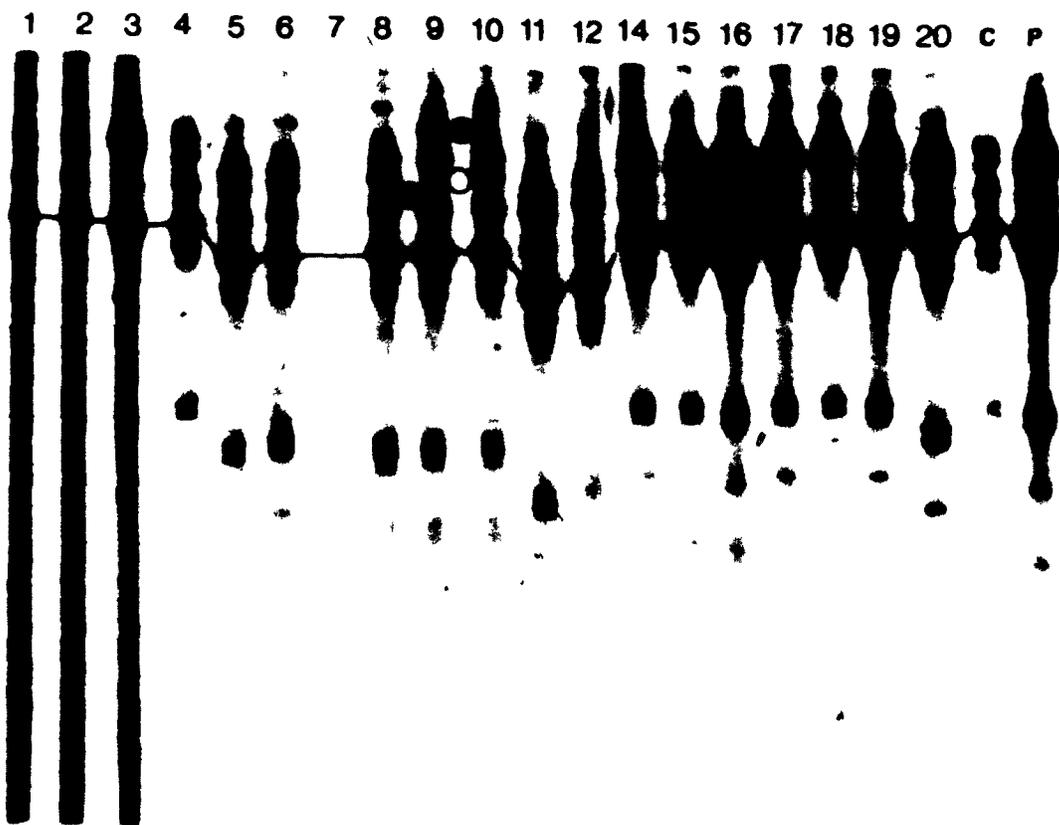


Fig. 28. Probe: clone 43.



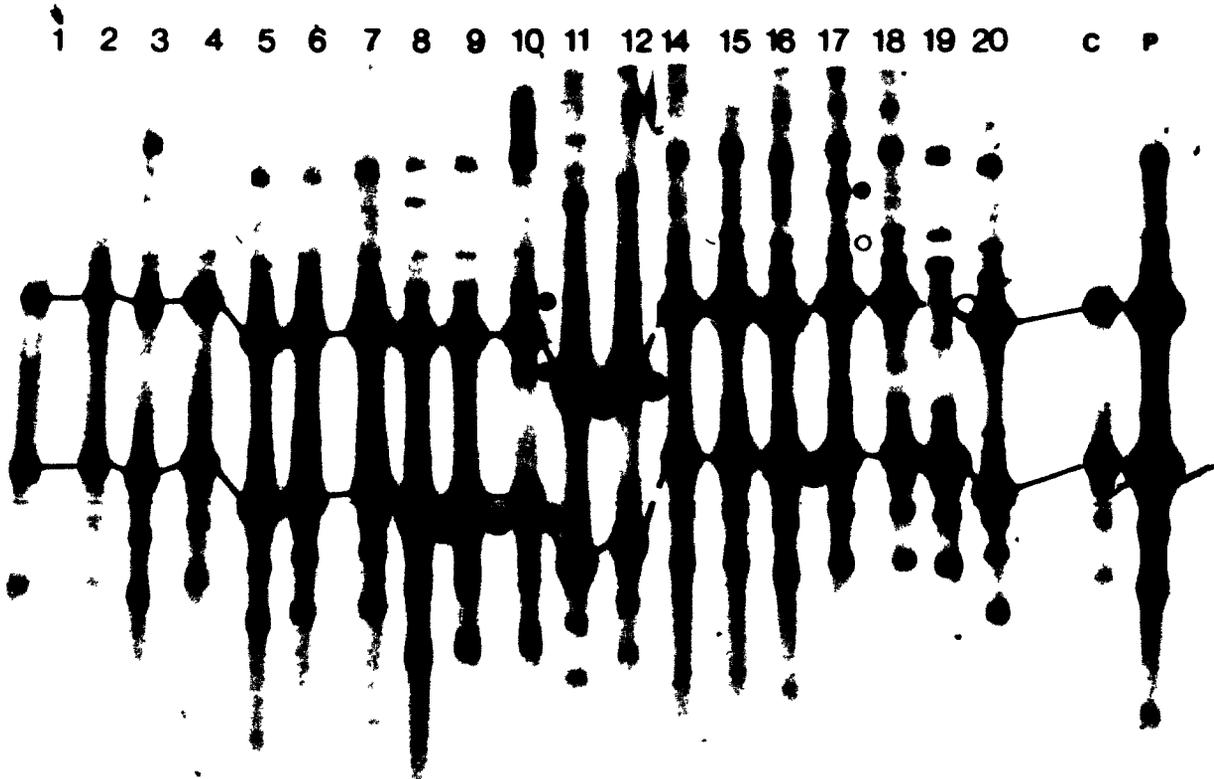
43

Fig. 29. Probe: clone 54.



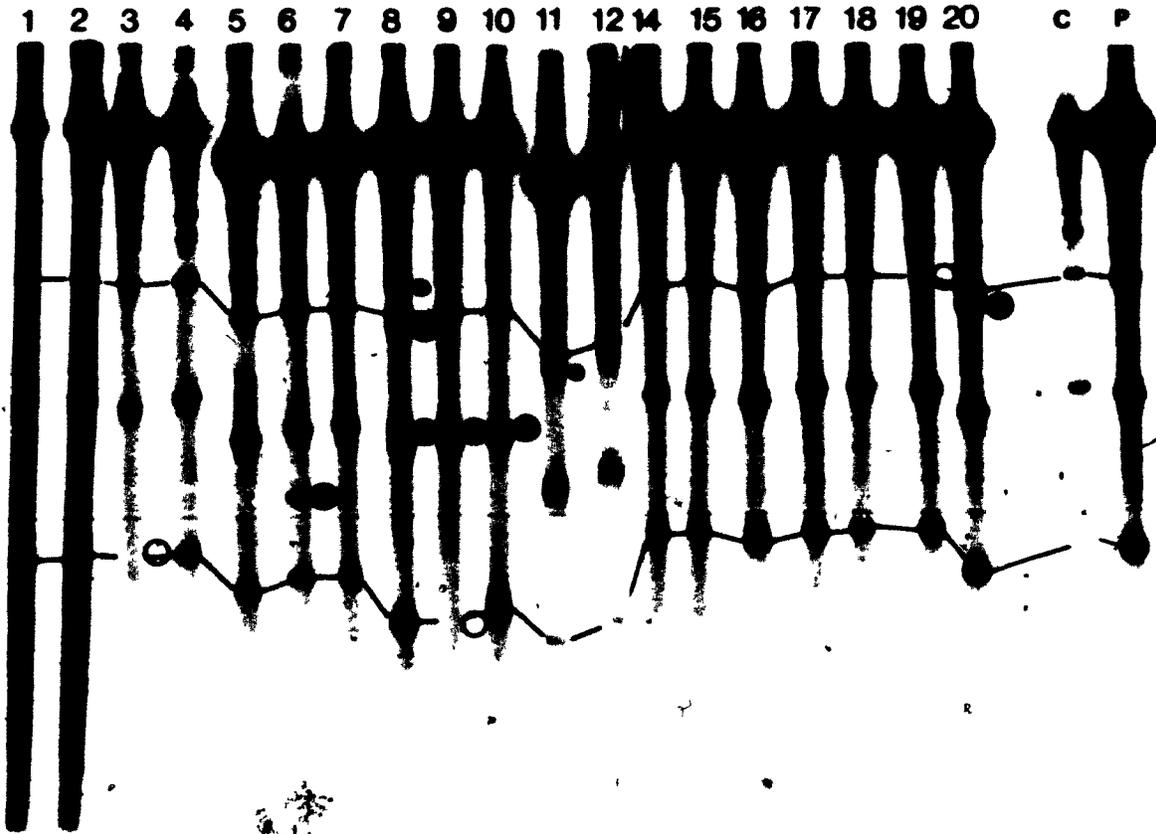
54

Fig. 30. Probe: clone 74.



74

Fig. 31. Probe: clone 84.



34

Fig. 32. Probe: clone 85.

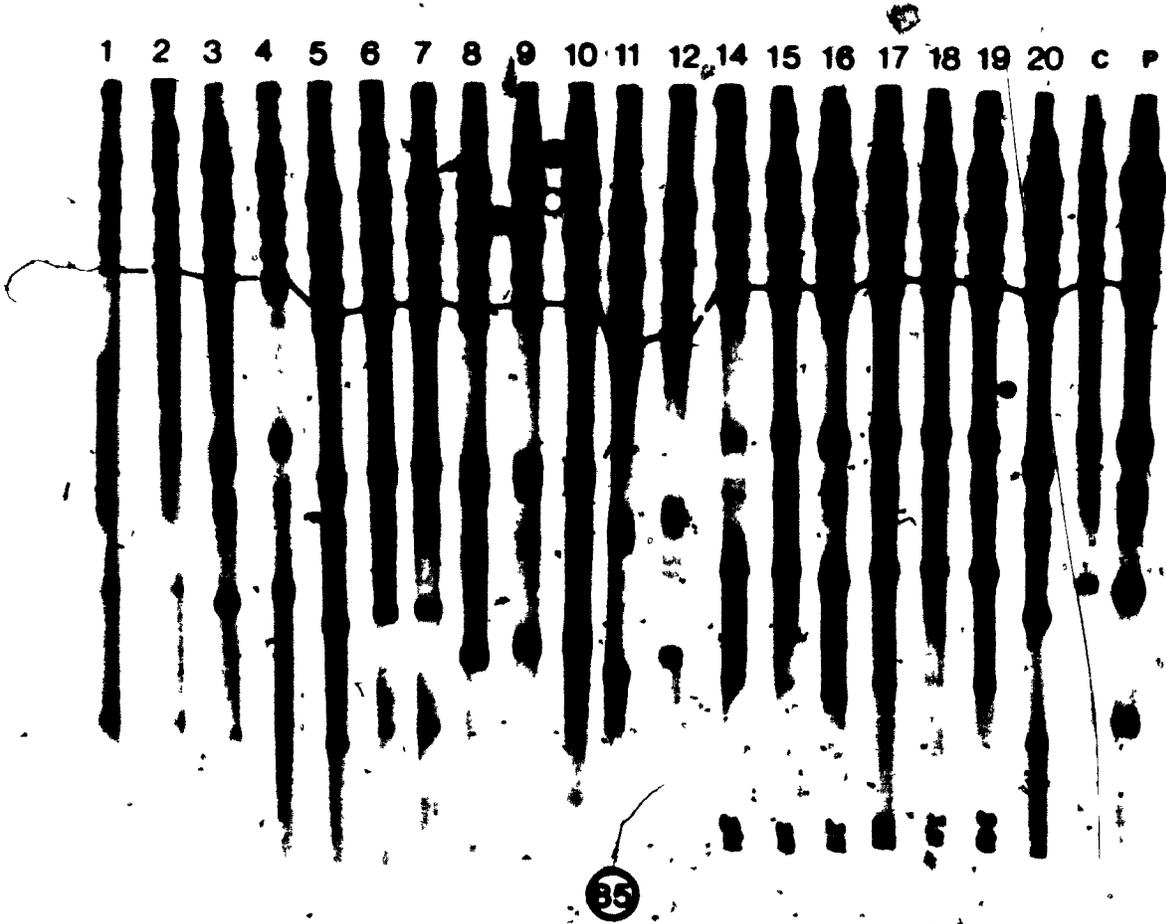


Fig. 33. Probe: clone 91.

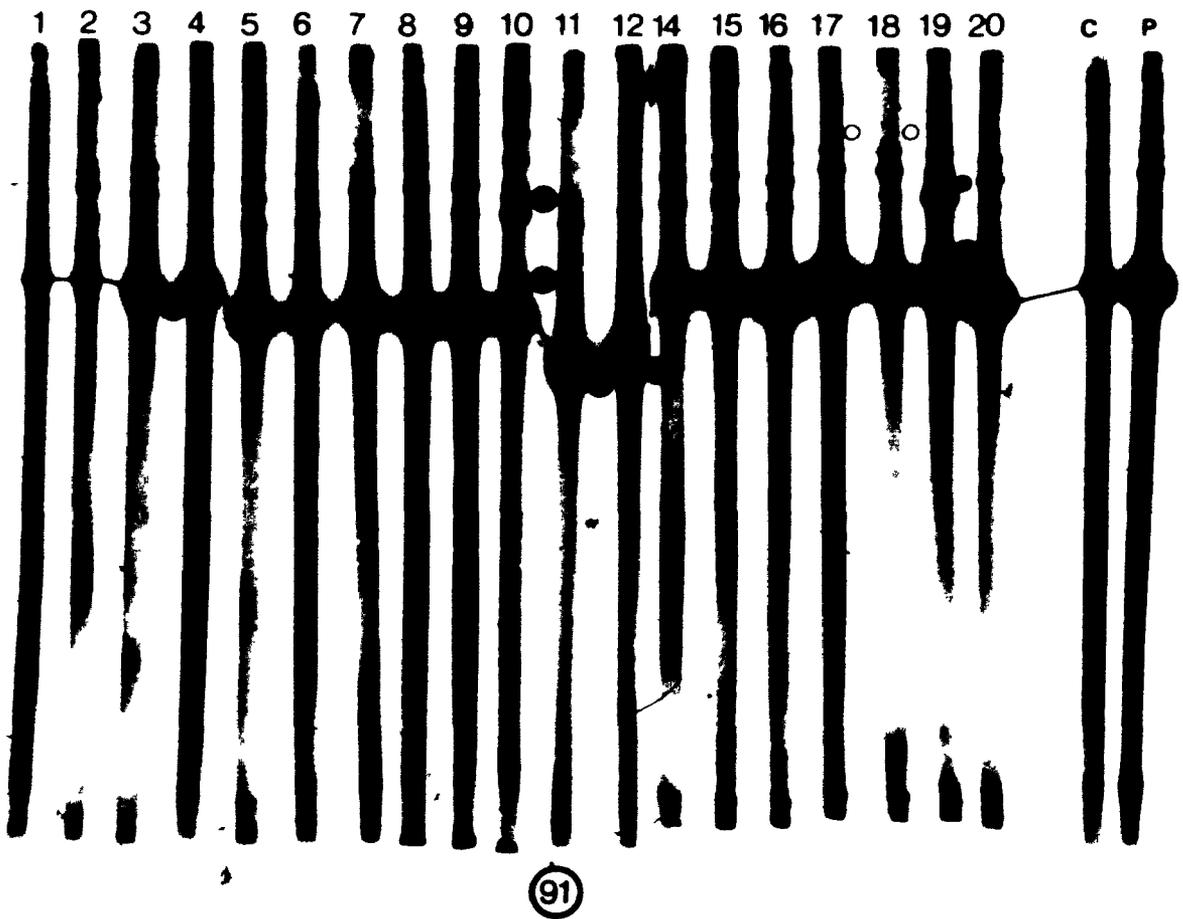


Fig. 34. Probes: clones p41 and p108.

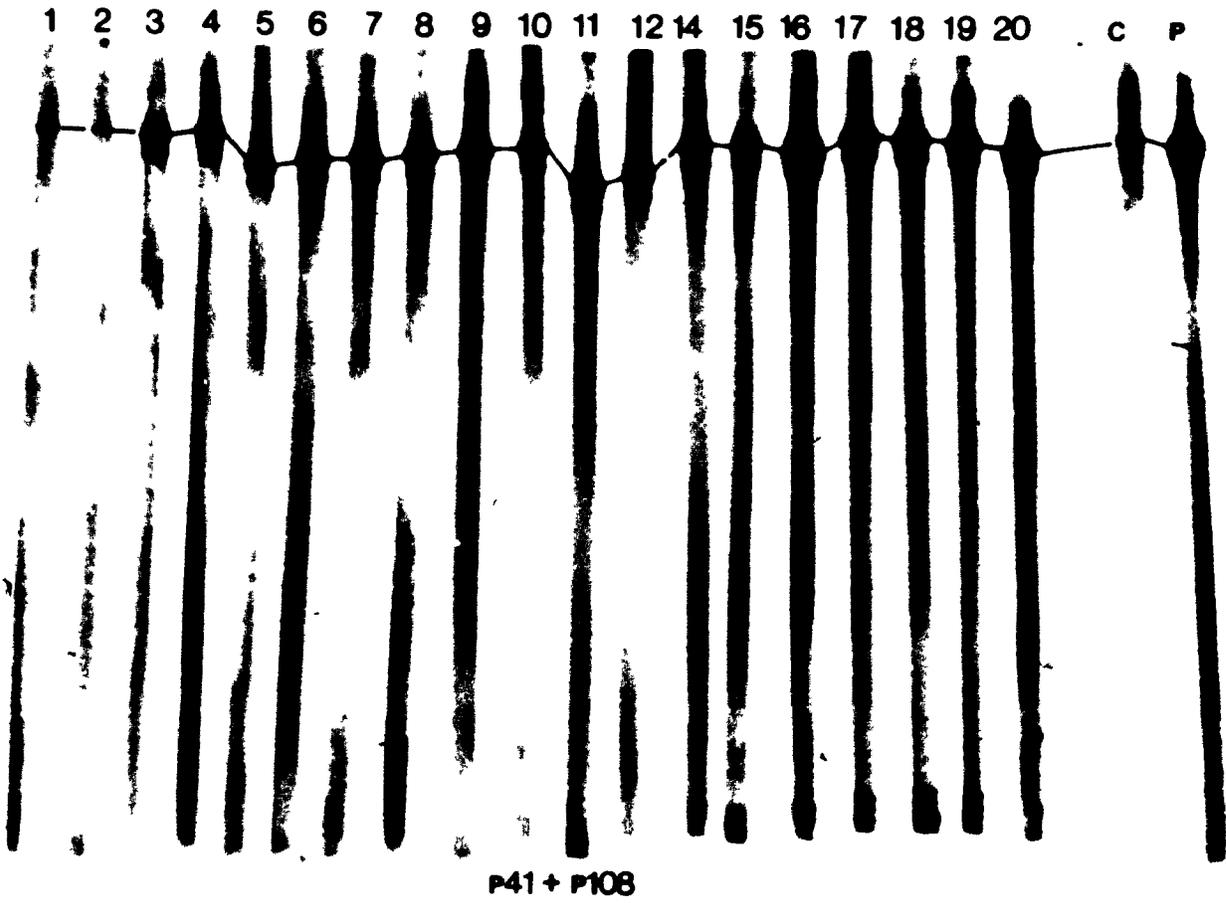


Figure 30 for examples) is formally analogous to excision and reinsertion of an element, but could also reflect independent transposition and deletion-like events. Reciprocal recombination would result in the appearance of two new fragments accompanied by the loss of two pre-existing fragments (of the same combined molecular weights). No events of this type have been observed.

If each appearance or disappearance of a fragment is defined as an event, there are 57 events among the 19 isolates which are uniquely detected by one or another of the 8 repeat sequence probes. All repeat sequence probes detect probe-specific alterations in at least one of the isolates, and there is no obvious reason to doubt that all repeat sequences are affected by, or effect, genomic rearrangements at roughly comparable frequencies. On the other hand, the two unique-sequence DNA probes (Figure 34) detected no changes in any of the 19 isolates.

If genomic rearrangements involving repeat sequences occur independently, then the observed frequency of events ($57 \cdot 8 \text{ repeat sequence families}^{-1} \cdot 19 \times 34 \text{ cell generations}^{-1} = 0.01 \cdot \text{repeat sequence family}^{-1} \cdot \text{cell generation}^{-1}$) could be taken as the probability of such rearrangements in any cellular lineage. However, the 57 events may not be distributed randomly among the 19 isolates. Four of the isolates (8, 9, 10 and 19) together account for about half of the events and isolate 19 shows changes with six of eight repeat sequence probes.

If all events are completely independent, then the probability of not observing a change with any particular repeat sequence family probe is $(1-0.01) = 0.99 \cdot \text{repeat sequence family}^{-1} \cdot \text{cell generation}^{-1}$. The probability of not observing a change with eight repeat sequence family probes in 34 cell generations becomes $[(1-0.01)^8]^{34} = 0.065$. This

implies that only one isolate out of nineteen should have had no detectable changes when eight repeat sequence families were used as probes. In fact, seven of the isolates (1, 2, 4, 6, 14, 16 and 18) showed no changes with the eight probes. If it is assumed that one event is usually coupled with others, one may calculate the probability of that first event, p , by using the observed frequency of isolates exhibiting no changes with any of the eight repeat sequence family probes using the zero-term of the binomial distribution (Feller, 1968): frequency of isolates showing no changes = $7/19 = 0.368 = [(1-p)^{34}]^8$. This gives a value for p of 0.004 events·repeat sequence family⁻¹·cell generation⁻¹. A similar calculation, using the zero-term of the Poisson formula $p(0, \lambda t) = e^{-\lambda t}$, where λ gives the probability of an "event" occurring in one unit of time, yields a λ value of 0.00420 . This is effectively equivalent to the binomial distribution estimate of p as 0.00419 (Feller, 1968). This estimate is not remarkably lower than the previous estimate (0.01 events·repeat sequence family⁻¹·cell generation⁻¹) and may be considered a minimum value for p because events appear to be only partially coupled. Using a value for p of 0.004 ·repeat sequence family⁻¹·cell generation⁻¹, one would predict that about one-tenth of the isolates will experience rearrangements affecting at least one of the eight repeat sequence families during the first three or four divisions after their establishment from single cells. Such rearrangements will give rise to restriction site polymorphisms detectable as weakly hybridizing fragments in the DNA prepared after 34 generations. Such weak hybridization signals were indeed detected with several isolates and several probes (isolate 19 in Figure 28, for example; fragment denoted by small closed circle) and were, as far as possible, excluded

from the above calculations. Polymorphisms of this sort should be especially common in DNAs prepared from batch cultures established after serial passage of small liquid inocula, and probably account for the weak hybridization signals detected in Figure 15.

Of the nine EcoRI fragments probed strongly by the repeat sequence family member borne by clone 7, five have been deleted in one or another of the isolates shown in Figure 19A or Figure 26. This may suggest that all members of at least this repeat sequence family are equally liable to suffer genomic rearrangement. To address the related question of whether the "new" hybridization patterns produced by rearrangement are as stable as "old" ones, four of the nineteen isolates (5, 10, 15 and 17) were carried through an additional seven single colony isolations. DNA was prepared after the fourth (120 generations), samples labelled M in Figures 35-42, and seventh (215 generations), samples labelled F, platings. These DNAs and samples of DNA from the 34th generation (identical to those used in Figures 26-34 and labelled B in Figures 35-42) were probed with seven cloned H. halobium repeat sequence DNAs and three cloned repeat sequence DNAs from H. volcanii, previously shown to be homologous to three different H. halobium repeat sequence families (Figure 17B). Some of these results are shown in Figures 35-41. There are three instances in which a fragment hybridized in the parental sample was lost in one of the four isolates during the first 34 generations (and thus absent from B DNA samples). In none of these cases was a fragment of identical mobility regained during 215 additional cell generations. Vacated sites appear not to be selectively reoccupied (for example, see open triangles for isolate 5, Figure 35). There are four instances in which a hybridizable fragment was gained in one of the four

isolates during the first 34 generations and retained through the next 120 generations (and thus present in both B and M samples). In three of these four instances, the "new" fragment was retained through 215 generations (for examples, isolate 17 in Figure 35; isolate 10 in Figure 37; fragments indicated by filled triangles). This frequency of loss is not substantially greater than that expected from calculations presented above, and thus "new" fragments are not uniquely unstable. Figures 35-41 provide further examples of genomic rearrangements (hybridizing fragments present or absent from M or F DNAs which were not present or absent from the preceding DNA sample). No changes were detected in any of the isolates using three unique sequence DNA probes (Figure 42).

One other estimate of the frequency of rearrangements affecting repeat sequence families may be made from this experiment. No changes were detected in isolate 10 between the M and F samples (95 generations) with any of the seven repeat sequence family probes. p may be calculated from the observed frequency of no changes as $0.25 = [(1-p)^{95}]^7$, or $p = 0.002$ events·repeat sequence family⁻¹·cell generation⁻¹. This value is reasonably close to, but because of the small sample size, less reliable than the value of 0.004 events·repeat sequence family⁻¹·cell generation⁻¹ previously obtained.

J. Implications of genomic rearrangement at high frequency

Comparison of the estimate for H. halobium genomic rearrangements to transposition frequencies reported for E. coli transposable elements (10^{-4} - 10^{-7} per colony forming unit; Calos and Miller, 1980; Kleckner, 1981) is complicated by the fact that the frequencies reported for E. coli elements are derived from selected events at particular genetic loci. Read and Jaskunas (1980) reported rates for unselected

Figs. 35-42. Nitrocellulose filter-bound EcoRI cleaved DNAs from isolates 5, 10, 15 and 17 probed with ^{32}P -labelled cloned repeated (Figs. 35-41) or unique-sequence (Fig. 42) DNAs from strain NRC-1. Closed triangles indicate the position of fragments not present in the parental DNA sample (lanes P in Figs. 26-34) but present in B, M and F DNA samples. Open triangles indicate the absence of fragments in the B, M and F DNA samples which were present in the parental DNA sample. Filled squares denote hybridizing fragments in the M and F DNA samples not present in B DNA samples. Open squares denote fragments hybridized in B DNA samples but not present in M or F. Closed circles indicate fragments hybridized in only M or F DNA samples. Open circles denote fragments missing from F but present in B and M DNA samples.

Fig. 35. Probe: clone 37.

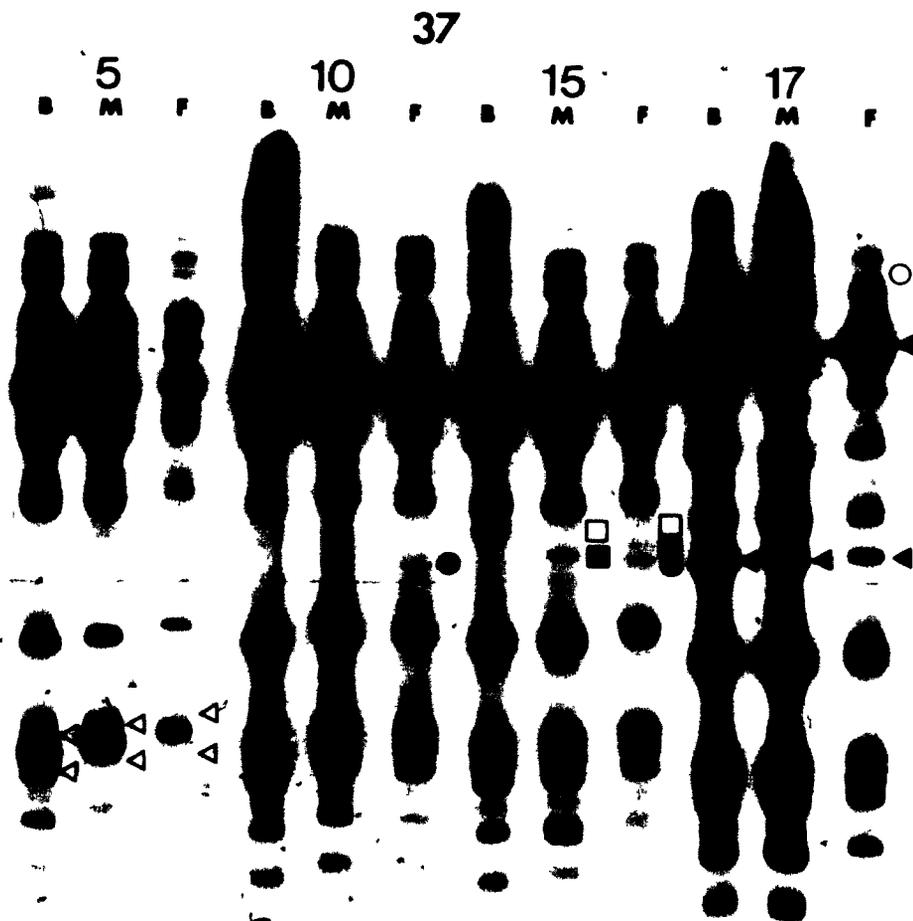


Fig. 36. Probe: clone 43.

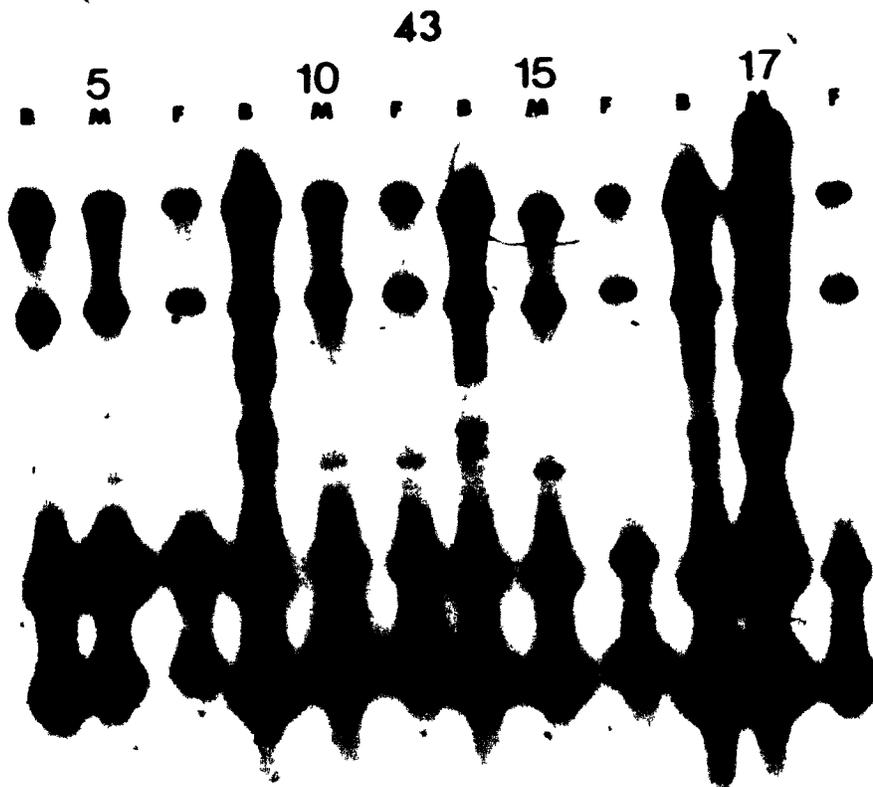


Fig. 37. Probe: clone 74.

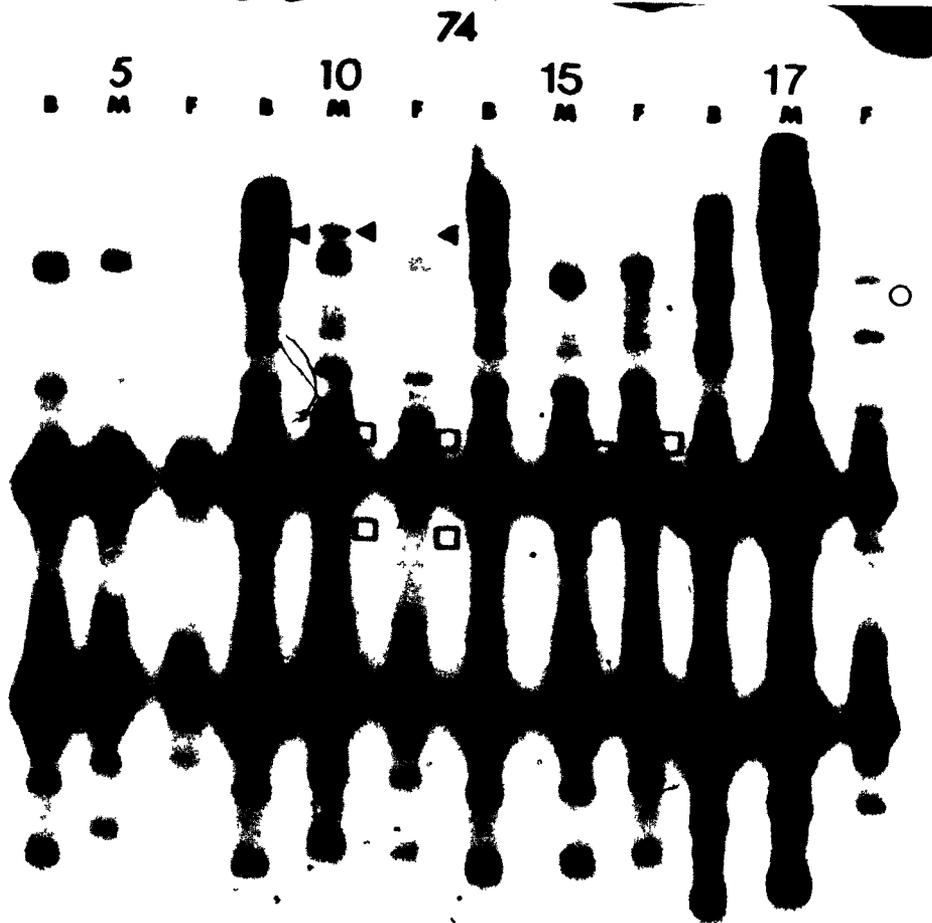


Fig. 38. Probe: clone 85.

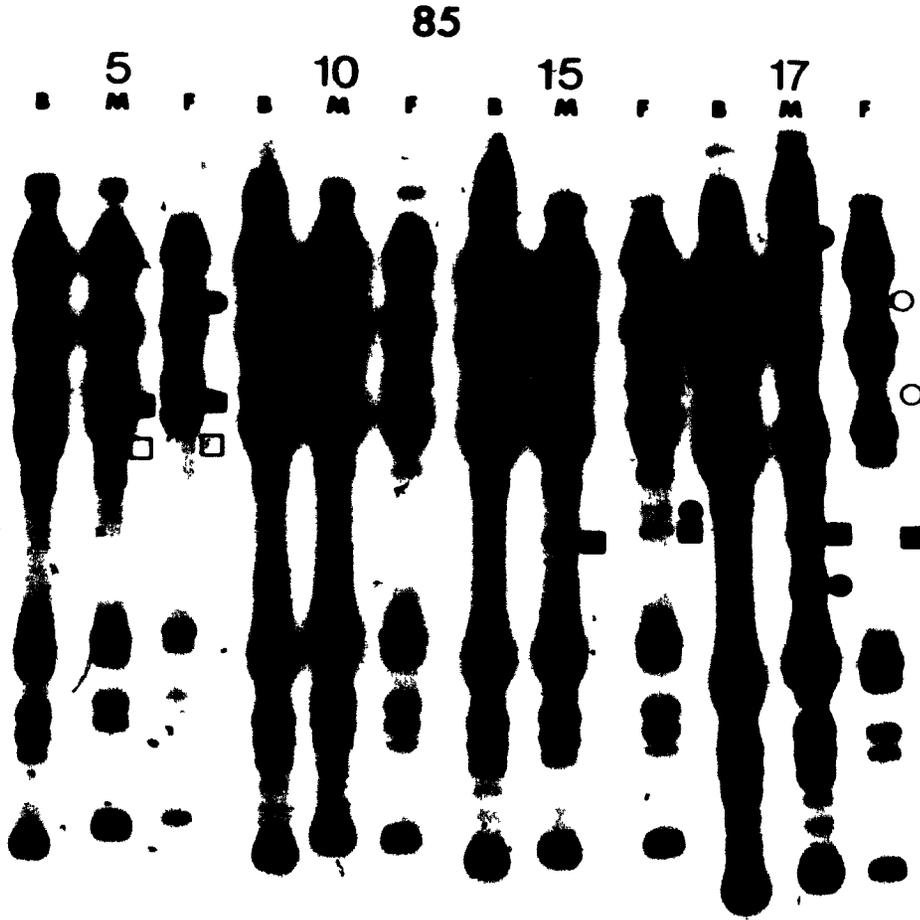


Fig. 39. Probe: clone v88.



Fig. 41. Probe: clone v122.

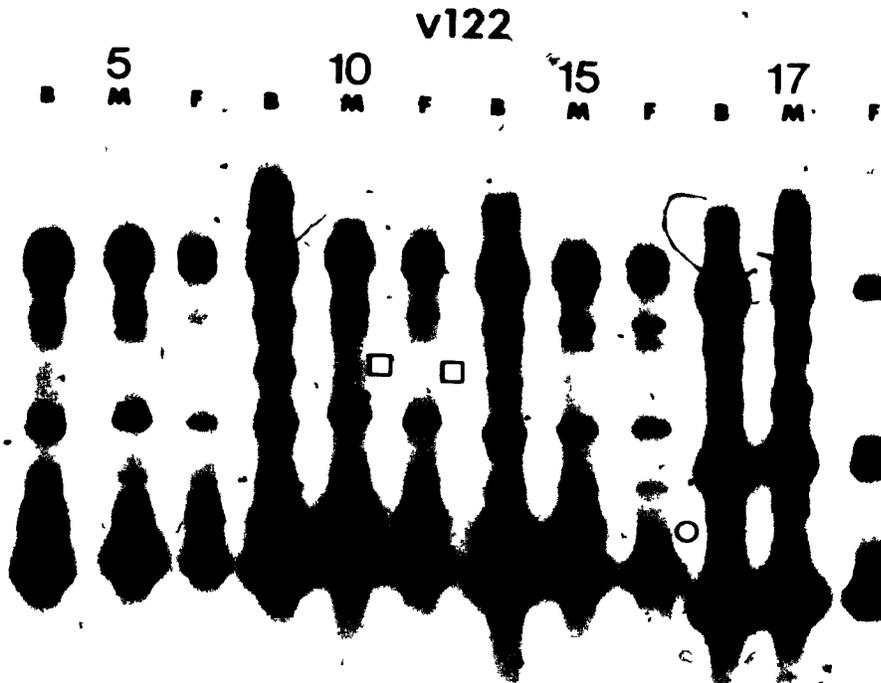
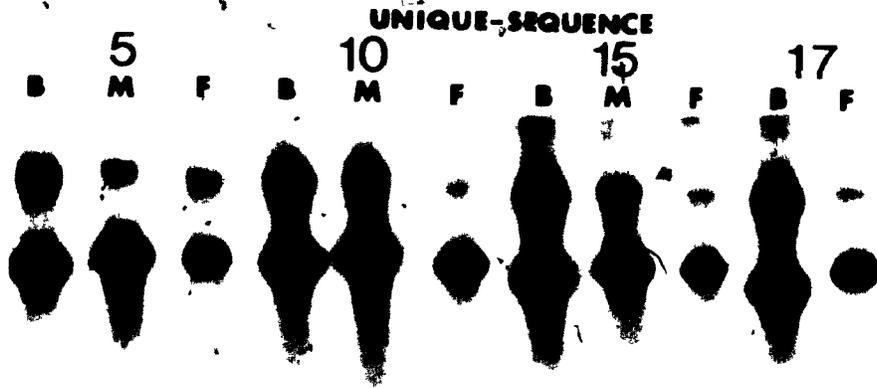


Fig. 42. Probes: clones p27, p41 and p97.



transposition of IS1 near those reported here for H. halobium repeat sequence families. Out of 40 mutants selected to contain an IS1 insertion into the spc operon, three had additional, unselected IS1 transpositions and one of these had three additional IS1 transposition events. Assuming the forty isolates they examined were separated by 25 generations, p may be calculated as: frequency of no unselected insertions = $37/40 = 0.925 = (1-p)^{25}$ which gives $p = 0.003$ transpositions-IS1 family⁻¹.cell generation⁻¹. These authors consider the observed level of unselected IS1 transposition unusually high and attribute it to bursts of transposition occurring from the same donor site or the possibility that one transposition event triggers others. If the latter possibility is the case, then the value of 0.003 for p would be an overestimate because all 40 isolates were selected on the basis of an IS1 transposition into the spc operon. The observation that one of the 40 isolates contained three unselected transpositions of IS1 does imply one event sometimes triggers others. A rough idea of the degree to which this estimate is in error may be obtained from a comparison of the spontaneous mutation frequencies of each organism. The frequencies of spontaneous mutation to vac⁻, pum⁻ and rub⁻ for H. halobium are 10^{-2} , 10^{-4} and 10^{-4} per colony forming unit respectively (Pfeifer et al., 1981). Various amino acid auxotrophs are obtained at frequencies of 10^{-4} - 10^{-5} per colony forming unit (Weber et al., 1982 and personal communication). In contrast, amino acid auxotrophs of E. coli are obtained at frequencies of approximately 10^{-7} per colony forming unit (Cox, 1976). The frequency of the first IS1 insertion (into the spc operon) measured by Read and Jaskunas [1980]) was 4×10^{-7} per colony forming unit. Even if only 1/10 of mutations in H. halobium are caused by genomic rearrangements, the higher frequency of spontaneous mutation implies at

least a 10 fold greater rate of rearrangements involving repeated sequences in H. halobium.

A rather striking consequence arises from such frequencies of genomic rearrangement involving 50 repeat sequence families. For the average one liter culture (started from a single cell) containing 10^{11} cells, the probability that any two cells are of the same genotype (undergoing no events with any of the 50 families) is 3×10^{-4} . This calculation assumes that no two events are identical (an "infinite allele" model - Ewens, 1979), which may not be strictly true if there are "hot spots" like those observed for some E. coli transposable elements (Johnsrud et al., 1978); nevertheless, such a level of genotypic variability is remarkably high.

The conclusion that H. halobium exhibits a high degree of genetic variability has also been reached by Weber et al. (1982 and personal communication) and has led to the unusual proposal that H. halobium does not regulate gene expression in the classical sense, but does so by mutation. More simply stated, the high degree of genetic variability maintained in H. halobium allows a large number of genotypes to be represented in any cell population. On the basis of chance alone, some of these genotypes will be better able to cope with any given change in the environment. These genotypes will therefore be selectively increased in the population, but will continue to generate new genotypes at a high level. When another change in the environment takes place, some of the newer genotypes will be of greater fitness.

Such a system can operate only if the rate of formation of new genotypes is great enough to ensure representation of all genetic loci in a viable, altered state in one or more members of any cell population. The size of the H. halobium cell population required to ensure

representation of 3000 genetic loci (assuming H. halobium and E. coli have similar numbers of loci) in an altered state may be calculated from the value of $p = 0.004$ events \cdot repeat sequence family⁻¹ \cdot cell generation⁻¹. If there are 50 families of repeated sequences, the probability of a change involving any repeat sequence family in any one generation is 0.2.

Assuming all events are different and lead to viable genotypes, the number of new genotypes created in generation n is $0.2 \sum_{i=1}^n 2^i$. The total number of new genotypes created in the cell population by generation n will be: $0.2 \cdot 2^1 + 0.2 \cdot 2^2 + 0.2 \cdot 2^3 + \dots + 0.2 \sum_{i=1}^n 2^i$ or $0.2 \sum_{i=1}^n 2^i$. Thus, the number of generations required to produce 3000 new genotypes is 13, or a total cell population of only ca. 8.2×10^3 . In reality, many or most events may give rise to non-viable genotypes; also the probability that new events are identical to previous events will increase as the number of genotypes in the population increases. As a worst case scenario, if only one out of ten events is functionally different and only one in a thousand unique events leads to a viable genotype, the number of generations required to obtain at least 3000 unique, viable genotypes is $0.2 \cdot 0.1 \cdot 10^{-3} \sum_{i=1}^n 2^i = 3000$ or $n=28$ generations (2.7×10^8 cells). This is a relatively modest number of cells for an organism capable of reaching very high population densities and correspondingly large total population sizes in nature (Kushner, 1978). In contrast, an E. coli population approximately 100 times this size would be required to generate 3000 unique, viable genotypes if p is assumed to be 0.0002 for each of ten families of repeated elements and similar assumptions are made about the number of unique, viable genotypes. Although cell populations greater than this size are commonly encountered in the laboratory, such

population sizes are very near those thought to exist in the human intestinal tract (Slack and Snyder, 1978).

Although it is possible, on numerical grounds, for H. halobium to "regulate" gene expression by mutation (i.e., not regulate gene expression), it is by no means clear that they do so. Only a few reports implying control of gene expression at any level have been published (Rogers and Morris, 1978; Spudich and Stoeckenius, 1980) and, in these cases, it is unclear what fraction of the cell populations examined are responsible for the responses observed. Studies on the control of gene expression at the level of transcription, using cloned halobacterial genes (of which only two defined genes are as yet available; the 16S-23S-5S-ribosomal RNA gene cluster of H. volcanii [Woese and Doolittle, unpublished] and the bacteriorhodopsin gene of H. halobium [Dunn et al., 1981]) should elucidate whether control mechanisms like those available in E. coli are also present in Halobacterium species.

A sequence capable of forming an inverted repeat structure reminiscent of the operator regions of several E. coli operons has been found at the beginning of the bacteriorhodopsin gene, and may function as a binding site for a regulatory protein (Dunn et al., 1981).

Dunn et al. (1981) have also found several regions 5' to the coding sequence of the bacteriorhodopsin gene which consist of alternating purines and pyrimidines. As they note, such sequences may adopt a Z-DNA structure (Wang et al., 1979; Arnott et al., 1980) under the high intracellular salt concentrations of H. halobium (4 M KCl). Z-DNA has been shown to exist in vivo in Drosophila by fluorescent antibody staining of polytene chromosomes (Nordheim et al., 1981). In these chromosomes, the Z-DNA is restricted to interband regions (Nordheim et al., 1981) and has

been implicated in the control of gene expression. Control of gene expression in H. halobium may be mediated by such structures.

The fraction of rearrangements involving repeated sequences which has significant phenotypic effect remains to be determined. This will be difficult in the absence of standard methods for genetic analysis, but there seem to be three alternatives, for each of which there is tentative experimental support. (i) Repeat sequence elements are confined to silent regions of the genome, and rearrangements affecting, or effected by, them are without phenotype effect. Indeed, most H. halobium repeat sequence elements reside in (presumably relatively A:T-rich) regions of both plasmid and chromosomal DNA which are relatively rich in EcoRI sites and relatively poor in PstI sites (see Section III E). (ii) Repeat sequence element-associated rearrangements are random and often detrimental, and H. halobium endures them because they are unavoidable or because group selection for high genetic variability maintains them (Sapienza and Doolittle, 1981). Pfeifer et al. (1981b) observed high rates of spontaneous mutation affecting gas vacuole and pigment production, some of these mutations being associated with complex and not easily interpreted alterations in plasmid restriction endonuclease digestion patterns. Schnabel et al. (1982) found restriction site polymorphisms in the DNA (most of which is presumably essential) of preparations of the halobacterial phage Φ H grown on H. halobium. (iii) Repeat sequence element-associated rearrangements reflect the operation of a multiplicity of complex and reversible transpositional switch mechanisms with specific effects on gene expression. Several of the independent gas vacuole-deficient variants of Pfeifer et al. (1981b)

seem to have suffered, among other events, specific insertions into plasmid DNA.

These three views are not mutually exclusive. Repeat sequence element-associated instability in the H. halobium genome seems astonishingly high. It is not unlikely that the majority of rearrangements are confined to silent regions of the genome (both plasmid and chromosome) while some result in random (presumably detrimental) gene inactivation and still others have been recruited to operate as regulatory switch mechanisms.

K. Evolutionary conservation of repeated sequences among the Halobacteria

Seven EcoRI fragments of H. halobium DNA (24.7 kbp total) bearing members of seven different repeat sequence families, three EcoRI fragments of H. volcanii DNA (4.3 kbp total) bearing different repeat sequence families common to both H. volcanii and H. halobium, and three PstI fragments (7.0 kbp total) bearing unique-sequence DNA of H. halobium, were selected from the pBR322 cloned genomic libraries described in previous sections. These recombinant plasmids were ³²P-labelled in vitro and used to probe EcoRI cleaved nitrocellulose filter-bound DNAs from H. halobium, H. salinarum, H. volcanii, H. trapanicum, H. vallismortis and H. saccharovororum. Hybridizations and subsequent washing were done under stringent conditions (Jeffreys and Flavell, 1977). The results of this experiment are shown in Figures 43-53.

Four factors can potentially complicate the interpretation of these results: evolutionary conservation of unique-sequence DNA flanking cloned H. halobium repeats; acquisition or loss of EcoRI sites within repeated elements; drift in element sequence past the point where

hybrids are stable at high stringency; and polymorphisms due to high-frequency genomic rearrangements within populations from which DNAs are made (genomic instability of repeated sequences is a diagnostic feature as shown in the previous section). These complicating factors cannot invalidate the following conclusions: (i) There is no halobacterial species which does not share several repeat sequence families with H. halobium. (ii) There is no halobacterial species which appears uniquely favoured in its possession of many different H. halobium repeats, with the possible exception of H. salinarum, often considered conspecific with H. halobium. (The H. volcanii results are biased by the inclusion of three probes (V88, V97 and V122) bearing repeat sequences already known to be shared with H. halobium.) (iii) There is no H. halobium repeat sequence family which is not represented in at least two other halobacterial species. (iv) There is no repeat sequence family which is uniquely favoured in its representation or uniformly high copy number. (v) Unique-sequence DNAs are less highly conserved than are repeat sequence DNAs. This is more convincingly demonstrated by the fact that none of 13 PstI clones bearing in sum approximately 50 kbp of H. halobium unique-sequence DNA hybridizes to H. volcanii DNA at high stringency, (data not shown) and neither do H. volcanii unique sequence DNAs hybridize to H. halobium DNA (Fig. 17B), while all but one or two of the H. halobium repeat sequence probes (in sum 24.7 kbp) hybridize to H. volcanii DNA, some of them quite strongly.

Given this variability in family size and phylogenetic distribution, it seems unlikely that any single halobacterial repeat sequence family is essential for survival. Their presence could be accounted for by interspecific transfer, although no mechanism for such transfer is

Figs. 43-53. Hybridization of ^{32}P -labelled cloned H. halobium strain NRC-1 repeated (Figs. 43-49), H. volcanii repeated (Figs. 50-52) or H. halobium strain NRC-1 unique (Fig. 53) sequence DNAs to EcoRI cleaved, nitrocellulose filter-bound DNAs from H. halobium strain NRC-1 (lanes A), H. salinarum (lanes B), H. volcanii (lanes C), H. trapanicum (lanes D), H. yallismortia (lanes E) and H. saccharovorum (lanes F).

Fig. 43. Probe: clone 7.

7
A B C D E F

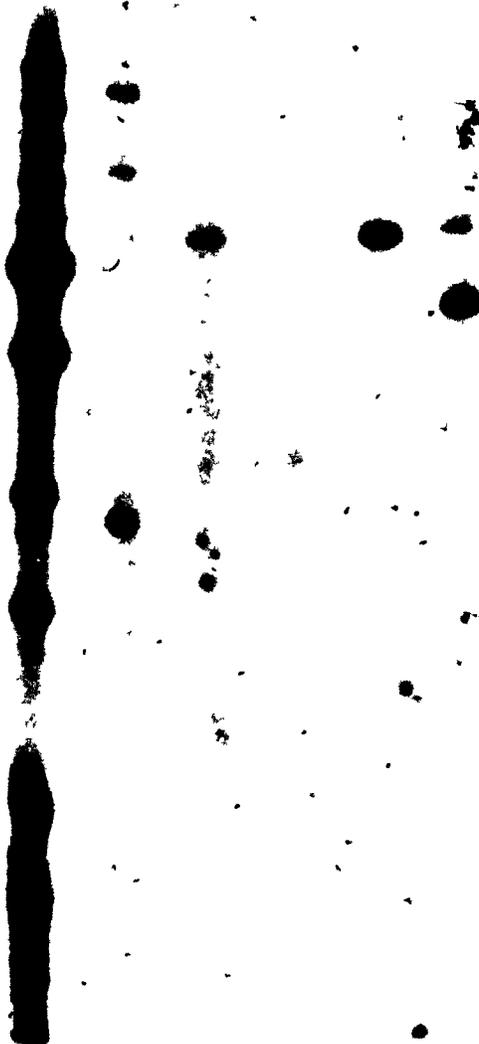


Fig. 44. Probe: clone 34.



Fig. 45. Probe: clone 37.



Fig. 46. Probe: clone 43.



Fig. 47. Probe: clone 54.



Fig. 48. Probe: clone 74.

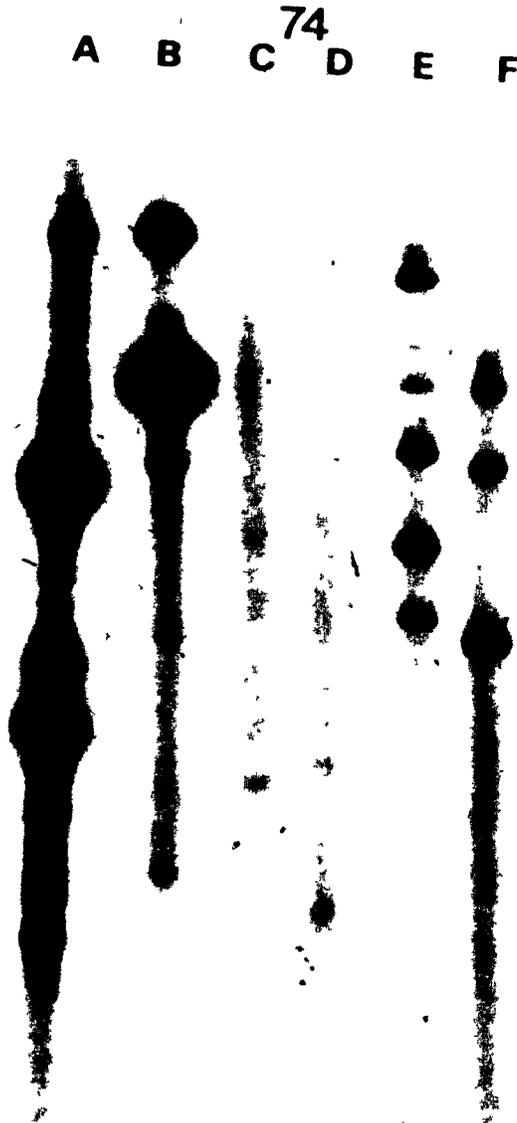


Fig. 49. Probe: clone 85.



Fig. 50. Probe: clone v88.

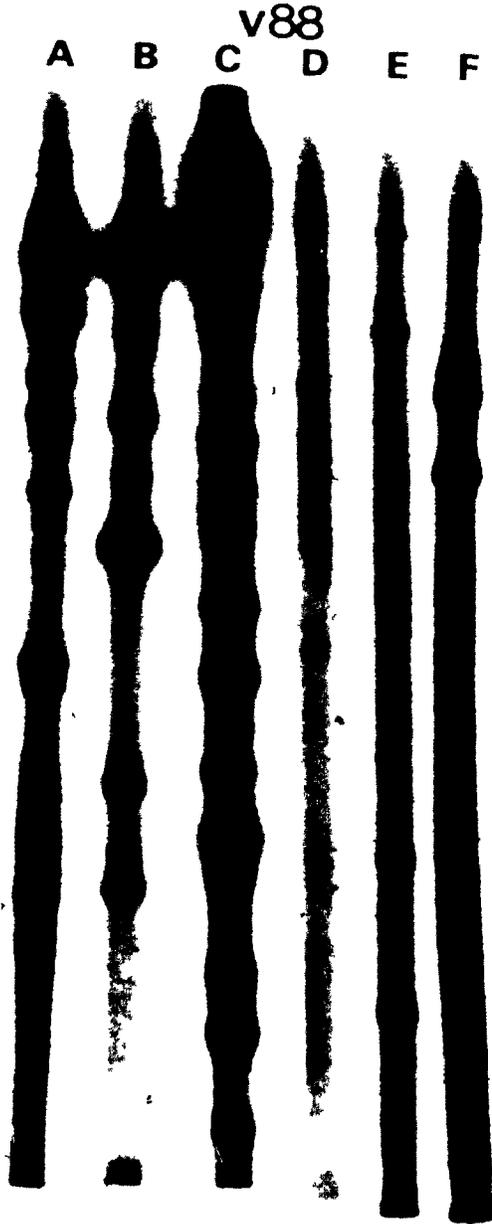


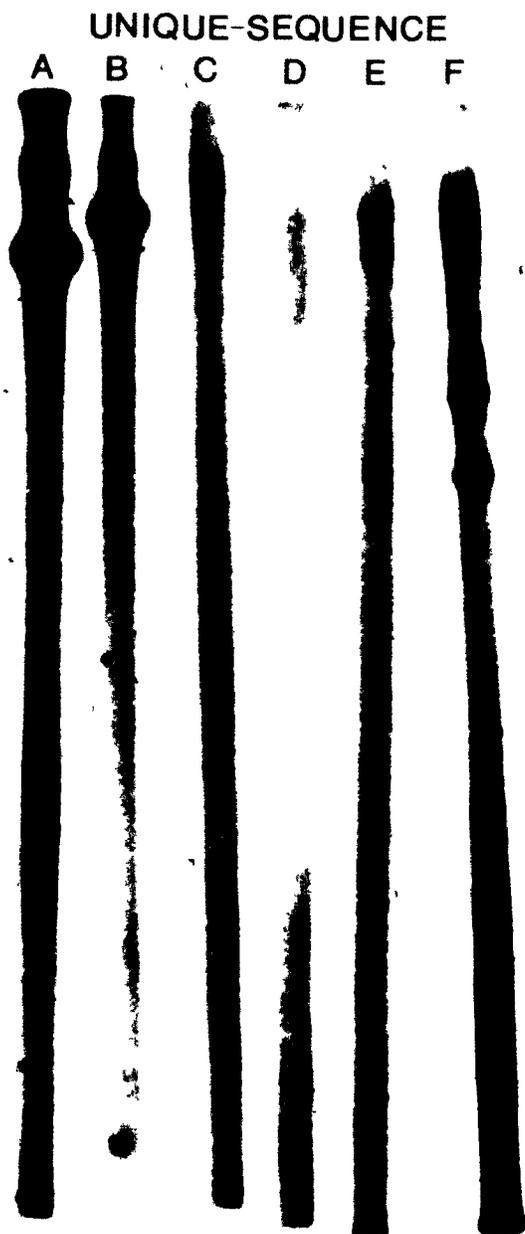
Fig. 51. Probe: clone v97.



Fig. 52. Probe: clone v122.



Fig. 53. Probe: clones p27, p41 and p97.



known in halobacteria. The species studied are confined by their obligate halophily to rigidly defined and geographically isolated habitats¹, and the lack of any strong correlations between the phylogenetic distributions of different repeat sequence families is not easily explained by such transfer. Alternatively, although subject to expansions, contractions and occasional loss, repeat sequence families have profited from stochastic or selective intragenomic processes (Doolittle and Sapienza, 1980) which have, within limits, maintained them throughout the long evolutionary diversification of this group of archaeobacteria.

B. Other archaeobacterial genomes

The existence of so many evolutionarily conserved repeated sequences in halobacterial genomes prompts the question of whether or not large numbers of repeated sequences are a hallmark of archaeobacterial genomes and thus makes them distinct from eubacterial genomes and more like those of eukaryotes.

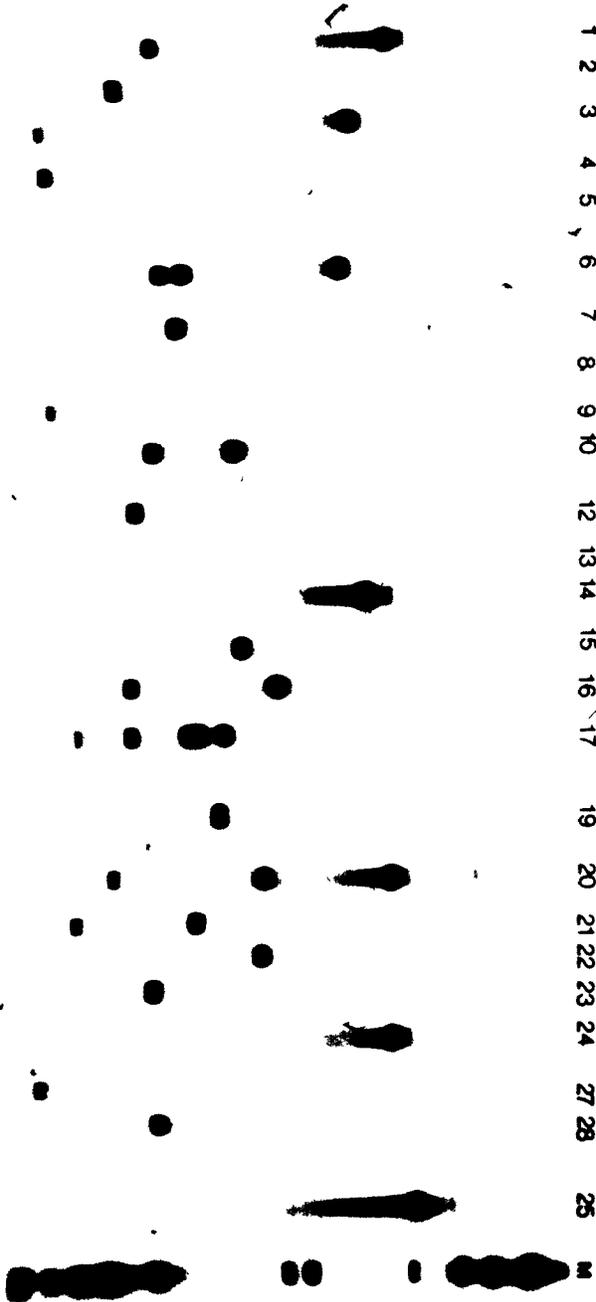
In order to investigate this possibility, an experiment similar to that shown in Figure 15 was conducted with the DNA of the Archaeobacterium Thermoplasma acidophilum. T. acidophilum DNA was obtained from D. Searcy (University of Massachusetts, Amherst), cleaved with EcoRI and ligated into pBR322. After screening of transformants to ensure cloned fragments were present, DNAs were prepared from 27 clones which contained inserts (several clones which contained more than one EcoRI

¹H. halobium strain NRC-1 was isolated from salted cod, H. salinarum from salted hides, H. trapanicum from "Trapani" salt at Bergen, Norway (Bergey's Manual of Determinative Bacteriology, seventh edition, The Williams and Wilkins Company, Baltimore, 1957). H. vallismortis was isolated from Death Valley, California. H. volcanii was isolated from the Dead Sea. H. saccharovororum was isolated from southern San Francisco Bay (Tomlinson and Hochstein, 1976).

Fig. 54. Hybridization of ^{32}P -labelled cloned EcoRI fragments of Thermoplasma acidophilum DNA to EcoRI cleaved, nitrocellulose filter-bound T. acidophilum DNA. Nitrocellulose strip numbers 1, 3, 6, 10, 16, 17, 20 were probed with recombinant plasmids bearing more than one T. acidophilum EcoRI fragment. No fragments other than those identical in mobility to the fragments cloned are hybridized.

1 2 3 4 5 6 7 8 9 10 12 13 14 15 16 17 19 20 21 22 23 24 27 28 25 M

T. acidophilum EcoRI clones



fragment were chosen in order to increase the fraction of the genome represented), nick-translated and hybridized to Southern blots of EcoRI cleaved T. acidophilum DNA. Figure 54 shows the result of this experiment. 53.6 kpb or 4.6% of the T. acidophilum genome is represented in these 27 clones. No fragments in addition to those identical in mobility to the cloned fragments were hybridized. This implies that not all archaeobacterial genomes are rich in repetitive sequences. A large number of repetitive sequences has, however, been found in the genome of Methanobacterium thermoautotrophicum by Reeve (personal communication). Thus repetitive sequences appear to be common in two of the three major groups of Archaeobacteria, the methanogens and the extreme halophiles.

M. Organisms: How low do they go?

The presence of repeated sequences in both eukaryotic and prokaryotic genomes has given rise to endless speculations as to the roles they play in the control of gene expression (Britten and Davidson, 1969; Davidson and Britten, 1979), differentiation and development (Constantini et al., 1980; Moore et al., 1980), and generating evolutionary novelty (Ohno, 1970; Cohen, 1976; Starlinger and Saedler, 1976; Kleckner, 1977; Nevers and Saedler, 1977). The realization that many of these repeated sequences are mobile, capable of differential replication and thus able to increase in number within genomes, led Ford Doolittle and me to propose a rather different explanation for their existence (Doolittle and Sapienza, 1980; Sapienza and Doolittle, 1981). The explanation follows logically from a consideration of the essentials of natural selection.

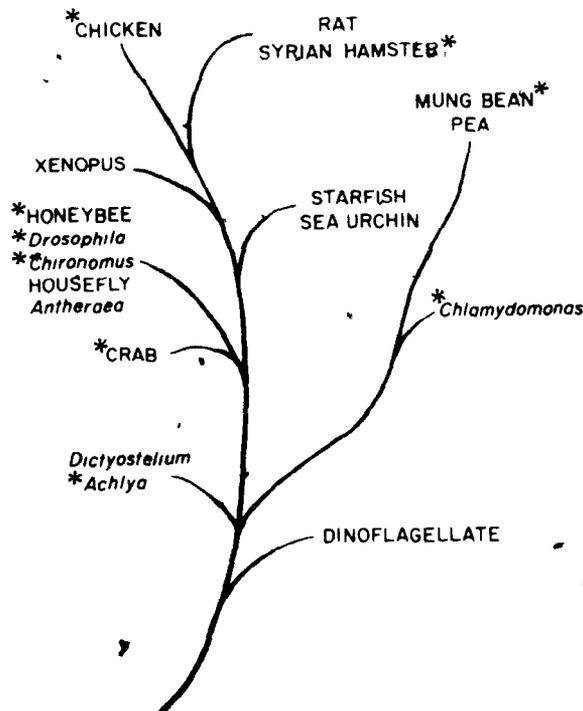
The only selection pressure which DNAs experience directly is the pressure to survive within cells. Cells themselves are environments in

which DNA sequences can replicate, mutate and evolve (Orgel, 1979). If there are ways in which mutation can increase the probability of a DNA sequence's survival within cells without drastically reducing the fitness of the organism harbouring such a sequence, natural selection will inevitably give rise to DNAs whose only "function" is survival within genomes. A corollary of this statement is that if a DNA sequence can be shown to have adopted such a survival strategy, no other (phenotypic) explanation for its continued existence is required. Transposition is one such strategy.

Prokaryotic insertion sequences and transposons can in general be inserted into a large number of chromosomal or plasmid sites, can be excised precisely or imprecisely and can engender deletions in adjacent DNAs (Cohen, 1976; Nevers and Sadler, 1977). Models for transposition require the synthesis of a new element for insertion, without loss of old elements (Shapiro, 1979; Arthur and Sherrat, 1979, Harshey and Bukhari, 1981). This behaviour ensures the survival of the transposed element, regardless of effect on phenotype (and this is generally destructive) unless such an effect is sufficiently negative.

Middle repetitive DNAs in Drosophila (Potter et al., 1979; Strobel et al., 1979; Young, 1979), yeast (Cameron et al., 1979) and possibly humans (Jagadeeswaran et al., 1981; Krowlewski et al., 1982) also appear to be transposable although the mechanism of transposition may well be different (Jagadeeswaran et al., 1981) than that proposed for prokaryotic elements. A large fraction of many eukaryotic genomes consists of middle-repetitive DNA (Lewin, 1975), and the variety and patterns of their interspersions with unique-sequence DNA makes no particular phylogenetic (Figure 55) or phenotypically functional sense. Britten,

Fig. 55. Phylogenetic tree of some organisms in which genome organization has been studied. Asterisk denotes that the middle repetitive DNA and unique-sequence DNA is arranged as in the Drosophila, or long-period interspersion, pattern; others have genome organization characteristic of the Xenopus, or short-period interspersion, pattern. Data from Davidson et al. (1973), Graham et al. (1974), Firtel and Kindle (1975), Manning et al. (1975), Crain et al. (1976), Efstratiadis et al. (1976), Howell and Walker (1976), Wells et al. (1976), Valau et al. (1977), Hudspeth et al. (1977), Moyzis et al. (1977), Arthur and Strauss (1978), Smith and Boal (1978), Christie and Skinner (1979), Hinnebusch et al. (1980), Smith et al. (1980), and Murray et al. (1979, 1981).



Davidson and collaborators have elaborated models which ascribe regulatory functions to middle-repetitive DNAs, and evolutionary adaptability to the quantitative and qualitative changes in middle-repetitive DNA content observed even between closely related species (Britten and Davidson, 1969; 1971; Davidson et al., 1977; Davidson and Britten, 1979). The evidence for a phenotypically functional role for middle repetitive DNAs remains dishearteningly weak (Klein et al., 1978; Scheller et al., 1978; Kuroiwa and Natori, 1979; Kimmel and Firtel, 1979), and middle-repetitive DNAs together comprise too large a fraction of most eukaryotic genomes to be kept homologous in sequence by Darwinian selection operating on phenotype (Kimura, 1968; Salser and Isaacson, 1976). If it is assumed that middle-repetitive DNAs are transposable elements or degenerate descendants of such elements, then the observed spectra of sequence divergence within families and changes in middle-repetitive DNA family sequence and abundance may all be explained by non-phenotypic selection operating within genomes.

The "Selfish DNA" explanation for the existence of transposable elements is not simply a variant of Occam's Razor (Smith, 1980) in the sense that because transposable elements may be explained by non-phenotypic selection, they must be explained in this way. Rather, other explanations for the existence of transposable elements suffer from lack of evidence or logical flaws. Most speculations on the function of prokaryotic transposable elements concentrate on the roles they may play in promoting the evolution of plasmid and bacterial chromosomes through rearrangements and the modular assembly of functional units (Cohen, 1976; Starlinger and Saedler, 1976; Nevers and Saedler, 1977) or in

facilitating gene transfer (Campbell, 1981). The functions of middle-repetitive DNAs have also been relegated to the generation of new chromosome primary structures and as cogs in ". . . the biochemical machinery by which organisms control their hereditary apparatus and regulate its transmission (changed or unchanged) over cell generations" (Shapiro and Cordell, 1982). Such explanations have at their base the implicit assumption that evolution is anticipatory. The notion that natural selection operating on phenotype can direct the formation of structures whose effects are felt only in the future runs counter to the usual laws of cause and effect. It seems unreasonable to endow the process of natural selection with foresight.

This does not mean transposable elements do not play such evolutionary roles. Genomic rearrangements effected by transposable elements will, of course, give rise to new chromosome structures and gene arrangements upon which natural selection may act. However, this is not their primary function. Part of the argument lies in the definition of "function". An evolutionary definition for "function" is best illustrated by an anecdote from one of John Maynard Smith's lectures (June 30, 1981 at Cambridge University):

Horses have very stiff spines. Because of their very stiff spines, people may sit upon their backs and ride them. However, natural selection certainly did not fashion the horse's stiff spine so that people could sit upon their backs.

Thus, a workable definition of function might be: that purpose served by a structure which natural selection fashioned that structure to serve. Within this framework, the function of transposable elements

must certainly be survival with genomes, although they, like the horse's spine, may have other effects.

Transposable elements, and the "Selfish DNA" proposal are also distinct from "junk DNA" (although some authors seem not to have noticed - Ohno, 1981; Grant, 1981). Junk DNA is merely a collection of ultimately to be disposed of, non-functional nucleotide sequences, carted along by genomes as excess baggage or evolutionary relicts. Such sequences, unlike transposable elements, are unable to influence their copy number or maintain themselves in sequence except by random processes.

Perhaps the most comfortable framework within which to view transposable elements (and many plasmids; Novick, 1980) is to think of them as the lower limit in a continuum of living organisms. They may be considered akin to defective viruses; organisms in their own right, but unable to live outside the realm of the genome which harbours them.

References

- Arnold, S., R. Chandrasekaran, D.L. Birdsall, A.G.N., Leslie and R.L. Ratliff. 1980. Left-handed DNA helices. *Nature* 283:743-748.
- Arthur, A. and D.J. Sherratt. 1979. Dissection of the transposition process: A transposon-encoded site-specific recombination system. *Molec. gen. Genet.* 175:267-274.
- Arthur, R.R. and N.A. Straus. 1978. DNA-sequence organization in the genome of the domestic chicken (Gallus domesticus). *Can. J. Biochem.* 56:257-263.
- Birnboim, H.C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7:1513-1523.
- Britten, R.J. and E.H. Davidson. 1969. Gene regulation for higher cells: A theory. *Science* 165:349-357.
- Calos, M.P. and J.H. Miller. 1980. Transposable elements. *Cell* 20:579-595.
- Campbell, A. 1981. Evolutionary significance of accessory DNA elements in bacteria. *Ann. Rev. Microbiol.* 35:55-83.
- Cameron, J.R., E.Y. Loh and R.W. Davis. 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* 16:739-751.
- Christie, N.T. and D.M. Skinner. 1979. Interspersion of highly repetitive DNA with single copy DNA in the genome of the red crab, Geryon quinquesdens. *Nucleic Acids Res.* 6:781-796.
- Clewell, D.B. and D.R. Helinski. 1969. Supercoiled circular DNA-protein complex in Escherichia coli: purification and induced conversion to an open circular DNA form. *Proc. Natl. Acad. Sci. U.S.A.* 62:1159-1166.
- Cohen, S.N. 1976. Transposable genetic elements and plasmid evolution. *Nature* 263:731-738.
- Constantini, F.D., R.J. Britten and E.H. Davidson. 1980. Message sequences and short repetitive sequences are interspersed in sea urchin egg poly(A)⁺ RNA's. *Nature* 287:111-117.
- Cox, E.C. 1976. Bacterial mutator genes and the control of spontaneous mutation. *Ann. Rev. Genet.* 10:135-156.
- Crain, W.R., E.H. Davidson and R.J. Britten. 1976. Contrasting patterns of DNA sequence arrangement in Apis mellifera (honeybee) and Musca domestica (housefly). *Chromosoma* 59:1-19.
- Davidson, E.H. and R.J. Britten. 1979. Regulation of gene expression: Possible role of repetitive sequences. *Science* 204:1052-1059.

- Davidson, E.H., B.R. Hough, C.S. Amenson and R.J. Britten. 1973. General interspersion of repetitive with non-repetitive sequence elements in the DNA of Xenopus. J. Mol. Biol. 77:1-23.
- Doolittle, W.F. and C. Sapienza. 1980. Selfish genes, the phenotype paradigm, and genome evolution. Nature 284:601-603.
- Dunn, R., J. McCoy, M. Simsek, A. Majumdar, S.H. Chang, U.L. RajBhandary and H.G. Khorana. 1981. The bacteriorhodopsin gene. Proc. Natl. Acad. Sci. U.S.A. 78:6744-6748.
- Efstratiadis, A., W.R. Crain, R.J. Britten, E.H. Davidson and F.C. Kafatos. 1976. DNA sequence organization in the lepidopteran Antheraea pernyi. Proc. Natl. Acad. Sci. U.S.A. 73:2289-2293.
- Ewens, W.J. 1979. Mathematical Population Genetics. Springer, Berlin.
- Feller, W. 1968. An Introduction to Probability Theory and Its Applications. Wiley, New York.
- Fox, G.E., E. Stackebrandt, R.B. Hespell, J. Gibson, J. Maniloff, T.A. Dyer, R.S. Wolfe, W.E. Balch, R.S. Tanner, L.J. Magrum, L.B. Zablen, R. Balkmore, R. Gupta, L. Bonen, B.J. Lewis, D.A. Stahl, K.R. Luehrsen, K.N. Chen and C.R. Woese. 1980. The phylogeny of prokaryotes. Science 209:457-463.
- Galau, G.A., M.E. Chamberlin, B.R. Hough, R.J. Britten and E.H. Davidson. 1975. Evolution of repetitive and non-repetitive DNA in Molecular Evolution, F.J. Ayala (ed.), Sinauer Associates, Sunderland, Massachusetts.
- Gochnauer, M.B. and D.J. Kushner. 1969. Growth and nutrition of extremely halophilic bacteria. Can. J. Microbiol. 15:1157-1163.
- Graf, H. 1979. Optimization of conditions for the in vitro formation of hybrid DNA molecules by DNA ligase. Biochim. Biophys. Acta 564:225-234.
- Graham, D.E., B.R. Neufeld, E.H. Davidson and R.J. Britten. 1974. Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. Cell 1:127-137.
- Grant, B. 1981. The safe-neighborhood hypothesis of junk DNA. J. Theor. Biol. 90:149-150.
- Harshey, R.M. and A.I. Bukhari. 1981. A mechanism of DNA transposition. Proc. Natl. Acad. Sci. U.S.A. 78:1090-1094.
- Hinnebursh, A.G., L.C. Klotz, E. Immergut and A.R. Loeblich III. 1980. Deoxyribonucleic acid sequence organization in the genome of the dinoflagellate Cryptocodinium cohnii. Biochem. 19:1744-1755.

- Hofman, J.D., R.H. Lau and W.F. Doolittle. 1979. The number, physical organization and transcription of ribosomal RNA cistrons in an archaeobacterium: Halobacterium halobium. Nucleic Acids Res. 7:1321-1333.
- Holmes, D.S. and M. Quigley. 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. 114:193-197.
- Howell, S.H. and L.L. Walker. 1976. Informational complexity of the nuclear and chloroplast genomes of Chlamydomonas reinhardtii. Biochim. Biophys. Acta 418:249-256.
- Hudspeth, M.E.S., W.E. Timerlake and R.B. Goldberg. 1977. DNA sequence organization in the water mold Achlya. Proc. Natl. Acad. Sci. U.S.A. 74:4332-4336.
- Jagadeeswaran, B., B.G. Forget and S.M. Weissman. 1981. Short interspersed repetitive DNA elements in eukaryotes: Transposable DNA elements generated by reverse transcription of RNA polIII transcripts? Cell 26:141-142.
- Jeffreys, A.J. and R.A. Flavell. 1977. A physical map of the DNA regions flanking the rabbit β -globin gene. Cell 12:429-439.
- Johnsrud, L., M.P. Calos and J.H. Miller. 1978. The transposon Tn^9 generates a 9 b.p. repeated sequence during integration. Cell 15:1209-1219.
- Kimmel, A.R. and R.A. Firtel. 1979. A family of short, interspersed repeat sequences at the 5' end of a set of Dictyostelium single-copy mRNAs. Cell 16:787-796.
- Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217:624-626.
- Kleckner, N. 1977. Translocatable elements in prokaryotes. Cell 11:11-23.
- Kleckner, N. 1981. Transposable elements in prokaryotes. Ann. Rev. Genet. 15:341-404.
- Klein, W.H., T.L. Thomas, C. Lai, R.H. Scheller, R.J. Britten and E.H. Davidson. 1978. Characteristics of individual repetitive sequence families in the sea urchin genome studied with cloned repeats. Cell 14:889-900.
- Kopecko, D. 1980. Involvement of specialized recombination in the evolution and expression of bacterial genes in Plasmids and Transposons: Environmental Effects and Maintenance Mechanisms, C. Stuttard and K. Rozee (eds.), Academic Press, New York.
- Krowlewski, J.J., A.H. Bertelsen, M.Z. Humayun and M.G. Rush. 1982. Members of the Alu family of interspersed, repetitive DNA sequences are in the small circular DNA population of monkey cells grown in culture. J. Mol. Biol. 154:399-415.

- Krumlauf, R. and Marzluf, G.A. 1980. Genome organization and characterization of the repetitive and inverted repeat DNA sequences in Neurospora crassa. J. Biol. Chem. 255:1138-1145.
- Kuroiwa, A. and S. Natori. 1979. Preferential expression of unique sequences adjacent to middle repetitive sequences in mouse cytoplasmic RNA. Nucleic Acids Res. 7:751-754.
- Kushner, D.J. 1978. Life in high salt and solute concentrations: halophilic bacteria in Microbial Life in Extreme Environment, D.J. Kushner (ed.), Academic Press, New York.
- Lau, R.H. and W.F. Doolittle. 1978. Covalently closed circular DNAs in closely related unicellular cyanobacteria. J. Bacteriol. 137:648-652.
- Lau, R.H., C. Sapienza and W.F. Doolittle. 1980. Cyanobacterial plasmids: Their widespread occurrence, and the existence of regions of homology between plasmids in the same and different species. Molec. gen. Genet. 178:203-211.
- Lewin, B. 1975. Units of transcription and translation: Sequence components of heterogeneous nuclear RNA and messenger RNA. Cell 4:77-93.
- Manning, J.E., C.W. Schmid and N. Davidson. 1975. Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome. Cell 4:141-155.
- Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Meth. Enzymol. 65:499-560.
- Messing, J., R. Crea and P.H. Seeburg. 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. 9:309-321.
- Miller, J.H. 1972 Experiments in Molecular Genetics. Cold Spring Harbor Laboratory.
- Mitchell, R.M., L.A. Loeblich, L.C. Klotz and A.R. Loeblich III. 1979. DNA organization in Methanobacterium thermoautotrophicum. Science 204:1082-1084.
- Moore, G.P., F.D. Costantini, J.W. Posakony, E.H. Davidson and R.J. Britten. 1980. Evolutionary conservation of repetitive sequence expression in sea urchin egg RNAs. Science 208:1046-1048.
- Moore, R.T. and B.J. McCarthy. 1969a. Characterization of the DNA of various strains of halophilic bacteria. J. Bacteriol. 99:248-254.
- Moore, R.T. and B.J. McCarthy. 1969b. Base sequence homology and renaturation studies of the DNA of extremely halophilic bacteria. J. Bacteriol. 99:255-262.
- Moyzis, R., J. Bonnet and P.O.P. Ts'0. 1977. DNA sequence organization of the Syrian hamster. J. Cell. Biol. 75:103a.

- Murray, M.G., J.D. Palmer, R.E. Cuellar and W.F. Thompson. 1979. Deoxyribonucleic acid sequence organization in the mung bean genome. *Biochem.* 18:5259-5266.
- Murray, M.G., D.L. Peters and W.F. Thompson. 1981. Ancient repeated sequences in the pea and mung bean genomes and implications for genome evolution. *J. Mol. Evol.* 17:31-42.
- Nevers, P. and H. Saedler. 1977. Transposable genetic elements as agents of gene instability and chromosomal rearrangements. *Nature* 268:109-115.
- Nordheim, A., M.L. Pardue, E.M. Lafer, A. Moller, B.D. Stollar and A. Rich. 1981. Antibodies to left-handed Z-DNA bind to interband regions of *Drosophila* polytene chromosomes. *Nature* 194:417-422.
- Norgard, M.V., K. Keem and J.J. Monahan. 1978. Factors affecting the transformation of *Escherichia coli* strain 1776 by pBR322 plasmid DNA. *Gene* 3:279-292.
- Norgard, M.V., K. Emigholz and J.J. Monahan. 1979. Increased amplification of pBR322 plasmid deoxyribonucleic acid in *Escherichia coli* K-12 strains RR1 and 1776 grown in the presence of high concentrations of nucleoside. *J. Bacteriol.* 138:270-272.
- Novick, R.P. 1980. Plasmids. *Scient. Am.* 243(6):102-127.
- Ohno, S. 1970. Evolution by gene duplication. Springer, New York.
- Ohno, S. 1981. (AGCTG)(AGCTG)(AGCTG)(GGGTG) as the primordial sequence of intergenic spacers: The role in immunoglobulin class switch. *Differentiation* 18:65-74.
- Orgel, L.E. 1979. Selection in vitro. *Proc. Royal Soc. B.* 205:435-442.
- Pfeifer, F., G. Weidinger and W. Goebel. 1981a. Characterization of plasmids in Halobacteria. *J. Bacteriol.* 145:369-374.
- Pfeifer, F., G. Weidinger and W. Goebel. 1981b. Genetic variability in *Halobacterium halobium*. *J. Bacteriol.* 145:375-381.
- Potter, S.S., W.J. Brorein, Jr., P. Dunsmuir and G.M. Rubin. 1979. Transposition of elements of the 412, copia and 297 dispersed repeated gene families in *Drosophila*. *Cell* 17:415-427.
- Read, H.A. and S.R. Jaskunas. 1980. Isolation of *E. coli* mutants containing multiple transpositions of IS sequences. *Molec. gen. Genet.* 180:157-164.
- Rigby, P.W.J., M. Dieckmann, C. Rhodes and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237-251.

- Rogers, P.J. and C.A. Morris. 1978. Regulation of bacteriorhodopsin synthesis by growth rate in continuous cultures of Halobacterium halobium. Arch. Microbiol. 119:323-325.
- Salser, W. and J.S. Isaacs. 1976. Mutation rates in globin genes: The genetic load and Haldane's Dilemma. Prog. Nucleic Acids Res. Molec. Biol. 19:205-220.
- Sapienza, C. and W.F. Doolittle. 1981. Genes are things you have whether you want them or not. Cold Spring Harb. Symp. Quant. Biol. 45:177-182.
- Scheller, R.H., F.D. Constantini, M.R. Dozowski, R.J. Britten and E.H. Davidson. 1978. Specific representation of cloned repetitive DNA sequences in sea urchin RNAs. Cell 15:189-203.
- Schnabel, H., W. Zillig, M. Pfaffle, R. Schnabel, H. Michel and H. Delius. 1982. Halobacterium halobium phage Φ H. EMBO Journal, in press.
- Searcy, D.G. and E.K. Doyle. 1975. Characterization of Thermoplasma acidophilum deoxyribonucleic acid. Int. J. Syst. Bacteriol. 25:286-289.
- Shapiro, J.A. 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. U.S.A. 76:1933-1937.
- Shapiro, J.A. and B. Cordell. 1982. Eukaryotic mobile and repeated genetic elements. Biology of the Cell, in press.
- Slack, J.M. and I.S. Snyder. 1978. Bacteria, and human disease. Year Book Medical Publishers, Inc., Chicago.
- Smith, M.J. and R. Boal. 1978. DNA sequence organization in the common Pacific starfish Pisaster ochraceus. Can. J. Biochem. 56:1048-1054.
- Smith, M.J., A. Lui, K.K. Gibson and J.K. Etkorn. 1980. DNA sequence organization in the starfish Dermasterias imbricata. Can. J. Biochem. 58:352-360.
- Smith, T.F. 1980. Occam's Razor. Nature 285:620.
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503-517.
- Spudich, J.L. and W. Stoeckenius. 1980. Light-regulated retinal-dependent reversible phosphorylation of Halobacterium proteins. J. Biol. Chem. 255:5501-5503.
- Starlinger, P. and H. Saedler. 1976. IS-elements in microorganisms. Curr. Topics Microbiol. Immunol. 75:111-152.

- Strobel, E., P. Dunsmuir and G.M. Rubin. 1979. Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 17:429-439.
- Sutcliffe, J.G. 1978. pBR322 restriction map derived from the DNA sequence: accurate DNA size markers up to 4361 nucleotide pairs long. Nucleic Acids Res. 5:2722-2728.
- Timberlake, W.E. 1978. Low repetitive DNA content in Aspergillus nidulans. Science 202:973-974.
- Tomlinson, G.A. and L.I. Hochstein. 1976. Halobacterium saccharovorum sp. nov., a carbohydrate-metabolizing, extremely halophilic bacterium. Can. J. Microbiol. 22:587-591.
- Walseth, T.F. and R.A. Johnson. 1979. The enzymatic preparation of [α -³²P] nucleoside triphosphates, cyclic [³²P]AMP, and cyclic [³²P]GMP. Biochim. Biophys. Acta 562:11-31.
- Wang, A.H., G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.F. vanBoom, G. van der Marel and A. Rich. 1979. Molecular structure of a left-handed helical DNA fragment at atomic resolution. Nature 282:680-686.
- Weber, H.J., R. Bogomolni and T.L. Leighton. 1982. Genetic instability of Halobacterium halobium. Zbl. Bakt. Hyg., I. Abt. Orig., in press.
- Weidinger, G., G. Klotz and W. Goebel. 1979. A large plasmid from Halobacterium halobium carrying genetic information for gas vacuole formation. Plasmid 2:377-386.
- Wells, R., H.D. Royer and C.P. Hollenberg. 1976. Non-Xenopus-like DNA sequence organization in the Chironomus tentans genome. Molec. gen. Genet. 147:45-51.
- Woese, C.R. 1981. Archaeobacteria. Scient. Am. 244(6):98-122.
- Woese, C.R. and G.E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. U.S.A. 74:5088-5090.
- Young, M.W. 1979. Middle repetitive DNA: A fluid component of the Drosophila gene. Proc. Natl. Acad. Sci. U.S.A. 76:6274-6278.
- Zillig, W., J.T. Tu and I. Holz. 1981. Thermoproteales - a third order of thermoacidophilic archaeobacteria. Nature 293:85-86.