INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

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

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Université d'Ottawa · University of Ottawa

Mapping the Genome of *Sulfolobus solfataricus* P2

Felix M. Young

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the

Ottawa-Carleton Institute of Biology

Thèse soumise à
l'École des études supérieures et de la recherche
Université d'Ottawa
en vue de l'obtention de la maîtrise en sciences à

L'Institut de biologie d'Ottawa-Carleton

Acknowledgements

I would like to thank Dr. Charlebois for providing me the opportunity to further my studies in biology. The knowledge and experience gained under his supervision were invaluable. I would also like to thank Dr. Johnson and Dr. Vierula for their support and input. I have received a lot of technical support and advice from Dr. Hickey and his lab, Dr. Drouin and his lab and the members of my lab and would like to thank them.

This thesis is dedicated to my mother, Uncle Mike, Aunt Ella, Paul and Chris who have supported me through the years. This is also dedicated to my friends.

Abstract

The goal of this project was to construct a contig map of the 3.0 +/- 0.1 Mbp genome of the thermoacidophilic archaeon *Sulfolobus solfataricus* P2. This high resolution physical map would guide the DNA sequencing of the entire genome on a cosmid-by-cosmid strategy. 60 to 70% of the genome has been cloned into cosmids which have been linked into 31 contigs by the landmark strategy and chromosome walking. The lack of coverage of the entire genome by cosmid clones may be the result of the incompatibility of the A-T rich DNA of *S. solfataricus* P2 with the *Escherichia coli* host system. The identification of insertion sequence elements in this archaeal genome suggests that it may be unstable.

Attempts have been made to generate an integrated contig/macrorestriction map because of the lack of genome coverage by the cosmid clones, but this effort has been unsuccessful because of unreliable hybridization results and a lack of suitable restriction enzymes. This cosmid library will be supplemented by λ phage clones with PCR for the final few pieces for complete coverage of the genome of *S. solfataricus* P2 so that DNA sequencing can be completed.

Table of Contents

List of Figures

List of Tables

Introduction

A) The Thermoacidophilic Archaeon, *Sulfolobus solfataricus*

In 1959, Whittaker proposed the five kingdoms of Animalia, Plantae, Fungi, Protista, and Monera for classifying living organisms (Whittaker, 1959). It has become apparent that this five-kingdom scheme is not phylogenetically correct. This is due to the discovery of the Archaea in 1977 by comparative analysis of the 16S ribosomal RNA (16S rRNA) molecule (Fox *et. al.*, 1977; Woese and Fox, 1977). The 16S rRNA molecule is readily isolated, ubiquitous and highly constrained in sequence therefore it is more revealing of evolutionary relationships than are classical phenotypes, particularly for microorganisms (Woese *et. al.*, 1975). In order to remedy this situation of classifying organisms to reflect their evolutionary relationships, Woese *et al.* (1990) have proposed a new taxon called a "domain" to be established above the level of kingdom.

Organisms are classified into one of three domains: Archaea, Bacteria and Eucarya (Woese *et al.*, 1990). The domain Archaea is composed of unicellular organisms that resemble Bacteria at a gross level, but possess their own unique characteristics such as lipids with ether links and polyisoprenoid (branched) chains rather than ester links and straight carbon chains, and intermediary biochemical pathways that allow them to live at extreme conditions of high temperatures, low pH or high salinity (Danson *et al.*, 1992; Kates *et al.*, 1993). This domain is divided into two kingdoms, Euryarchaeota and Crenarchaeota: the Euryarchaeota being a phenotypically diverse group

comprised of extreme halophiles, sulfate-reducers, thermophiles and methanogens, while the Crenarchaeota is a phenotypically homogeneous group comprised solely of thermophiles. This classification system proposed by Woese *et al.*(1990) was based on the analysis of 16S rRNA sequences in the existing database at that time.

Recent isolation of archaea from marine environments (Delong, 1992), the North Sea and Alaskan oil reservoirs (Stetter *et al.*, 1993), Antarctic marine picoplankton (Delong *et al.*, 1994) and archaeal DNA extracted directly from sediment of a hot spring in Yellowstone National Park (Barns *et al.*, 1994) have altered our understanding of the phylogenetic organization of the Archaea. The discovery of archaeal lineages that branch near the crenarchaeal stem and the deeply diverging Euryarchaeota, *Methanopyrus* (Burggraf *et al.*, 1991) blurs the prior sharp distinction between the two archaeal groups. The two distinct archaeal lineages proposed by Woese *et al.*, may become a bush of lineages arising from the root of the Archaeal domain as more Archaea are discovered.

Despite this ambiguous distinction between the two archaeal lineages, it is clearly evident that thermophily occurs in both lineages which would suggest that the ancestral phenotype of the Archaea was thermophilic. It has also been suggested that the Archaeal/Eucaryal ancestor was a thermophile (Woese, 1987). The analysis of duplicated genes show that the Archaea and Eucarya are sister groups (Gogarten *et al.*, 1989; Iwabe *et al.*, 1989; Brown and Doolittle, 1995; Lawson *et al.*, 1996). The Archaea and Eucarya share common features

such as the resistance or sensitivity to various antibiotics, the presence of tRNA introns, the lack of formylmethionyl-tRNA, and the lack of peptidoglycan (Zillig, 1981). In addition, they have similar transcription mechanisms, process the rRNA primary transcript with U3 and degrade proteins by proteasomes (Keeling *et al.*, 1994; Keeling and Doolittle, 1995). This would seem to support the idea that the eukaryotes and their nuclear genome arose from the genome of a thermophilic, ancestral Archaeon.

The genomes of Archaea are similar in size, shape and structure to Bacteria. Both are small, circular and compact with high gene densities and the genes are organized into operons with a gene order conserved between these two domains (Ramirez *et al.*, 1993) suggesting the common ancestor of all life had a similar form, and it is the eukaryotic nuclear genome which is derived.

Archaeal genomes appear to be complex structurally and genetically, in particular the extreme halophiles (Charlebois *et al.*, 1991; Pfeifer *et al.*, 1981; Sapienza and Doolittle, 1982; Pfeifer, 1986; Charlebois and Doolittle, 1989). The halobacterial genomes are partitioned and not homogeneous in composition. Plasmids are prevalent and contribute as much as 30% of the size of the genome. The G-C content and oligonucleotide composition differ for plasmids, portions of plasmids and islands within the chromosome form the bulk of the DNA with 10-30% of the genome made up of this minor-fraction (FII) DNA. High rates of disruptive transposition and recombination caused by the presence of insertion sequence (IS) elements have been observed in some halobacterial

genomes (Pfeifer *et al.*, 1981; Sapienza *et al.*, 1982; Pfeifer *et al.*, 1988; Pfeifer *et al.*, 1989). Interestingly, these IS elements are found clustered within the FII DNA regions (Pfeifer and Betlach, 1985; Hofman *et al.*, 1986; Cohen *et al.*, 1992) and are responsible for the dynamic nature of the halobacterial genomes.

Perhaps, the environment of a living Archaeon may be quite similar to that of the ancestral Archaeon. One such Archaeon, *Sulfolobus solfataricus*, a Crenarchaeote, is found in hot sulfur springs (solfataras) in Yellowstone National Park, Italy, Dominica and El Salvador (Brock *et al.*, 1972). Its thermoacidophilic characteristic and its important position in the universal phylogenetic tree make *S. solfataricus* an ideal model organism for genomic studies. So began "The *Sulfolobus solfataricus* P2 Genome Project" in 1993 funded by the Canadian Genome Analysis and Technology Program. It is a collaboration of the labs located at the University of Ottawa, Dalhousie University and the National Research Council in Halifax with the objective to map and sequence the entire genome of this Archaeon (Charlebois *et al.*, 1996; Sensen *et al.*, 1996; Sensen *et al.*, in press).

This effort to obtain the complete DNA sequence of *S. solfataricus* P2 will assist in our understanding of this species' biology. In particular, revealing the entire gene complement will permit a thorough study of the many functions, processes and regulatory pathways possessed by this Archaeon. The complete sequence will be the starting point for many biochemical, physiological and genetic studies. The complete genome sequence of the Euryarchaeote,

Methanococcus jannaschii has been recently reported and provides the first opportunity to compare complete genetic complements and biochemical pathways among the three domains of life (Bult *et. al.*, 1996).

Sulfolobus may be useful for industrial applications with its ability to oxidize sulfur (Kargi and Robinson, 1982) and to dissolve pyrite (Lindstrom *et al.*, 1993) and as a source of a complete set of thermostable enzymes. Its thermostable enzymes may be of commercial value especially in the cases of degradation of organosulfur compounds such as rubber tires and coal and the extraction of metals from ores (Norris, 1992).

The unique position of *Sulfolobus* in the phylogenetic tree is important for the study of genome evolution. A completed physical map of the genome of this thermoacidophilic Archaeon, *Sulfolobus solfataricus* P2 can be used for genomic comparisons at various taxonomic levels to allow evolutionary questions to be addressed. The genes and their organization in this archaeon will enable us to formulate and test hypotheses about the evolution of genes and of genomes. The comparison of complete sequences from several genomes may give us insight about the common ancestor of prokaryotes and eukaryotes, or at a deeper level, the last common ancestor of all extant life.

The genome of *S. solfataricus* P2 is among the largest in the Archaea being 3.0 +/- 0.1 Mbp and A-T rich (about 36% G-C content). The low G-C content makes it easier to sequence without having to deal with band compressions which is a common problem associated with the sequencing of

G-C rich DNA. This particular strain is found growing aerobically in the hot, sulfur springs of Naples, Italy. It grows optimally at acidic conditions of pH 2 to 4 and temperatures of 75 to 80°C, with a maximum at 87°C. It is able to oxidize sulfur into sulfuric acid. Isolated colonies can be easily grown on plates of gellan gum (Grogan, 1989). There exists a potential to develop tools for genetic analysis with the discovery of viruses (Schleper *et al.*, 1992), small curable plasmids (*Zillig et al.*, 1994) and a large conjugative plasmid (Schleper *et al.*, 1995). A recent article by Grogan (1996) has provided the first genetic evidence of chromosomal exchange and recombination in *Sulfolobus.*

B) Physical Mapping of a Genome

Top-down or macrorestriction mapping and bottom-up or contig mapping methods are commonly used to construct a physical map of a genome. These two methods offer different levels of resolution for the constructed physical maps. Low-resolution macrorestriction maps are generated by top-down methods whereas high-resolution maps represented by a set of overlapping clones are generated by the bottom-up approach.

The top-down mapping approach has produced the most publications (for reviews, see Smith and Condemine, 1990; Cole and Saint Girons, 1994; Derkacheva and Kagramanova, 1994; Fonstein and Haselkorn, 1995) because it is easier and faster, but these maps lack detail and usually none of the genome gets cloned in the process. Bottom-up mapping has generated fewer complete maps (Kohara *et al.*, 1987; Charlebois *et al*, 1991; Fonstein and Haselkorn,

1993; St. Jean *et al.*, 1994) because they are more laborious to construct. The advantage of using the bottom-up method is that the genome gets subdivided into manageable pieces and the ordered set of clones is valuable for producing genetic maps, determining the organization of the genome and DNA sequencing.

Bottom-up approaches depend on an efficient means to detect overlaps between clones. There are a variety of methods that have been used. Some are based on DNA hybridization (Craig *et al.*, 1990) or PCR (Green and Olson, 1990). Others used chromosome walking (Wenzel and Herrmann, 1989). There are those that are based on restriction-digest analysis such as the fingerprinting method (Coulson *et al.*, 1986; Olson *et al.*, 1986), partial digest mapping (Kohara *et al.*, 1987) and landmark analysis (Charlebois *et al.*, 1989).

There are few genome maps of Archaea compared with Bacteria. At present, eight archaeal genomes have been physically mapped (Table 1). Physical maps have been generated for *Halobacterium* sp. GRB (St. Jean *et al.*, 1994), *Halobacterium halobium* NRC-1 (Bobovnikova *et al.*, 1994), *Haloferax volcanii* DS2 (Charlebois *et al.*, 1991), *Haloferax mediterranei* (Lopez-Garcia *et al.*, 1992), *Thermococcus celer* (Noll, 1989), *Methanococcus voltae* (Sitzmann and Klein, 1991), *Methanobacterium thermoautotrophicum* Marburg (Stettler and Leisinger, 1992), and *Sulfolobus acidocaldarius* 7 (Kondo *et al.*, 1993; Yamagishi and Oshima, 1990). Most of these are low resolution macrorestriction maps except for *H. volcanii* DS2 and *Halobacterium* sp. GRB. High resolution contig maps have been generated for these two archaeal genomes.

Table 1 Physical maps of archaeal genomes and their G-C content.

Organism	Genome Size(Mbp)	Mapping Method	Reference	DNA(mole% GC)	Reference
Halobacterium halobium NRC-1	2.4	PFGE	Bobovnikova *et al.*, 1994	66-71	Staley *et al.*,1989
Halobacterium sp. GRB	2.47	Cloning	St. Jean *et al.*, 1994	66-71	Staley *et al.*, 1989
Haloferax mediterranei ATCC33500	3.84	PFGE	Lopez-Garcia *et al.*, 1992	60	Staley *et al.*, 1989
Haloferax volcanii DS2	4.14	Cloning	Charlebois *et al.*, 1991	65	Gutierrez *et al.*, 1989
Methanobacterium thermoautotrophicum Marburg	1.6	PFGE	Stettler and Leisinger, 1992	48	Jones *et al.*, 1987
Methanococcus voltae PS	1.9	PFGE	Sitzmann and Klein, 1991	30	Klein and Schnorr, 1984
Sulfolobus acidocardarius 7	2.76	PFGE	Kondo *et al.*, 1993	40	Yamagishi and Oshima, 1990
Thermococcus celer Vu13	1.89	PFGE	Noll, 1989	57	Zillio *et al.*, 1983

Summary of archaeal genomes that have been mapped by the two different methods: cloning and PFGE, pulsed-field gel electrophoresis. References for these physical maps are given. Genome sizes have been determined by mapping. G-C content of the respective genomes are listed along with their references.

C) Landmark Analysis

The landmark strategy, a bottom-up strategy, has been successfully used in generating contig maps of the Archaeal genomes of *Haloferax volcanii* DS2 (Charlebois *et al.*, 1991) and *Halobacterium* sp. GRB (St. Jean *et al.*, 1994).

The landmark strategy involves comparing the gel patterns of cosmid clones produced from a digest using the cloning enzyme with a digest using the cloning enzyme in combination with an infrequently-cutting enzyme. If two clones possess this rare site (recognition site of the infrequent-cutting enzyme), they can be linked by this "landmark". This rare site or "landmark" can be recognized because it cuts a standard-frequency restriction fragment (generated by the cloning enzyme) of a certain size into subfragments of distinctive size (Charlebois, 1993). The detection of overlaps between clones depends on the ability to identify these "landmarks". The identified overlaps between clones can be confirmed using additional infrequent-cutting enzymes.

Although the landmark strategy is a simple and efficient way to map a genome, it has its limitations. Gaps in the physical map may result from a lack of a landmark which prevents an overlap between clones to be detected or there are regions of the genome that are unclonable.

There is a chance that a landmark is not found which prevents an overlap between two clones to be determined. The identified contigs can be linked up by hybridizing cosmid DNA from each of the contig ends to the cloning enzyme digests of the collection of end cosmids. In the mapping projects of *H. volcanii*

DS2 and *Halobacterium* sp. GRB, Charlebois *et al.* (1991) and St. Jean *et al.* (1994) had to contend with unclonable DNA. There are three causes for uncloned regions: 1) Too few clones were examined; 2) A biased partial-digest clone library resulting from local deficiencies or excesses of sites for the cloning enzyme in certain regions of the genome; and 3) DNA that is unstable or incompatible within the host strain.

Uncloned DNA may pose a problem, but can be minimized by two approaches. One method is to examine as many clones as possible from two independent partial-digest libraries generated with different cloning enzymes, each enzyme having its own peculiar bias. The other method involves studying one library extensively by landmark analysis and then performing chromosome walking in a second library. Chromosome walking is an efficient method to extend contigs and fill gaps, but is more laborious and time-consuming than the landmark approach. In the *Halobacterium* sp. GRB mapping project, ten contigs were identified after landmark analysis of over 45 genome equivalents of cosmid DNA from the first library and the contigs were either extended or joined by chromosome walking into a second library. In the end, 99% of the genome was cloned as cosmids, in contrast to only 96% of the *H. volcanii* genome using a single-library approach.

The two approaches of examining two independent partial-digest libraries by landmark analysis and examining one library by landmark analysis and chromosome walking in the second library will resolve the problem of uncloned

regions caused by the examination of too few clones and a biased partial-digest clone library, however these two methods cannot resolve the problem of uncloned regions caused by DNA that is unstable or incompatible with the host strain.

Regions of a genome may not be clonable as a result of its instability or incompatibility with the host system. Foreign DNA cloned into vector can be introduced into a host for propagation by the mechanisms of transformation and transduction. Once the foreign DNA is inside the host, its viability and propagation will be under host-control therefore its compatibility with its host is paramount.

Most molecular cloning has been done in *Escherichia coli* because it grows well and rapidly, remains stable in culture, is transformable by DNA and there exists an extensive knowledge of its genetics and biochemistry (Neidhardt *et al.*, 1996). Wild type *E. coli* employs a restriction and modification system to protect itself from the presence of foreign DNA. It codes for a restriction enzyme, *Eco* K, that cleaves DNA containing unmodified *Eco* K recognition sites, but there are strains of *E. coli* that have been engineered to be defective in restriction and restriction and modification functions to accommodate foreign DNA (Yuan, 1981).

E. coli codes for several pathways such as RecA, RecBCD, RecE and RecF that are involved in the recombination of homologous sequences of DNA (Lloyd and Low, 1996). Roth *et al.* (1996) suggest the recombination system is

most heavily used for DNA repair and to maintain gene order by minimizing chromosomal rearrangement. Rearrangement of the bacterial chromosome occurs either by homologous recombination between repeat and insertion sequence elements or by site-specific recombination (Krawiec and Riley, 1990). Deletions, duplications and inversion are the causes of rearrangement. Recombinants that may be vulnerable to rearrangement and deletion of cloned sequences should be propagated in a host that is defective in one of the recombination pathways. Typically, RecA mutants are used as hosts because they are defective in this main recombination pathway. The instability of the cloned segments of a genome are attributed to the presence of repeated and insertion sequence (IS) elements which are sites for homologous recombination. The mobility of insertion sequence elements are also responsible for genome instability (Berg and Howe, 1989).

Incompatibility of cloned DNA with its host will result in deletions and the inability to clone a segment(s) of a genome. Reasons have been put forth to explain this cloning predicament. It has been suggested that the growth of host cells are affected by the increased dosage of a particular gene(s) on the cloned segment (Kohara *et al.*, 1987) and the toxicity of the gene products to the host (Bukanov and Berg, 1994).

The capacity to construct a high resolution genome map depends on the ability to clone the entire genome. The instability of cloned DNA or its incompatibility with the host system may be a detriment to a genome mapping

project, but it should not prevent the undertaking of such a venture. Kohara *et al.* (1987) were able to clone approximately 99% of the *E. coli* genome into lambda phage vector overcoming these problems associated with cosmid contig mapping (Bukanov and Berg, 1994; Fonstein *et al.*, 1995).

D) Thesis Objective

The goal was to construct a contig map of the genome of *S. solfataricus* P2 using the technique of landmark analysis (Charlebois *et al.*, 1989; Charlebois *et al.*, 1991; Charlebois, 1993; St. Jean *et al.*, 1994). In generating a high resolution contig map, the whole genome is cloned into a minimal set of clones. This ordered set of cosmid clones serves as an archive of all the information about the structure of the genome of this thermoacidophilic archaeon and thus facilitates easy communication between researchers regarding the disposition of the clones. The use of cosmid vectors has the advantage of cloning segments of DNA with an approximate size of 40 kb and this allows for the cloning of operons without disruption. Ordered cosmids also facilitate a more directed approach for determining the sequence of the genome; an entire genome is cloned and ready to be sequenced. The strategy employed to sequence the entire genome is on a clone-by-clone, one-cosmid-at-a-time basis. The cosmid clones used in the construction of the physical map have been provided to our sequencing team to be DNA sequenced.

Materials and Methods

A) Sizing the Genome of *Sulfolobus solfataricus* P2 by Pulsed-Field Gel Electrophoresis (PFGE)

Genomic DNA of *Sulfolobus solfataricus* P2 embedded in agarose gel plugs were prepared according to Yamagishi and Oshima (1990) from 20 ml of *S. solfataricus* P2 cell culture and provided to us by Ford Doolittle's laboratory at Dalhousie University in Halifax, Nova Scotia. The genomic DNA was cut with the following restriction enzymes: *Apa* I, *Bss* HII, *Eag* I, *Mlu* I, *Nar* I, *Ngo* MI, *Not* I, *Rsr* II, Sac II, *Sma* I and *Xho* I.

Each restriction digest was performed in a 200 µl reaction volume with 40 units (U) of each enzyme and 1x of each corresponding 10x restriction enzyme buffer. 2 µl (100 µg/ml) bovine serum albumin (BSA) was also added to the *Apa* I restriction digest. The plugs were equilibrated in 410 µl of distilled (d)H_2O and 50 µl of 10x restriction enzyme buffer for 30 minutes on ice. 1 µl (100 µg/ml) BSA was added to the *Apa* I tube for equilibration. After this equilibration time, the buffer was removed gently with a pipette. Once the digestion reactions were set up, they were gently mixed and set on ice for 20 minutes. The plugs were then digested for 3 hours at the optimum temperature of each restriction enzyme.

Each digested agarose plug was loaded into a 1% SeaKem GTG agarose gel [0.7 g SeaKem GTG Agarose + 70 ml TBE buffer (45 mM Tris, 45 mM boric acid and 1 mM EDTA)] and resolved at 170 Volts (V), 12°C, and a 22 second

switch time for 24 hours in a pulsed-field gel apparatus manufactured by Tyler Scientific Instruments filled with TBE buffer. ProMega-Markers Yeast Chromosomal DNA and ProMega-Markers Delta-39 were resolved alongside the digested DNA plugs. After this 24 hour period, the gel was then stained in ethidium bromide for 30 minutes and then destained in water for 15 minutes. The gel was then photographed with a Polaroid camera under ultraviolet (uv) lighting. The genomic DNA digests with *Bss* HII and *Eag* I gave a number of well resolved fragments and the respective fragments were sized relative to the DNA size markers.

B) Construction of the *Hind* III and *Bam* HI Random Cosmid Clone Libraries

Dr. Charlebois performed the partial digestions of the genomic DNA of *S. solfataricus* P2 for both cosmid clone libraries. Dr. Charlebois prepared the cosmid vector, Tropist 3 for the *Hind* III library and I prepared the vector for the *Bam* HI library. Dr. Charlebois carried out the ligation, host cell preparation, packaging and infection reactions for the cosmid clones designated sh01a01 to sh04h12 of the *Hind* III library. I carried out these reactions for the clones designated sh05a01 to sh26h12 of the *Hind* III library and for the clones of the *Bam* HI library.

i) Determination of Partial Digest Conditions of *S. solfataricus* Genomic DNA

In constructing the *Hind* III cosmid library, S. *solfataricus* P2 genomic DNA was digested under partial digest conditions beginning with 2 Units (U) and

decreasing by a factor of two down to 1/64 U for 30 minutes at 37°C and zero Units as a control. The partial digest conditions of 1/4 and 1/8 U were chosen for the first 384 cosmid clones, designated sh01a01 to sh04h12. The partial conditions of 1/8 and 1/16 U were chosen for the cosmid clones designated sh05a01 to sh26h12. Each numeric series designated sh01 to sh26 is comprised of 96 cosmid clones.

A different restriction enzyme dilution series was used in constructing the *Bam* HI library, starting from 32 U and decreasing by a factor of 2 down to 1/64 U. The partial digest conditions of 1/2, 1 and 2 U were chosen for this cosmid clone library.

ii) Preparation of Cosmid Vector, Tropist 3

Dr. Charlebois prepared the cosmid vector, Tropist 3 by digesting with *Sca* I, and following phosphatase treatment, digesting it with *Hind* III for the *Hind* III library . I prepared the cosmid vector, Tropist 3 for the *Bam* HI library. Tropist 3 was prepared as follows:

a) Linearizing Tropist 3 with *Sca* I

50 µl (50 µg) of cosmid vector, Tropist 3 was digested in a reaction volume of 100 µl with 40 U of *Sca* I and 1x *Sca* I buffer for 3 hours at 37°C. It was then treated with 100 µl of 1:1 ratio of phenol/chloroform and centrifuged at 2500g for 5 minutes. The aqueous phase containing the DNA of this wash was then extracted and placed into a new 1.5 ml Eppendorf microcentrifuge tube.

9.5 µl of 3 M sodium acetate pH 5.5 and 200 µl of 95% ethanol were added to precipitate the DNA. This tube was then centrifuged at 2500g for 7 minutes. The supernatant was removed by aspiration. The pellet was washed with 150 µl of 80% ethanol and then centrifuged at 2500g for 3 minutes. The supernatant was removed by aspiration. The ethanol-washed pellet was allowed to air dry on the bench for 30 minutes. It was then resuspended in 50 µl of TE [10 mM Tris-HCl pH 7.6; 1 mM EDTA pH 8.0] at 37°C for 30 minutes. 2 µl of the *Sca* I-digested vector was checked for complete digestion by resolving it in a 1.1% agarose gel at 2 V/cm for 18 hours in TAE buffer [50 mM Tris, 2 mM EDTA, 20 mM sodium acetate, 1.8 ml/L acetic acid]. 10 µl of DNA size marker, Mappers' Marker Mix [MMM: 100 µl (15 µg) of *Bst*E II-cut λ DNA, 100 µl (5 µg) of *Xho* I-cut λ DNA, 100 µl (5 µg) of *Xba* I-cut λ DNA, 100 µl of TE, 100 µl of 0.25M EDTA and 150 µl of loading buffer (stock solution: 50% glycerol, 10 mM Tris-HCl pH 7.6 to 8, 20 mM EDTA and a pinch of bromophenol blue)] was resolved alongside the digested vector. The gel was stained with ethidium bromide for 40 minutes and destained in water for 20 minutes. It was then photographed with a Polaroid camera under uv lighting.

b) Removal of 5' end phosphate group of the *Sca* I-digested Tropist 3 vector

The remaining 48 µl of *Sca* I-digested vector was treated with 1 U of phosphatase and 25 µl of 10x phosphatase buffer in a reaction volume of 250 µl for 30 minutes at 37°C and then for 15 minutes at 60°C. The *Sca*I-digested and

phosphatased vector was treated with a 1:1 ratio of phenol/chloroform. The aqueous phase containing the DNA was then placed into a new 1.5 ml Eppendorf tube. The DNA was precipitated with 24.5 µl of 3 M sodium acetate pH 5.5 and 500 µl of 95% ethanol. The tube was then centrifuged at 2500*g* for 7 minutes. The supernatant was removed by aspiration. The pellet was treated with 375 µl of 80% ethanol. The ethanol was removed by aspiration and the pellet was allowed to dry on the bench for 30 minutes. The pellet was then resuspended in 86 µl of TE for 30 minutes at 37°C.

c) Digestion of the *Sca* I-digested and phosphatased Tropist 3 vector with *Bam* HI

The *Sca* I-digested and phosphatased vector was then digested with 40 U of *Bam* HI with 1x of 10xRE buffer [500 mM Tris-HCl pH 7.6, 500 mM KCl, 100 mM $MgCl_2$ and 100 mM dithiothreitol] in a 100 µl reaction volume for 4 hours at 37°C. The digestion product was recovered following the procedure described in section B) ii) a).

The *Sca* I-linearized, phosphatased and *Bam* HI-digested vector was ligated with 0.3 U of T4 DNA ligase and 1x of 10x ligase buffer in a 10 µl reaction volume for 2 hours at room temperature. This ligation reaction was performed in order to determine whether the vector would ligate itself.

The ligated and unligated products of the *Sca* I-linearized, phosphatase treated and *Bam* HI-digested vector were resolved in a 1.1% agarose gel at

9 V/cm for 4 hours in TAE buffer. The gel was stained, destained and photographed following the procedures in section B) ii) a).

iii) Ligation of treated Tropist 3 vector with *S. solfataricus* partial digests

The DNA partial digestion conditions of 1 U, 2 U and 1/2 U were selected for ligation with the *Sca* I-linearized cosmid vector treated with phosphatase and digested with *Bam* HI. The ligation reaction was composed of 1.1 µl (0.5 µg) of *Bam* HI-, *Sca* I-digested and phosphatased Tropist 3 vector, 2 µl of 10x ligase buffer, 1 µl (1 U) of T4 DNA ligase, 8 µl (0.6 µg) of 1 U-partial, 4 µl (0.6 µg) of 2 U-partial and 4 µl (0.6 µg) of 1/2 U-partial and resolved in a 1.1% agarose gel at 2V/cm overnight at 12°C.

The ligation reaction was performed by Dr. Charlebois for the cosmid clones designated sh01a01 to sh04h12 in the *Hind* III library. 7 µl of 1/8 U partials and 4.5 µl of 1/4 U partials were combined with 1 U T4 DNA ligase, 1x of 10x ligase buffer and 1.1 µl (0.5 µg) of phosphatase treated, *Sca* I and *Hind* III digested Tropist 3 cosmid vector in a 20 µl reaction volume for 12 hours at 14 to 14.5°C.

I performed the ligation reactions for the cosmid clones designated sh05a01 to s26h12 of the *Hind* III library. 6 µg of 1/8 U and 1/16 U partially digested genomic DNA of *S. solfataricus* were each separately ligated to 0.5 µg of Tropist 3 cosmid vector linearized with *Sca* I, treated with phosphatase and

digested with *Hind* III. Ligations were performed at 14°C for 12 hours in a reaction volume of 20 µl with 1 U of T4 ligase and 1x of 10x ligase buffer.

iv) Preparation of Host Cells, *Escherichia coli* ED8767

Host cells were prepared by Dr. Charlebois for the cosmid clones designated sh01a01 to sh04h12 of the *Hind* III library. I prepared the host cells for the cosmid clones designated sh05a01 to sh26h12 of the *Hind* III library. Host cells were prepared as follows:

A single colony of the host, *E. coli* ED8767 was picked from a YT agar plate grown overnight at 37°C. This colony was inoculated into 3 ml of YT broth and was grown overnight at 37°C with shaking. 1/2 ml of this overnight culture was used to inoculate 50 ml of YT-maltose-$MgSO_4$ in a sterilized 250 ml Erlenmeyer flask. 8 ml samples of this culture were taken for monitoring the growth of the cells with a Bausch & Lomb Spectronic 20 spectrophotometer set at an absorbance with a wavelength of 600 nm. The desired absorbance reading is 1.0. The culture was grown for 6 1/2 hours at 37°C with shaking to give this desired measurement. The cells were centrifuged at 400*g* for 10 minutes. The supernatant was discarded and the pellet was resuspended in 10 mM $MgSO_4$ to give an absorbance reading of 0.5.

In the *Bam* HI library host cells were prepared according to the procedure stated for the *Hind* III library section with the following changes: 1 ml samples of the overnight culture were taken for monitoring the growth of the cells and the

culture was grown for 3 1/2 hours at 37°C with shaking to give the desired absorbance measurement of 1.0.

v) Packaging of DNA

DNA was packaged, according to the instructions of the Stratagene Gigapack II XL kit, by Dr. Charlebois for the cosmid clones sh01a01 to sh04h12 of the *Hind* III library. I followed the same instructions for packaging DNA for the cosmid clones sh05a01 to sh26h12 of the *Hind* III library and the clones of the *Bam* HI library.

vi) Infection

The infection of the host cell, *Escherichia coli* ED8767, with packaged cosmid DNA was performed by Dr. Charlebois for the clones sh01a01 to sh04h12 of the *Hind* III library. Cell suspensions of 30 and 100 µl were spread onto three and ten YT-kanamycin agar plates respectively and incubated overnight at 37°C.

I performed the infection of *E. coli* ED8767 with packaged cosmid DNA for the clones sh05a01 to sh26h12 of the *Hind* III library, 150 µl of 1/8 U- and 1/16 U-partial condition, packaged cosmid DNA was each mixed separately with 1.5 ml of prepared host cells for 20 minutes at 37°C with gentle shaking. 3.75 ml of YT broth warmed to 37°C was added to each mixture and gently shaken for 40 minutes at 37°C. Each mixture was then centrifuged at 650*g* for 5 minutes. The supernatant was discarded and the pellet was resuspended in 1.3 ml of YT

broth. Cell suspensions of 40 and 100 µl were spread onto four and six YT-kanamycin agar plates respectively for both partial digest conditions and incubated overnight at 37°C.

I also performed the infection of the host cell with cosmid DNA for the *Bam* HI library following the same procedure used for the *Hind* III library except that 150 µl of a phage DNA mixture was used and cell suspensions of 30, 40 and 100 µl were plated onto three, five and ten YT-kanamycin agar plates respectively. Single colonies were picked from these plates and inoculated into the 96 wells of a Titertek plate each containing 100 µl of YT-kanamycin broth. A total of 1728 colonies were inoculated into 18 Titertek plates. These plates were labelled sb50 to sb67. 25 µl of 75% glycerol was then added to each well to preserve the cells during storage at -80°C.

C) Alkaline Extraction of Cosmid DNA from the Clones of the *Hind* III Library

In constructing the random *Hind* III cosmid clone library, single colonies from the initial spread plates were grown overnight in 10.5 ml YT-kan broth at 37°C. 0.4 ml samples from these cultures were placed into Titertek tubes, preserved with 0.2 ml of 50%glycerol/50%YT and stored at -80°C. The remaining cell culture was centrifuged at 1200g for 7 minutes. The media was discarded and the cell pellet was resuspended in 200 µl of TEG [20 mM Tris-HCl pH 8.0, 50 mM EDTA, 1% glucose]. The cells were lysed with 400 µl of 0.2 M NaOH/1% SDS and neutralized with 300 µl of -20°C, 7.5 M ammonium acetate.

The cell suspensions were mixed by rocking the tubes. The neutralized lysate was transferred to a 1.5 ml Eppendorf tube using a disposable polypropylene transfer pipette. The tube was centrifuged at 18000g for 7 minutes. The supernatant was decanted into a 2 ml BioRad tube. 500 µl of 1:1 ratio of phenol/chloroform was added to the supernatant and this suspension was vortexed, then centrifuged at 11000g for 5 minutes. Approximately, 800 µl of the aqueous phase containing the DNA was transferred to a 1.5 ml Eppendorf tube. 500 µl of isopropanol was added and the tube was mixed by shaking. The tube was then centrifuged at 18000g for 7 minutes. The supernatant was removed by two rounds of aspiration. The DNA pellet was dissolved in 300 µl of TE at 37°C for 20 minutes with occasional vortexing. 1 ml of a 1:6 ratio of 7.5 M ammonium acetate/ethanol mixture was added to the tube. The tube was mixed by shaking and then centrifuged at 18000g for 7 minutes. The supernatant was removed by two rounds of aspiration. 1 ml of 80% ethanol was added and the tube was shaken and centrifuged at 18000g for 3 minutes. Two rounds of aspiration were used to remove the supernatant. The DNA pellet in the tube was allowed to air dry on the bench. It was then dissolved in 50 µl of TE by vortexing and stored at 4°C.

D) <u>DNA Digestion of Cosmid Clones from the *Hind* III Library</u>

Each cosmid clone was digested under three different conditions: *Hind* III,

Hind III/*Mlu* I and *Hind* III/*Xho* I. 2 µl (200-600 ng) of cosmid clone DNA was digested in a 10 µl total reaction volume with 2 U of each enzyme and 0.02 µl of RNase, DNase-free (Boehringer Mannheim) at 37°C for 3 hours. The individual digests were stopped by adding 3µl of loading dye and loaded into the wells of a 1.1% agarose gel in the order: *Hind* III, *Hind* III/*Mlu* I and *Hind* III/*Xho* I. 12 clones were loaded into each gel. The procedures for resolving DNA, staining, destaining and photographing the gel are as stated in section B) ii) a).

E) Landmark Analysis of Cosmid Clones from the *Hind* III Library

The cosmid clones were sorted into groups according to their *Hind* III, *Hind* III/*Mlu* I and *Hind* III/*Xho* I digestion patterns by visual inspection of the gels. *Hind* III fragment(s) that contained *Mlu* I and/or *Xho* I restriction sites were identified as the fragments containing the "landmark(s)". These landmarks identify region(s) of overlap between cosmid clones. Clones with different digestion patterns that possessed these landmarks were linked into contiguous stretches of DNA called contigs. The identification of landmarks was performed by visual inspection of the gels of restriction digests.

F) Constructing an Integrated Contig/Macrorestriction Map

i) Immobilizing *Bss* HII and *Eag* I Macrorestriction Fragments onto Membrane

Agarose gel plugs containing genomic DNA of *S. solfataricus* P2 were digested with *Bss* HII and *Eag* I separately. The procedures for restriction digests, pulsed-field gel electrophoresis, staining, destaining and viewing of the

gel are the same as stated in section A). Each of the *Bss* HII and *Eag* I macrorestriction fragments were individually cut out of the gel and GeneCleaned using the GeneClean II kit by Bio 101 Inc. The individual *Bss* HII and *Eag* I macrorestriction fragments were immobilized onto their respective membranes. 3.5 µl of DNA was combined with 3.5 µl of 0.8M NaOH to give a final concentration of 0.4 M NaOH in a well of a microtiter plate. They were mixed in the well by pipetting and then blotting onto GeneScreen hybridization transfer membrane. The spots were allowed to air dry on the bench. Once dried, the membrane was rinsed in 2xSSC [0.3 M NaCl, 0.035 M tri-sodium citrate]. The membrane was allowed to partially dry on the bench. It was then uv-irradiated for 5 minutes and then stored in a Ziplock bag.

ii) Identification of Junction Clones

Cosmid clones from the minimum set were digested separately with *Bss* HII at 50° and *Eag* I at 37°C following the procedures outlined in section D).

iii) Hybridization of Junction Clones Against *Bss* HII and *Eag* I Fragments Immobilized on Membrane

a) Making Probes of *Hind* III Fragments Containing Either *Bss* HII or *Eag* I site(s)

Clones possessing the restriction site(s) for *Bss* HII or *Eag* I were digested with *Hind* III and *Hind* III in combination with either *Bss* HII or *Eag* I following the procedures stated in section D). Each digested clone was loaded into the wells of a 1.1% agarose gel in the order: *Hind* III, *Hind* III/*Bss* HII or

Hind III/Eag I. The *Hind* III fragment containing either the *Bss* HII or the *Eag* I site(s) was cut out of the gel and placed into a 1.5 ml Eppendorf tube. The agarose block containing the *Hind* III fragment was GeneCleaned using the GeneClean II kit manufactured by Bio 101 Inc.

b) Hybridization of *Hind* III Fragments Containing Either *Bss* HII or *Eag* I site(s) Against *Bss* HII and *Eag* I Macrorestriction Fragments Immobilized onto Membrane

Separate membranes containing immobilized *Bss* HII and *Eag* I macrorestriction fragments were pre-hybridized at 36°C for at least an hour in a hybridization tube containing 15 ml of hybridization solution [29.2 g of sodium chloride (1M), 25 ml of 1M Tris-HCl pH 7.6 (50 mM), 25 g of SDS (5%) and 250 ml of formamide in 500 ml (50%)] and 750 μg of herring sperm DNA that was boiled for 5 minutes. This was carried out in a Tyler Research Instruments HI-16000 Hybridization Incubator.

2.5 μl of the GeneCleaned *Hind* III fragment possessing either *Bss* HII or *Eag* I site(s) was combined with 5 μl of water and 1 μl of (2.5 mg/ml) random hexamer primer in a 0.5 ml Eppendorf tube. The tube was then boiled for 5 minutes. 1.25 μl of 10x RP-C (Random Priming using [α-^{32}P]dCTP) buffer [200 mM Tris-HCl pH8.0; 100 mM $MgCl_2$ $6H_2O$; 50 mM dithiothreitol; 600 μM dGTP; 600 μM dTTP; 600 μM dATP; 25% glycerol], 1.5 μl of BSA, 1.0 μl of (10 μCi/μl) [α-^{32}P]dCTP and 0.5 μl of (2 U/μl) Klenow were added to the tube. The tube was incubated at room temperature for 5 minutes and then at 37°C for 45 minutes. 6 μl of ice-cold 7.5 M ammonium acetate and 32 μl of ethanol was

then added to the tube. It was vortexed and then centrifuged at 11000g for 5 minutes. The supernatant was removed by pipette and the DNA probe was resuspended in 100 µl of TE. The tube of DNA probe was boiled for 2 minutes and then immediately placed on ice. The tube of probe was measured for the amount of radioactivity by a Geiger counter and then recorded. The probe was added to the hybridization tube of pre-hybridizing dot blots. The hybridization reaction ran overnight at 36°C.

The following day, the probe-containing hybridization solution was discarded and the membranes were rinsed twice with 2xSSC solution. The 2xSSC solution was discarded after both rinses. The hybridization tube was then filled 3/4 of the way with 2xSSC/1%SDS solution and the dot blots were washed for 1 hour at 65°C . The membranes were rinsed twice with 2xSSC solution to ensure that all nonspecifically bound probe was removed and after each rinse the solution was discarded. They were air dried for a short period, wrapped with FISHERbrand all-purpose laboratory wrap and then placed into a film cassette along with Kodak Scientific Imaging X-OMAT AR 35 cm X 43 cm film and an intensifying screen. The cassette was placed into a -80°C freezer. Film exposure time varied depending on the signal of the probe. It varied from 2 days to a week or more. The film was developed in Kodak Film Developer for 3 minutes and then fixed in Kodak Film Fixer for 45 seconds. The film was rinsed in water after being placed in developer and fixer and was hung for air drying.

c) Stripping of Probe from Membrane with *Bss* HII or *Eag* I Macrorestriction Fragments Immobilized on it

Membrane containing bound probe was placed into a half-filled hybridization tube containing 70°C heated stripping solution [5.0 g of SDS; 10.46 ml of 1M sodium phosphate, dibasic; 4.06 ml of 1 M sodium dihydrogen orthophosphate (monobasic); 480 ml of H_20 and 500 ml of formamide per litre]. Removal of DNA probe from membrane was carried out in a Tyler Research Instruments HI-16000 Hybridization Incubator for a period of 60 to 90 minutes at a temperature range of 65 to 75°C. The membranes were rinsed twice with 2xSSC solution and after each rinse the solution was discarded. They were air dried for a short period, wrapped with FISHERbrand all-purpose laboratory wrap and then placed into a film cassette along with Kodak Scientific Imaging X-OMAT AR 35 cm X 43 cm film and an intensifying screen. The cassette was placed into a -80°C freezer. The film was exposed for about a day and then developed to ensure that most of the probe had been removed. The film was developed according the procedures stated in section F) iii) b).

G) Extension of the Identified Contigs from the *Hind* III Library

i) Hybridization using Dot Blots of the Minimal Set of Cosmid Clones

a) Making Dot Blots of the Minimal Set of Cosmid Clones

Dot blots of the minimal set of cosmid clones were made following the procedure for immobilizing *Bss* HII and *Eag* I macrorestriction fragments onto

membrane except that 1 μl of cosmid DNA was combined with 2 μl of 0.6 M NaOH to give a final concentration of 0.4 M NaOH in a well of a microtiter plate.

b) Making Probes of *Hind* III Fragments Representing the Ends of Identified Contigs

3 μl of cosmid DNA of clones from the end of identified contigs were digested following the procedures stated in section D). The digestion reaction was resolved and the gel was stained, destained and photographed following the procedures stated in section B) ii) a). The *Hind* III fragment representing one of the ends of the identified contigs was cut out of the gel and placed into a 1.5 ml Eppendorf tube. The agarose block containing the *Hind* III fragment was GeneCleaned using the GeneClean II kit manufactured by Bio 101 Inc.

c) Hybridization of *Hind* III Fragments Against Dot Blots

Hybridization of *Hind* III fragments against dot blots were performed following the previously mentioned hybridization procedure except that pre-hybridization and hybridization occurred at 40°C and the washing occurred at 70°C. Film exposure time was also different varying from 2 hours to 2 days.

d) Stripping of Probe from Dot Blots

Removal of probe from dot blots were performed according to the previously mentioned protocol.

ii) Landmark Analysis of the Minimal Set of Cosmid Clones with *Eag* I

The same procedure was followed as stated in sections D) and E) except that *Hind* III and *Eag* I were used in the double digest and 3 μl of cosmid DNA was used for the single and double digests.

iii) Hybridization using Southerns of the Minimal Set of Cosmid Clones

This procedure was carried out by Ghislaine Allard.

Cosmid clones from the minimum set were digested with *Hind* III and resolved in agarose gels. The digested clones were examined under uv lighting. The gels were then vacuum transferred onto GeneScreen Hybridization Transfer membrane with a Tyler Research Instruments Vacuum Transfer Apparatus. Probes of *Hind* III fragments representing ends of identified contigs as well as complete cosmid clones were used to hybridize against the Southerns of the minimal set of cosmid clones.

iv) Hybridization against *Bam* HI Cosmid Clone Library

This procedure was carried out by Christina Chan.

Colony blots were made from the *Bam* HI cosmid clone library. *Hind* III fragments representing the ends of identified contigs were used as probes to hybridize against these blots. The colonies that hybridized with the probes were grown in YT-kanamycin broth and alkaline extracted for their cosmid DNA. The extracted DNA was digested with *Bam* HI and resolved in an agarose gel to examine their digestion pattern. These clones that potentially extend the linked

contigs had their ends DNA sequenced with an automated sequencer to confirm an overlap by sequence homology.

Results

A) Sizing the Genome of *Sulfolobus solfataricus* P2 by Pulsed-Field Gel Electrophoresis (PFGE)

Genomic DNA of *S. solfataricus* P2 was digested with 11 restriction enzymes: *Apa* I, *Bss* HII, *Eag* I, *Mlu* I, *Nar* I, *Ngo* MI, *Not* I, *Sac* I, *Sma* I, *Rsr* II and *Xho* I. The restriction enzymes, *Eag* I and *Bss* HII cut the genome into a number of relatively well resolved bands which were used to determine its size. The 16 *Bss* HII macrorestriction fragments in Figure 1 totalled up to 2940 kbp while the 17 *Eag* I macrorestriction fragments in Figure 1 totalled up to 3000 kbp. These results estimate the genome size of *S. solfataricus* P2 to be 3.0 +/- 0.1 Mbp with a 5% error in sizing the individual fragments against the DNA size markers, ProMega-Markers Yeast Chromosomal DNA (size range of 245 to 2000 kb) and ProMega-Markers Delta-39 (size range of 39 to 858 kb).

B) *Hind* III Random Cosmid Clone Library

The *Hind* III library consists of 2496 cosmid clones of which, 1296 clones have been analyzed by the landmark strategy and the rest remain frozen and unanalyzed. 59 out of the 1296 clones yielded no cosmid DNA. These clones did not yield any cosmid DNA probably as a result of failed DNA extractions. It is unlikely these clones did not contain cosmid DNA because the agar plates contained the antibiotic, kanamycin which selected for kanamycin-resistant colonies and kanamycin resistance is conferred upon colonies that have cosmid DNA. Of the 1296 cosmid clones, 36 have deleted or are in the process of

deleting their inserts. Deleted or deleting clones can be identified by the characteristic bright vector band or the bright vector band and faint DNA insert bands respectively. The remaining 1201 cosmid clones are relatively stable and amount to 16 genome equivalents of DNA with an average of 40 kb inserts.

The first 384 cosmid clones of this library with the series designation sh01 to sh04 had a bias for the cloning of large *Hind* III fragments because of the chosen partial-digestion conditions of 1/4 and 1/8 units which resulted in the exclusion of the cloning of smaller *Hind* III fragments. In order to overcome this bias for the cloning of large *Hind* III fragments, the partial-digestion conditions of 1/8 and 1/16 units were chosen for the clones with the series designation sh05 to sh26. Although this DNA insert bias was overcome for these latter series of clones, some clones from the series sh07 to sh12 showed signs of deletion. Most of the clones from series sh13 to sh24 appear stable, but recent results from restriction digests of some of these clones have shown them to be in the process of deleting their inserts. The alkaline DNA extraction of the clones from series sh05 to sh06 were lost during the extraction process, but frozen cell cultures are still available.

i) Identification of Unstable Cosmid Clones

The colonies that grew on the YT-kanamycin agar plates varied in size despite being exposed to the same growth conditions. The DNA extraction yields also varied. Certain clones resulted in very good DNA extraction yields demonstrated by the bright DNA bands resolved in the agarose gels. Other

cosmid clones had deleted or were in the process of deleting their insert *Hind* III band(s) evidenced by the faint and/or missing bands resolved in the agarose gels. 2.77% (36 out of 1296 clones) of the clones in the *Hind* III library are deleting or have deleted parts of or the entire insert.

ii) Identification of Contigs

(As of May 9, 1996)

The 1201 cosmid clones of the *Hind* III library were grouped according to their *Hind* III, *Hind* III/*Mlu* I and *Hind* III/*Xho* I restriction digest patterns. Clones that produced the same digestion patterns were grouped together. An individual clone from these groupings were then used for landmark analysis to construct contigs. Chromosome walking was also used to extend the identified contigs by landmark analyses. The clone collection of 1201 was reduced down to a minimal set of 105 cosmid clones which represent approximately 60 to 70% of the *S. solfataricus* P2 genome.

The minimal set of 105 cosmid clones have been linked into 31 contigs (Table 2) by landmark analysis with *Mlu* I and *Xho* I (Figure 2) and chromosome walking. Landmark analysis with *Mlu* I and *Xho* I identified one unusual, six variant and nine polymorphic clones which are a part of the minimal set (Table 2). There are clones categorized as "to be placed" in Table 2. These clones are those whose placement within a contig that have to be determined.

The unusual clone, sh02h03U maps to contig 2, but its digest pattern was unusual because some of the identical *Hind* III fragments from the *Hind* III and

Hind III/*Mlu* I restriction digests migrated differently in the *Hind* III/*Mlu* I lane compared with those in the *Hind* III lane; the bands of the *Hind* III/*Mlu* I lane were shifted down slightly. It does not appear that the *Mlu* I enzyme cuts any of the *Hind* III fragments of this clone so this shifting of the bands may be explained by the different consistencies in the agarose concentration between the lanes.

The clones, sh16c05; sh20f08U; sh14b10; sh19a12, sh01g03 and sh19g12U from contigs 7; 10; 12 and 15 respectively, were designated variant clones because they do not fit properly within the contig. These variant clones are uncharacterized; they may be polymorphic or may result from coligation. Recent testing by Ghislaine Allard have shown insertion sequence elements to be responsible for some of these variant clones.

Restriction fragment length polymorphisms (RFLPs) were identified in 9 clones that mapped to one end of contig 1 represented by cosmid clone, sh01e07. This region of contig 1 overlapped with clones, sh15e10U, sh02d04, sh13d04, sh15c01U, sh20b05U, sh22a12U, sh22g07, sh03f07 and sh02f03, but these clones did not extend each other. The polymorphisms are attributed to insertion sequence elements. An honours student, Todd Monkman studied the polymorphic nature of these clones.

Certain regions of the genome are either over-, under- or not represented in this cosmid clone library. This uneven distribution of clones is attributed to their stability within the *E. coli* ED8767 host (Figure 3). One particular region of the *S. solfataricus* P2 genome represented by contig 1 is well represented with

29.8% of the clones from the *Hind* III library being localized there. This over-representation of the clones found in contig 1 indicate their stability within *E. coli*. On the other hand, contigs 26, 27, 28 and 29 represent regions of the genome that are under-represented. These regions are represented by unique clones. 24.8% (26 out of 105 clones) of the clones within the minimum set are unique. With only 60 to 70% of the genome cloned, 30 to 40% is still missing. The fact that such a high percent of the genome is not cloned and the regions that are cloned show such an uneven distribution would suggest that some regions of *S. solfataricus* P2 DNA are toxic to *E. coli* and have contributed to the uneven representation of its genome.

iii) Extension of the Identified Contigs from the *Hind* III Library

Chromosome walking had limited success in extending contigs identified by shared landmarks because of the presence of insertion sequence and repeat elements in the genome. Figure 4 A shows one of the few successful chromosome walks. Repeat elements appear to be abundant in the genome of *S. solfataricus* P2 (Figure 4 B). Most of the results from chromosome walking were similar to the result shown in Figure 4 B. The differences in signal strengths (dot sizes) seen among the various clones which hybridized to the repeat-containing probe indicate sequence divergence between the repeat in the probe and the repeat(s) in the clones. There was no success in extending any of the identified contigs by landmark analysis using the enzyme, *Eag* I, but some of the overlaps between cosmid clones were confirmed.

The hybridizations performed by Ghislaine Allard using Southerns of the minimal set of cosmid clones extended and linked a few identified contigs. These experiments also confirmed the order of the clones and the overlaps between clones of linked contigs. Some contigs became shorter in size because of the identification of redundant clones within the linked contig.

The chromosome walking experiments performed by Christina Chan using colony blots made from the *Bam* HI cosmid clone library had limited success. The cosmid clones from the *Bam* HI library did not extend any of the identified contigs very far. This is probably due to the toxicity of *S. solfataricus* P2 DNA to *E. coli* resulting in uncloned regions of the *S. solfataricus* P2 genome and this would explain why there is limited extension of linked contigs with the *Bam* HI clones.

C) Construction of an Integrated Contig/Macrorestriction Map

Attempts to construct an integrated contig/macrorestriction map have been hampered by a lack of acceptable hybridization results, lack of restriction enzymes that are able to cut the genome into a number of well resolved fragments and the inability to separate some of the doublet and triplet bands of these digests.

The lack of acceptable hybridization results were caused by the variable amounts of target DNA bound to membrane. It seems that a large amount of the individual *Bss* HII or *Eag* I macrorestriction fragment DNA was lost during the purification process with the GeneClean kit. In some cases, the DNA recovery

was enough to be detected by hybridization and in other cases, there were insufficient amounts. This resulted in a limited amount of data obtained through this method.

Bss HII and *Eag* I were the only restriction enzymes that were able to cut the *S. solfataricus* genome into a number of well resolved fragments, but some of these were doublet and triplet fragments which needed to be resolved further. The discovery with *S. solfataricus* P2 DNA sequence data that some of the *Bss* HII restriction sites were located within an insertion sequence element rendered this enzyme to be useless for macrorestriction mapping.

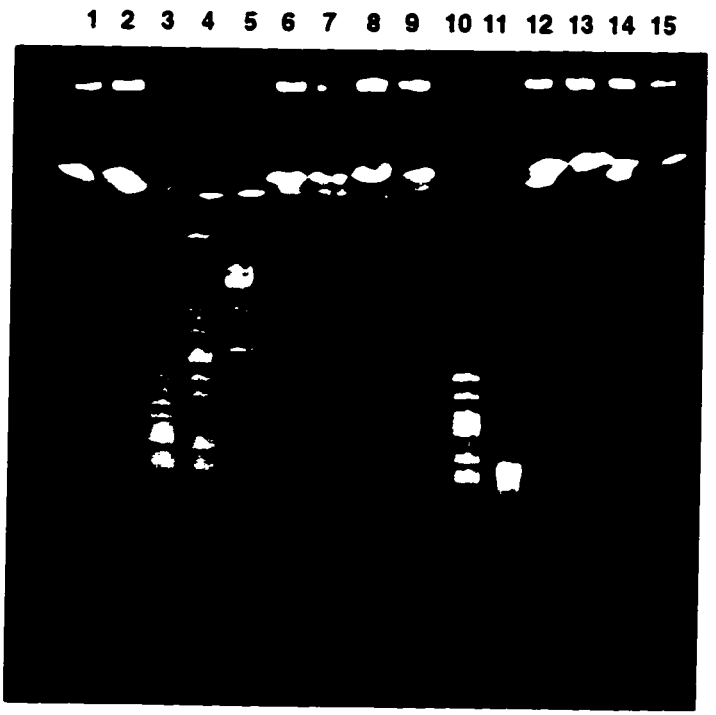
D) DNA Sequenced Cosmid Clones

156 kb of sequence from two contigs have been published (Sensen *et al.*, 1996) (Figure 5). The DNA sequence data identified novel insertion sequence elements and confirmed a previously reported IS element in the genome of *S. solfataricus* P2. Open reading frames (ORFs) that potentially encode proteins of certain biochemical pathways have been identified along with different classes of repeated sequences. Clones from the other 22 contigs have been or are in the process of being sequenced.

Figure 1 **A** **A pulsed-field gel of restriction digested genomic DNA of *Sulfolobus solfataricus* P2 embedded in agarose plugs.** The genomic DNA of *S. solfataricus* P2 was digested with a number of G-C recognition enzymes for the purposes of sizing the genome and to determine candidate enzymes to be used for macrorestriction mapping. The restriction enzymes, *Mlu* I and *Xho* I are not G-C recognition enzymes. DNA size markers, ProMega-Markers Yeast Chromosomal DNA (245 to 2000 kb) and ProMega-Markers Delta-39 (39 to 858 kb) are shown in lanes 1 and 15 and in lanes 2 and 14 respectively. Genomic digests are shown in the following lanes: 3, *Apa* I; 4, *Bss* HII; 5, *Eag* I; 6, *Nar* I; 7, *Ngo* MI; 8, *Not* I; 9, *Sac* II; 10, *Sma* I; 11, *Mlu* I; 12, *Xho* I; 13, *Rsr* II. *Bss* HII and *Eag* I digests were used for sizing the genome of *S. solfataricus* P2 and these enzymes were chosen for macrorestriction mapping because they cut the genome into a number of relatively well resolved bands. Restriction digests and pulsed-field gel electrophoresis conditions are as stated in materials and methods. Genomic digests of *S. solfataricus* P2 DNA with these enzymes were performed many times to ensure that reproducible results were obtained.

B ***Eag* I macrorestriction fragments with their respective designations according to size.**
Panel B is lane 5 taken from Panel A and enlarged. The *Eag* I macrorestriction fragments were sized against the DNA size marker, ProMega-Markers Delta-39. The size and alpha-numeric designation of each *Eag* I macrorestriction fragment are shown. The genome size of *S. solfataricus* P2 was estimated by totalling up the 17 *Eag* I macrorestriction fragments. The 16 *Bss* HII macrorestriction fragments were used to size the genome of *S. solfataricus* P2 by the same method (data not shown).

A


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

B

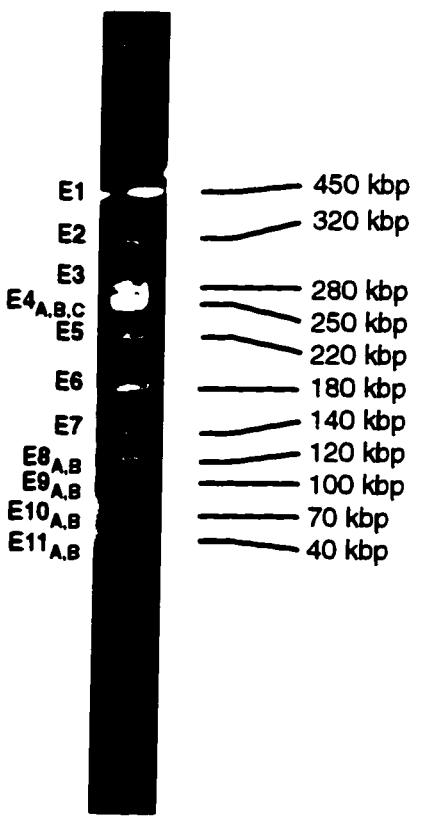
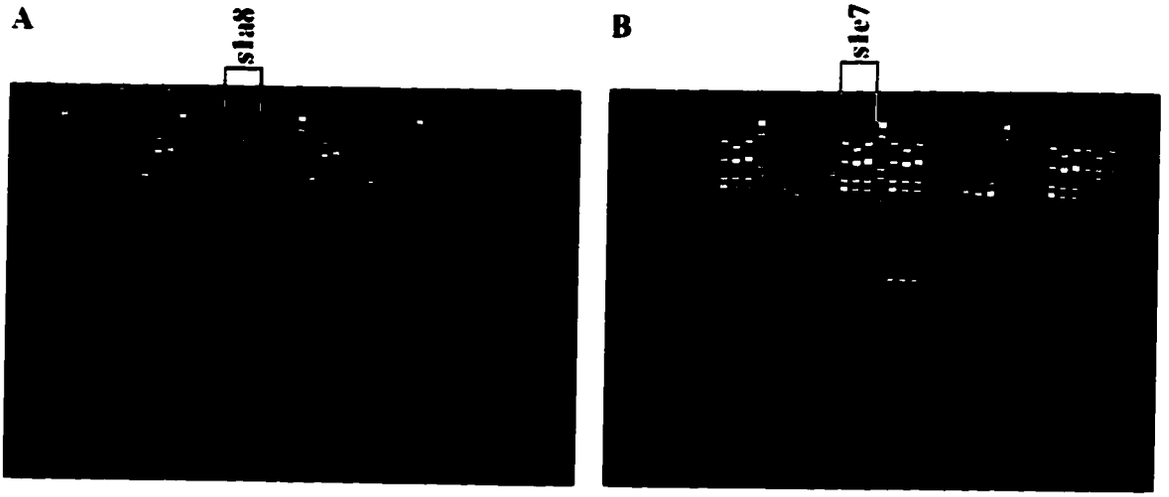

E1
E2
E3
E4A,B,C
E5
E6
E7
E8A,B
E9A,B
E10A,B
E11A,B
450 kbp
320 kbp
280 kbp
250 kbp
220 kbp
180 kbp
140 kbp
120 kbp
100 kbp
70 kbp
40 kbp

Figure 2 **A, B** **Samples of agarose gels of cosmid clones restriction digested with *Hind* III (H), *Hind* III/*Mlu* I (H/M) and *Hind* III/*Xho* I (H/X) for landmark analysis.**

Each one of the 1296 cosmid clones from the *Hind* III library was digested with *Hind* III, *Hind* III/*Mlu* I and *Hind* III/*Xho* I. Twelve restriction digested cosmid clones were resolved in an agarose gel along with the DNA size marker, Mappers' Marker Mix (MMM). Five lanes of the DNA size marker were loaded in the gel and each marker lane flanks the digests of three cosmid clones. The three lanes of restriction digests of cosmids s1a8 (should be designated sh01a08) and s1e7 (should be designated sh01e07) are marked in panels A and B respectively. Restriction digests and agarose gel electrophoresis conditions are as stated in materials and methods. At least 108 agarose gels were used for resolving the restriction digests of the 1296 clones of the *Hind* III library. An album of photos of these digests is kept so that recent restriction digest patterns of clones can be checked for clone identification and deletion of insert band(s).

C, D **The respective restriction digests of the overlapping cosmid clones sh01a08 and sh01e07.**

Panels C and D are taken from Panels A and B respectively and enlarged. Panels C and D show the Hind III (H), Hind III/Mlu I (H/M) and Hind III/Xho I (H/X) digests of clones sh01a08 and sh01e07 respectively. The photographs highlight the region of overlap between the two cosmid clones. Black arrowheads point to the *Hind* III fragment containing the "landmark"; white arrowheads point to the *Hind* III fragment containing the "landmark" which is cut by the restriction enzyme, *Mlu* I and the black arrows point to the Tropist 3 vector band. The DNA size marker, MMM is shown alongside the respective restriction digests of the overlapping cosmid clones sh01a08 and sh01e07.


A
s1a8
B
s1c7

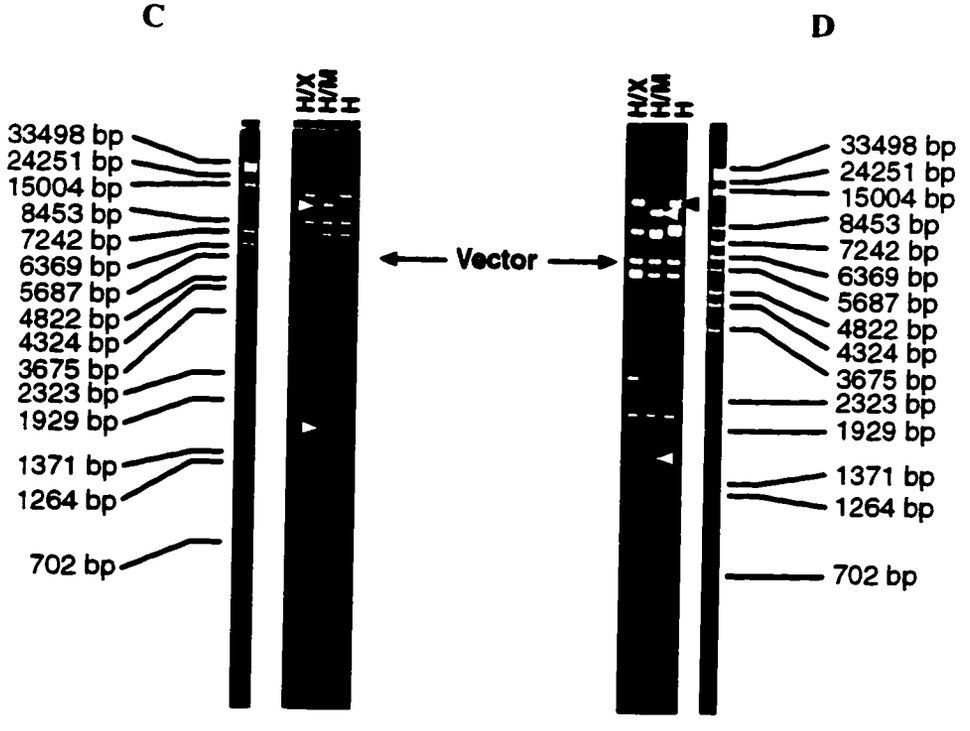

C
D
H/X
H/M
H
33498 bp
24251 bp
15004 bp
8453 bp
7242 bp
6369 bp
5687 bp
4822 bp
4324 bp
3675 bp
2323 bp
1929 bp
1371 bp
1264 bp
702 bp
Vector

Figure 3

Distribution of cosmid clones from the *Hind* III library among the linked contigs.

1201 of the 1296 clones of the *Hind* III cosmid clone library were grouped according to their *Hind* III, *Hind* III/*Mlu* I and *Hind* III/*Xho* I digestion patterns. Clones that produced the same digestion patterns were grouped together. An individual clone from these groupings were then used for landmark analysis to construct contigs. 59 out of the 1296 clones yielded no cosmid DNA probably a result of failed DNA extractions. 36 out of the 1296 clones were identified as deleted or deleting clones by the characteristic bright vector band or the bright vector band and faint DNA insert bands respectively. 1201 clones remain in the library after removing the 59 non-DNA yielding and the 36 deleted or deleting clones. The total number of clones found in each contig was calculated by adding up all the clones of the groups of clones that link into a contig. The graph shows the number of clones found in each contig (in parentheses) and the percent cloned per kbp. The percent cloned per kbp was calculated by dividing the proportion of clones found within each contig by the size of the linked contig. The proportion of clones found within each contig was determined by dividing the number of clones found in each contig by the total number of analyzed clones (1201) in the *Hind* III library. The histogram highlights the uneven distribution of cosmid clones among the linked contigs with a large proportion of clones situated in contig 1 and unique clones in contig 26,27,28 and 29. It appears that certain regions of the *S. solfataricus* P2 genome are stably maintained in *E. coli* and other regions are not.

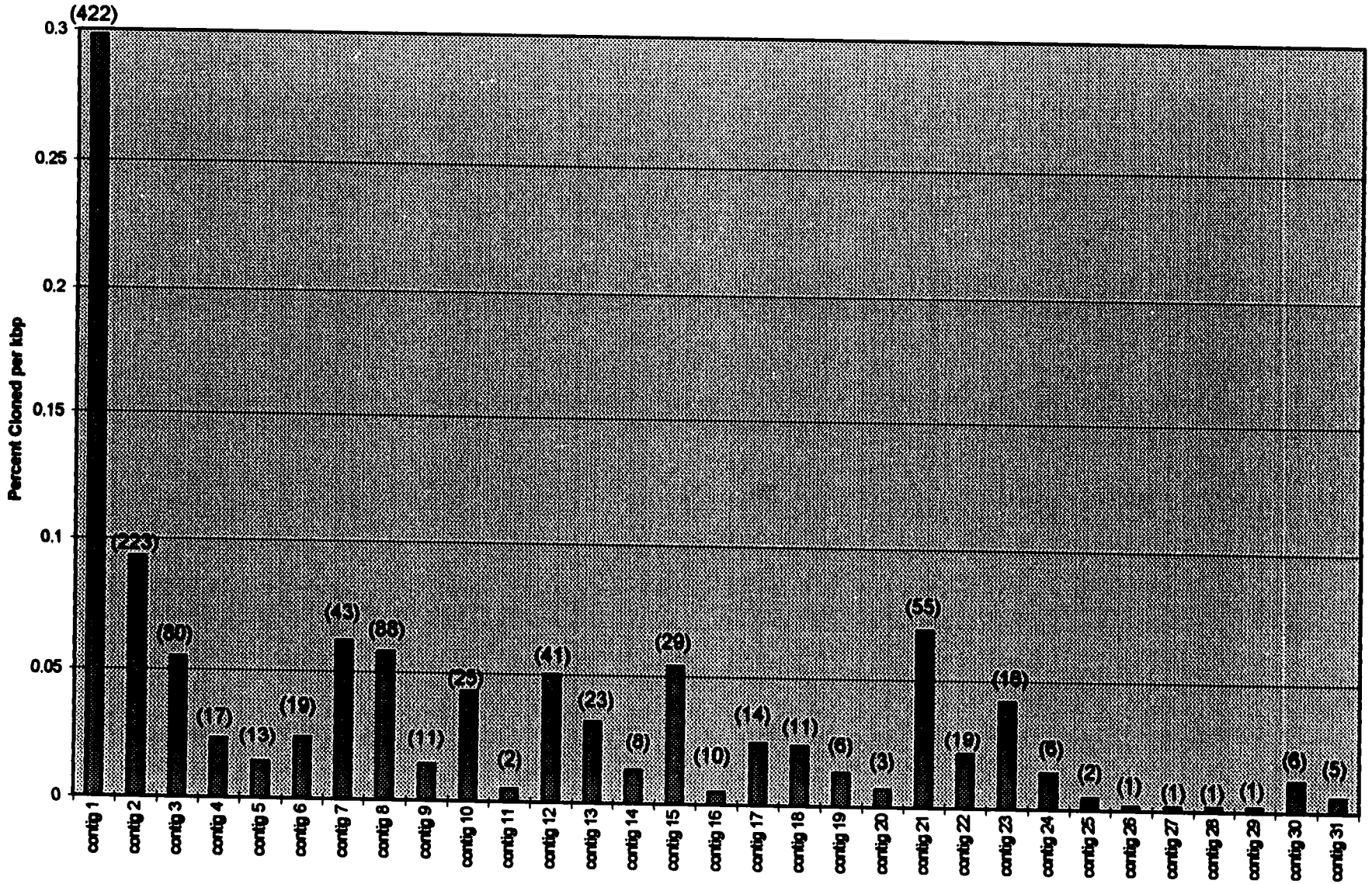

Distribution of Clones Among the Linked Contigs
Percent Cloned per kbp
0.3
0.25
0.2
0.15
0.1
0.05
0
(422)
(223)
(89)
(17)
(13)
(19)
(43)
(68)
(11)
(25)
(2)
(41)
(23)
(8)
(29)
(10)
(14)
(11)
(6)
(3)
(55)
(19)
(18)
(6)
(2)
(1)
(1)
(1)
(1)
(6)
(5)
contig 1
contig 2
contig 3
contig 4
contig 5
contig 6
contig 7
contig 8
contig 9
contig 10
contig 11
contig 12
contig 13
contig 14
contig 15
contig 16
contig 17
contig 18
contig 19
contig 20
contig 21
contig 22
contig 23
contig 24
contig 25
contig 26
contig 27
contig 28
contig 29
contig 30
contig 31

Figure 4 **A** **Dot blot hybridization for extending contig 7 using *Hind* III fragments 4 and 6 from cosmid clone, sh04b07 as probe.** *Hind* III fragments 4 and 6 representing the end of clone, sh04b07 were hybridized against a dot blot of the minimal set of cosmid clones. The dot blot contains the minimal set of clones minus the unusual clone, sh02h03U and five variant clones, sh20f08U, sh14b10, sh19a12, sh01g03 and sh19g12U. There are 99 clones immobilized on the membrane. The 9 polymorphic clones are on this membrane The results show a link between clones sh04b07 and sh03d02 (small dot on left). This is one of the few successful chromosome walks. The probe also identified itself (large dot on right) and a variant clone, sh16c05 (smaller dot on right). The hybridization conditions are as stated in materials and methods.

B **Dot blot hybridization for extending contig 1 using all the *Hind* III fragments from cosmid clone sh13d03 as probe.** All the Hind III fragments from clone, sh13d03 were hybridized against the same dot blot used in panel A. These results show the extent of repeated sequences within the genome of *S. solfataricus* P2. The repeat-containing probe identified itself (large dot in middle of blot) and 42 other clones. The differences in signal strength (dot sizes) seen among the various clones which hybridized to the repeat-containing probe indicate sequence divergence between the repeat in the probe and the repeat(s) in the clones. *Hind* III fragments from the ends of the 31 linked contigs were used as probes in attempts to extend the contigs. In the case of cosmid clone sh13d03, all the *Hind* III fragments were used as probes because initially it had not been linked to any of the contigs. The hybridization conditions are as stated in materials and methods.

A

B

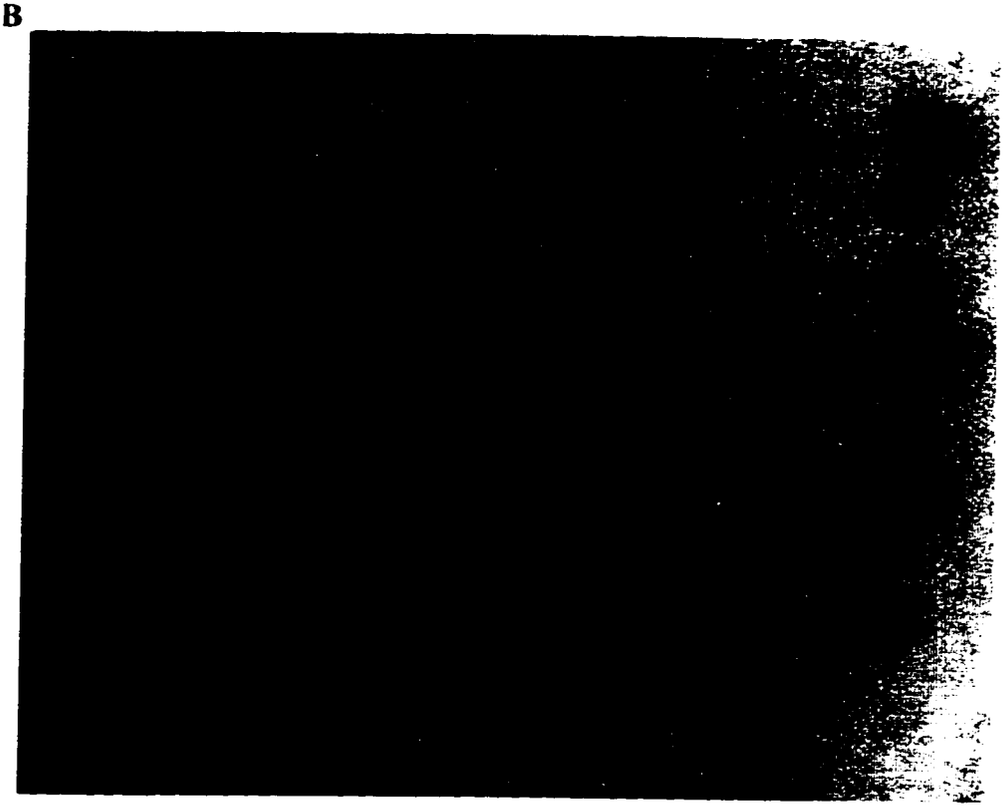
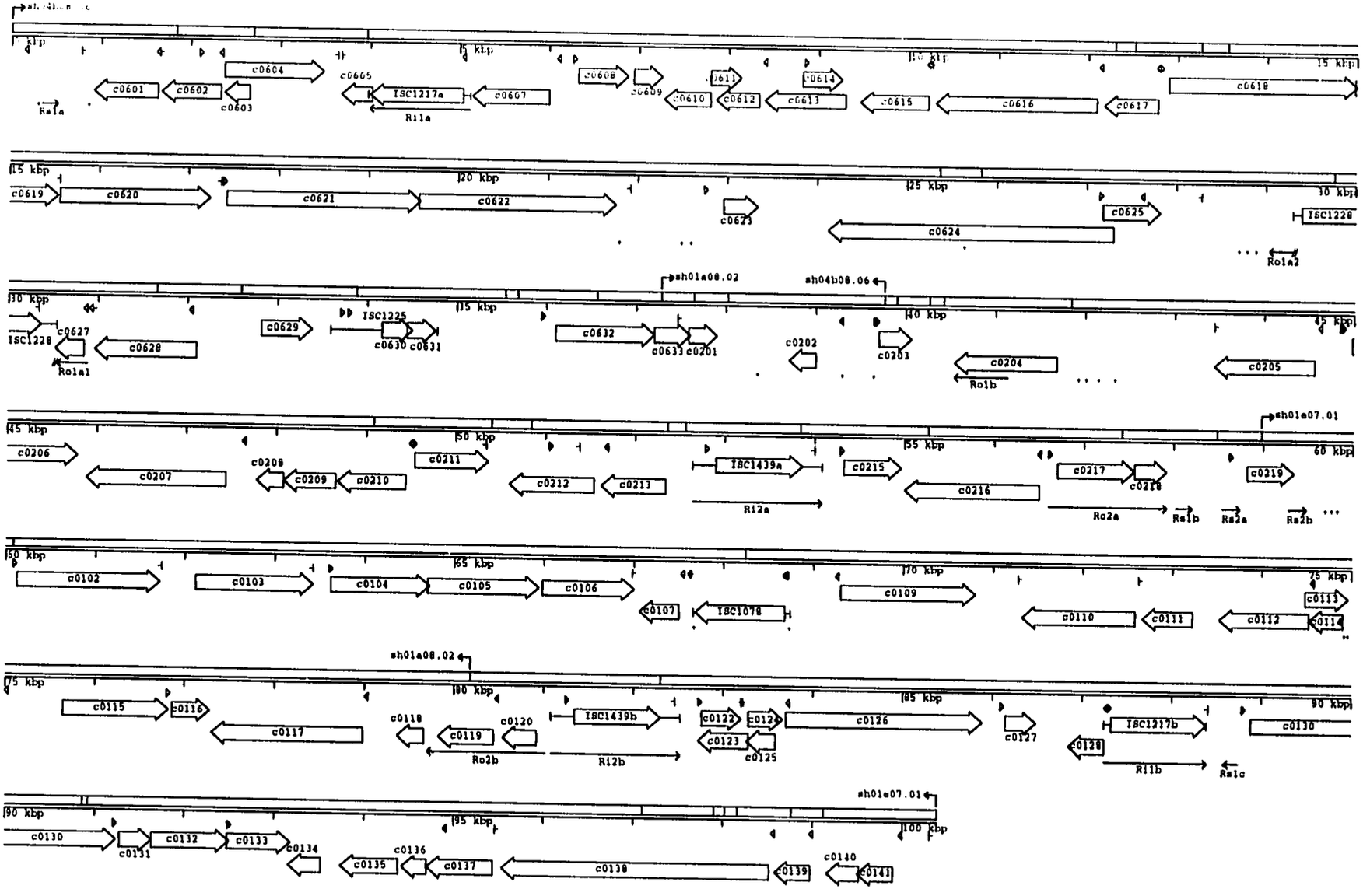


Figure 5 **A** **Sequence features of the 100 389 bp derived from cosmid clones sh04b08.06, sh01a08.02 and sh01e07.01 of contig 1.**

Open reading frames (ORFs) >300 bp are shown as open arrows and are numbered from left to right from each cosmid clone. Selected ORFs <300 bp (see text of Sensen *et al.*, 1996) and RNA genes are also indicated. Potential promoters and terminators (see text of Sensen *et al.*, 1996) are shown above the ORFs as small triangles and sideways letters T, respectively. Insertion-sequence boundaries are indicated by bars extending their respective ORFs. Repeated sequences of three types (Rs: short; Ro: ORF-containing; and Ri: IS) are mapped as thin arrows; duplicated or triplicated oligonucleotides (18-24 bp) are shown as anonymous dots. *Hin*dIII sites are displayed above the scale bar. GenBank accession numbers of the 100- and 56 kb contigs are X00000 and X00000, respectively. Both unmodified figures and the slightly modified figure legend were taken from Sensen *et al.*, 1996.

B **Sequence features of the 56 105 bp derived from cosmid clones sh03d02.04 and sh04b07.05 of contig 7.**

Refer to legend in A.


c0604
c0605
c0601
c0602
c0603
ISC1217a
c0607
Ri1a
Rs1a
c0608
c0609
c0610
c0611
c0612
c0613
c0614
c0615
c0616
c0617
c0618
c0619
c0620
c0621
c0622
c0623
c0624
c0625
ISC1220
Ro1a2
sh01a08.02
sh04b08.06
c0627
ISC1220
c0628
Ro1a1
c0629
ISC1225
c0630
c0631
c0632
c0633
c0201
c0202
c0203
c0204
Ro1b
c0205
sh01a07.01
c0206
c0207
c0208
c0209
c0210
c0211
c0212
c0213
ISC1439a
Ri2a
c0215
c0216
c0217
c0218
c0219
Ro2a
Rs1b
Rs2a
Rs2b
c0102
c0103
c0104
c0105
c0106
c0107
ISC1078
c0109
c0110
c0111
c0112
c0113
c0114
sh01a08.02
c0115
c0116
c0117
c0118
c0119
c0120
Ro2b
ISC1439b
Ri2b
c0122
c0123
c0125
c0126
c0127
c0128
ISC1217b
Ri1b
Rs1c
c0130
sh01a07.01
c0130
c0131
c0132
c0133
c0134
c0135
c0136
c0137
c0138
c0139
c0140
c0141

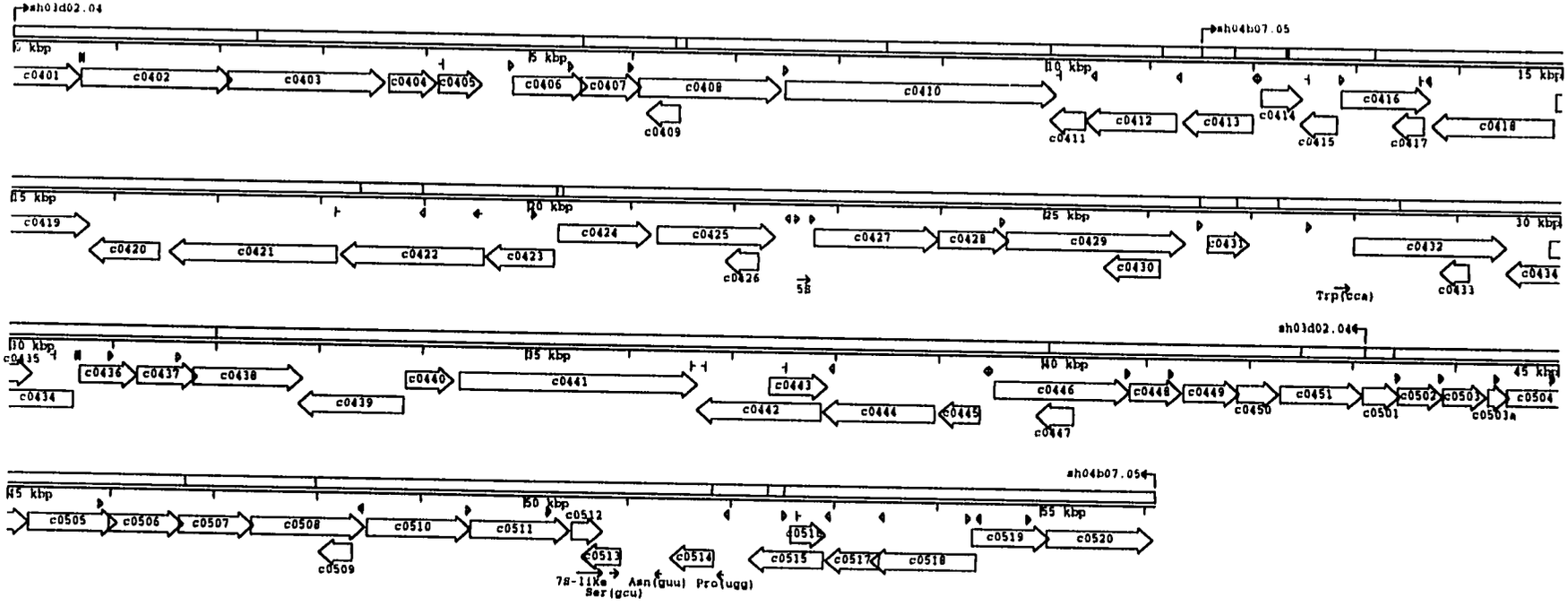

sh01d02.04
sh04b07.05
c0401
c0402
c0403
c0404
c0405
c0406
c0407
c0408
c0409
c0410
c0411
c0412
c0413
c0414
c0415
c0416
c0417
c0418
c0419
c0420
c0421
c0422
c0423
c0424
c0425
c0426
c0427
c0428
c0429
c0430
c0431
c0432
c0433
c0434
Trp(tcca)
sh03d02.04
c0435
c0436
c0437
c0438
c0439
c0440
c0441
c0442
c0443
c0444
c0445
c0446
c0447
c0448
c0449
c0450
c0451
c0501
c0502
c0503
c0503a
c0504
sh04b07.05
c0505
c0506
c0507
c0508
c0509
c0510
c0511
c0512
c0513
c0514
c0515
c0516
c0517
c0518
c0519
c0520
Ser(gcu)
Asn(guu)
Pro(ugg)

Table 2 **Linked contigs of minimal set cosmids.**

105 minimal set *Hind* III cosmid clones were linked into 31 contigs by landmark analysis and chromosome walking. Clones linked into contigs are shown with dashes between the linked clones. The order of the clones within the contigs are shown. The size of the linked contigs are indicated and were estimated by adding up all the *Hind* III fragments unique to a particular clone and then subtracting all common *Hind* III fragments shared between the two overlapping clones. This method may underestimate the actual size of each linked contig. One unusual, six variant and nine polymorphic clones were identified and are listed under their respective contig. The placement of certain clones within a contig that have to be determined are indicated as "to be placed". Clones with brackets around them are those whose order and position that still have to be determined. DNA sequencing clones are designated with a decimal point followed by a two digit number. Unique clones are identified with the capital letter U.

Contig 1 (118 kbp)

sh13d03.31-sh04b08.06-sh01a08.02-sh01e07.01

polymorphic clones: sh15e10U, sh02d04, sh13d04, sh15c01U, sh20b05U, sh22a12U, sh22g07, sh03f07.03, sh02f03

Contig 2 (200 kbp)

sh02g08.20-sh01d05.10-sh03f09.16-sh04c08.09-sh02a07.08-sh22g09.21-sh18e04.22-sh04g11.23-sh22e04.19

unusual clone: sh02h03U

Contig 3 (121 kbp)

sh01e04.42-sh19c10U-sh20c07.48-sh01d08.97-sh02d06.49

sh13b06 may extend sh01e04.42

Contig 4 (61 kbp)

sh13a06.29-sh19h12.30

to be placed: sh13b01

Contig 5 (75 kbp)

sh13b04.15-sh03g11.50

Contig 6 (65 kbp)

sh14b02.36-sh03g08.37

Contig 7 (56 kbp)

sh03d02.04-sh04b07.05

variant clone: sh16c05

Contig 8 (127 kbp)

sh22f04U-(sh01e09.99,sh19a04U)-sh14c08-(sh04h09, sh22e02)

to be placed: sh18b03, sh02b07

Contig 9 (64 kbp)

sh04a08.27-sh02e12.28

Contig 10 (48 kbp)

sh13g06.45-sh02a09.46

variant clone: sh20f08U

Contig 11 (34 kbp)

sh21b06U.44

to be placed: sh22b07U

Contig 12 (68 kbp)

sh03a02.13-sh03a06.26

to be placed: (sh01a07U, sh19e10, sh13a11U), sh02g02, sh20a10

variant clone: sh14b10

Contig 13 (60 kbp)

sh01c11.34-sh15d05.35

Contig 14 (52 kbp)

sh18f04.33

to be placed: sh04a10, sh14a11U

Contig 15 (45 kbp)

sh03b07.17

to be placed: sh21e08U, sh22c10U, sh01f03

variant clones: sh19a12, sh01g03, sh19g12U

Contig 16 (156 kbp)

sh04b05.52-sh03e02.14-sh03g07.53

Contig 17 (49 kbp)

sh01e03.38-sh13e08.51

to be placed: **sh18f05U**

Contig 18 (39 kbp)

sh13a04.25

to be placed: **sh22b10, sh03a08**

Contig 19 (38 kbp)

sh19b01.32

Contig 20 (35 kbp)

sh18h07.43

Contig 21 (66 kbp)

sh02c09.98-sh13a02.24

to be placed: **sh04a01, sh01f12, sh22d09, sh02g11U**

Contig 22 (76 kbp)

sh03f01.39-sh02d09.40

to be placed: **sh13e06U**

Contig 23 (32 kbp)

sh20b01.41

to be placed: **sh15h12U**

Contig 24 (36 kbp)

sh13e12.47

Contig 25 (36 kbp)

sh03a09

Contig 26 (35 kbp)

sh19b10U

Contig 27 (33 kbp)

sh19g02U

Contig 28 (35 kbp)

sh17g10U

Contig 29 (36 kbp)

sh17h09U

Contig 30 (40 kbp)

sh17f08

Contig 31 (72 kbp)

sh03h10-sh20c02U

Discussion

A) The Genome of *Sulfolobus solfataricus* P2

The genome of *S. solfataricus* P2 has been estimated to be 3.0 +/- 0.1 Mbp by pulsed-field gel electrophoresis. It is similar in size with another species of *Sulfolobus*, *S. acidocaldarius* strain 7 which is 2.76 Mbp (Kondo *et al.*, 1993). The difference in genome size between these two *Sulfolobus* species can be attributed to the presence of repeated genetic elements in the genome of *S. solfataricus* P2. When compared with other Archaeal genomes such as *Methanobacterium thermoautotrophicum* Marburg (1.6 Mbp, Stettler and Leisinger, 1992), *Methanococcus voltae* PS (1.9 Mbp, Sitzmann and Klein, 1991), *Methanococcus jannaschii* (1.66 Mbp, Bult *et al.*, 1996), *Thermococcus celer* Vu13 (1.89 Mbp, Noll, 1989), *Halobacterium* sp. GRB (2.47 Mbp, St. Jean *et al.*, 1994), and *Halobacterium halobium* strain NRC-1 (2.4 Mbp, Bobovnikova *et al.*, 1994), the genome of *S. solfataricus* P2 is quite large; but it is small relative to *Haloferax mediterranei* ATCC 33500 (3.84 Mbp, Lopez-Garcia *et al.*, 1992) and *Haloferax volcanii* DS2 (4.14 Mbp, Charlebois *et al.*, 1991).

This genome is not partitioned into a chromosome and a few plasmids like that of most halobacteria. Its genome consists of a chromosome which is assumed to be circular resembling that of *S. acidocaldarius* (Yamagishi and Oshima, 1990). Although plasmids have not been observed in the genome of *S. solfataricus* P2, a multicopy plasmid has been reported in *Sulfolobus* sp. strain NOB8H2 (Schleper *et al.*, 1995).

B) Problems Encountered While Mapping

i) Host System Bias

It appears that some of the DNA of *S. solfataricus* P2 is incompatible with the host, *E. coli* ED8767; the DNA may be toxic. This observation is supported by the presence of 36 deleted or deleting cosmid clones in the *Hind* III library, the varying colony sizes and mostly by the severe underrepresentation of many parts of the genome despite a large library. The A-T rich DNA of *S. solfataricus* P2 may have promoter-like functions which may cause extraneous transcription with energetic costs to the cell. Apparently, deletion of DNA is characteristic of cosmid clones even when they are stably maintained in a *recA* strain such as DH1 (Tabata *et al.*, 1989). This may suggest that *recA*-independent recombination is involved in deletion as suggested by Symington *et al.* (1985) and Ishiura *et al.* (1990). It may be the combination of the A-T rich DNA and *recA*-independent recombination that causes the deletion of DNA.

Uncloned regions of a genome are common obstacles in the construction of gene encyclopedias and appear as gaps. It was for this reason that Kohara *et al.* decided to use a lambda phage vector instead of a cosmid vector in constructing a high-resolution physical map of *Escherichia coli* K-12 W3110 (Kohara *et al.*, 1987). Kohara *et al.* suggested that cloned segments in cosmid frequently accumulate deletions that may result from selection for a shorter size for faster replication and/or metabolic imbalances caused by the increased dosage of a particular gene(s) on the cosmid that affects the growth of host cells.

The situation of uncloned regions of a genome were encountered in the generation of cosmid contig maps for *E. coli* strain BHB2600 (Birkenbihl and Vielmetter, 1989), *Helicobacter pylori* strain NCTC11638 (Bukanov and Berg, 1994) and *Rhodobacter capsulatus* (Fonstein *et al.*, 1995). These unknown genes in these uncloned regions are probably absent from these clone libraries because of the toxicity of their products in the host (Bukanov and Berg, 1994). These unclonable regions can be studied by PCR amplification and subsequent sequencing (Fonstein *et al.*, 1995).

ii) Clonability

Approximately 60 to 70% of the genome of *S. solfataricus* P2 is cloned into the cosmid vector, Tropist 3. Tropist 3 was chosen for its stability, yield of DNA from cosmid preparations and cloning capacity of 35 to 50 kb (De Smet *et al.*, 1993). It is a derivative of the loric cosmid which is capable of maintaining a higher and a constant copy number with increasing size than that of ColE1 replicon-based cosmids which appear to have an inverse relationship between size and copy number (Little and Cross, 1985). It would appear that certain regions of this genome can be maintained in this vector and others are not. There is an uneven distribution of the cosmid clones among the linked contigs in the *Hind* III library: some regions are over-represented and others are under-represented as shown in Figure 3. Most of the cosmid clones seem to cluster in contig 1; this region of the genome is highly represented with 35% of the clones localized there. Contigs 26, 27, 28 and 29 represent regions of the genome that

are under-represented by cosmid clones with 0.08% share of the clones. Even when it appeared that the clones are stably maintained, some began to delete. 36 out of 1296 clones are deleted or in the process of deleting. 26 out of the 105 clones from the minimum set are unique; they are the only representatives of that region of the genome. The number of unique and deleting or deleted clones reflects the incompatibility of *Sulfolobus* DNA with *E. coli.*

iii) Lack of Restriction Enzymes to Digest the Genomic DNA

A number of restriction enzymes, *Apa* I, *Bss* HII, *Eag* I, *Mlu* I, *Nar* I, *Ngo* MI, *Not* I, *Sac* II, Rsr II, *Sma* I and *Xho* I were tested for their ability to digest the genomic DNA of *S. solfataricus* P2. Some of these enzymes, *Not* I, *Bss* HII, *Eag* I and *Rsr* II were used successfully to generate a low resolution physical map of the chromosome of *S. acidocaldarius* 7 (Kondo *et al.*, 1993). All the enzymes were able to cut the genomic DNA of *S. solfataricus* with the exception of *Not* I, but they cut too frequently, resulting in a smear of bands. The enzymes, *Bss* HII and *Eag* I, were the only ones that cut the genomic DNA into a manageable number of discernible bands, 10 and 11 respectively. Some of these bands contained fragments of similar size that appeared on the pulsed-field gel as doublets or triplets. It was quite surprising that *Not* I and *Rsr* II were not useful considering they had been used successfully in constructing the physical map of *S. acidocaldarius* 7 which is similar in genome size (2.7 Mbp) and G-C content (40%) with *S. solfataricus* P2 (Kondo *et al.*, 1993).

The cloning of the genomic DNA of *S. solfataricus* being a problem, a concerted effort was undertaken to generate an integrated contig/macrorestriction map using the linked contigs and the *Bss* HII and *Eag* I macrorestriction fragments. The strategy involved hybridizing and thereby localizing the linked contigs onto the macrorestriction fragments. The macrorestriction fragments would be linked into a circle by junction clones, clones that contained *Bss* HII and/or *Eag* I sites. Linking up of macrorestriction fragments would also be facilitated by single and double digests with *Bss* HII and *Eag* I.

This effort to generate an integrated contig/macrorestriction map has encountered a few problems that would hinder the construction of this physical map. The hybridization of junction clones to *Bss* HII and *Eag* I macrorestriction fragments has not been fruitful. It appears that there is not enough target DNA bound to the membranes so that results are usually undetectable except in a few cases. Attempts have been made to concentrate the amount of target DNA immobilized onto membrane by pooling of samples but the results are still undetectable. Compounding this problem is that some of these macrorestriction fragments still have to be separated properly. It was also discovered with DNA sequence data that some of the *Bss* HII sites were located within an insertion sequence (IS) element rendering this enzyme to be useless for map construction because of the known mobility of these IS elements in *Sulfolobus*.

Currently work is underway to rectify some of these problems: the use of concentrated cultures of *S. solfataricus* to help increase the amount of genomic DNA in each agarose gel plug and the doublet and triplet bands of *Eag* I are being resolved further to give better band separations. Previous genomic restriction digests with *Not* I have shown the inability of this enzyme to cut the genomic DNA of *S. solfataricus* P2, but recent genomic restriction digests with *Not* I performed by Ghislaine Allard suggest it to be a candidate for macrorestriction mapping. It is possible that the tube of *Not* I, which was used in the original genomic digests, contained inactivated enzyme. A new tube of *Not* I had been used by Ghislaine Allard in the recent genomic digests.

iv) Number of Bands to Analyze for the Landmark Strategy

It is ideal to have a manageable number of cloning enzyme fragments per cosmid clone for landmarking. Too many fragments makes it very difficult to find landmarks and too few bands would result in a cloning bias. An appropriate number of fragments would be ten given the fact that the DNA insert size of a cosmid vector is 40 to 50 kb.

The partial digest conditions of 1/4 and 1/8 units resulted in the cloning of large DNA fragments in the first 384 cosmid clones of the *Hind* III library with the series designation sh01 to sh04. The cloning of smaller size *Hind* III fragments would be excluded from this library resulting in uncloned regions of the genome.

In order to overcome this DNA insert size bias, partial restriction digest conditions of 1/8 and 1/16 units were chosen for the clones in the sh05 to sh26

series. The clones from series sh07 to sh12 showed signs of deletion emphasizing the need to process cosmids promptly; these were stored for a period of one to three weeks prior to DNA extraction. Clones from the series sh13 to sh24 were stably maintained, but some of them began to delete. Under these partial digestion conditions, the clones possessed smaller and many more *Hind* III fragments. Each clone contained about 15 *Hind* III fragments on average which made it difficult for landmark analysis; there were many comigrating bands to analyze.

v) Repeat Elements

Hybridization results identified repeat elements within this genome and DNA sequence data confirmed these findings. There are 200 base-pair repeat sequences and insertion sequence (IS) elements in this archaeal genome (Sensen *et al.*, 1996). Database searches identified novel IS elements and the previously reported IS element, ISC1217 in *Sulfolobus* (Schleper *et al.*, 1994).

The identification of possibly active repeat elements in this genome suggest it to be unstable. These repeat elements have a strong influence on the structure and stability of the genome. They may be sites of intrachromosomal homologous recombination leading to inversions and deletions (Krawiec and Riley, 1990; Roth *et al.*, 1996).

Of the Archaea, halophiles are known to have dynamic genomes; genomic rearrangements and phenotypic variability is caused by the movement of insertion sequence elements (IS*H*, *H* for Halophiles). IS*H* elements are

abundant and highly active in *Halobacterium halobium* (now *H. salinarium*, Staley *et al.*, 1989) (Sapienza and Doolittle, 1982; Sapienza *et al.*, 1982). Movement of IS*H* elements by transposition causes frequent disruptions of gas vacuole protein (*gvp*) and bacterioopsin (*bop*) genes and results in easily detectable phenotypic variants (DasSarma *et al.*, 1983; Pfeifer *et al.*, 1989). High mutation rates of 10^{-2} have been observed in the plasmid encoded gas vacuole protein gene (*p-vac*) of *H. salinarium* (Pfeifer *et al.*, 1989).

Many of the IS*H* elements have been cloned and sequenced and range in size from 521 bp (ISH2, DasSarma *et al.*, 1983; Pfeifer and Blaseio, 1989) to 3kb (ISH24, Pfeifer *et al.*, 1984) and are generally A-T richer than the rest of the genome (Charlebois and Doolittle, 1989). All but a few IS*H* elements are flanked by short target site duplications and bounded by short terminal inverted repeats; many also contain long open reading frames (ORF). These features are also characteristic of bacterial IS elements. IS*H* elements are unevenly distributed among the genome with some found on the chromosome, small and large plasmids.

Within the completely sequenced genome of *Methanococcus jannaschii*, three families of repeated genetic elements have been identified (Bult *et al.*, 1996). One is a family of putative insertion sequence elements with the designation, IS*AMJ1*. Another is a family of multicopy repetitive elements and the third family consists of two open reading frames (ORFs) that are 23%

identical at the amino acid sequence level to the carboxyl terminus of a transposase from *Lactococcus lactis*.

Before this report by Bult *et al.* (1996), only two IS elements were documented in the methanogenic Archaea. There was one confirmed insertion element, ISM1, from *Methanobrevibacter smithii* (Hamilton and Reeve, 1985) and a putative insertion element, FR-1 had been identified in plasmid pFV1 from *Methanobacterium thermoformicum* THF and chromosomal DNA from *M. thermoformicum* THF, Z-245, FTF, FF1, FF3, CSM3, HN4, and *M. thermoautotrophicum* ΔH (Nolling *et al.*, 1993). These two elements possess the characteristic open reading frame and short target site duplications of insertion elements.

Schleper *et al.* reported the first IS element, ISC1217 (C for Crenarchaeota) within the genome of *Sulfolobus solfataricus* P1 (Schleper *et al.*, 1994). It was found in the *lacS* gene of *S. solfataricus* with features that are typical of insertion sequence elements such as the possession of terminal inverted repeats, flanked by a direct repeat of 6 bp and an open reading frame that might code for a transposase. There are approximately 8 copies of ISC1217 in the genome of *S. solfataricus* and their transposition activities have been associated with two spontaneous mutants with a frequency of 10^{-4} per plated cell from independent cultures. This insertion mutation rate is comparable to those that have been described in *H. salinarium*.

It would appear that the Archaeal genomes just like those of bacteria are not immune to the infectivity of IS elements. There have been documented cases of insertion elements within the genomes of halophiles and methanogens and the current data support and confirm the presence of IS elements in the genome of at least one thermophile, *S. solfataricus* P2.

vi) Polymorphic Clones

A branch point was identified in contig 1 while landmarking. One end of contig 1 represented by cosmid clone, sh01e07 was found to overlap with a number of clones, but these clones did not extend each other. These clones were later confirmed to be polymorphic in restriction fragment lengths known as restriction fragment length polymorphisms (RFLPs). The polymorphisms of the clones are caused by the presence and/or absence of specific insertion elements. The polymorphic nature of these clones was studied by the honours student, Todd Monkman. He employed a strategy that involved extensive restriction mapping in attempts to isolate the smallest fragment and then sequencing in towards the site of the polymorphism with primers synthesized to flank the region. A number of novel IS elements in *Sulfolobus* which are implicated in the polymorphisms have been identified by the sequence data.

C) Solutions to Mapping this Genome

i) Cloning Systems

This mapping endeavour has been hampered by the presence of repeat and insertion sequence elements and the problem with cloning A-T rich DNA. Apparently, cosmid cloning is not the appropriate system to use for generating a physical map of *S. solfataricus* P2 with only 60-70% of genome cloned. This type of cloning efficiency was also observed in the construction of the *E. coli* K-12 W3110 chromosome using cosmid vectors (Tabata *et al.*, 1989). Despite the inability to clone the DNA of *S. solfataricus* P2, this cloning system has been used successfully in covering greater than 90% of the genome of *E. coli* strain BHB2600 (Birkenbihl and Vielmetter, 1989), *Mycoplasma pneumoniae* (Wenzel and Herrmann, 1988; Wenzel and Hermann, 1989), *Mycoplasma genitalium* (Lucier *et al.*, 1994; Peterson *et al.*, 1995), *Rhodobacter capsulatus* SB1003 (Fonstein and Haselkorn, 1993), *Helicobacter pylori* strain NCTC11638 (Bukanov and Berg, 1994), *H. volcanii* DS2 (Charlebois *et al.*, 1991) and *Halobacterium* sp. GRB (St. Jean *et al.*, 1994).

There exist other cloning systems such as yeast artificial chromosomes (YACS), bacterial artificial chromosomes (BACs), plasmids and bacteriophages (P1and λ). Each cloning system offers a different level of resolution for the constructed physical map because each vector has a DNA insert size limit: YACs (>500 kb, Burke *et al.*, 1987), BACs (>300 kb, Shizuya *et al.*, 1992), plasmids (2 kb), λ phage (20 kb) and P1 (100 kb, Sternberg, 1990). The DNA

insert size will determine the number of clones to screen in order to have whole coverage of the genome.

In constructing a physical map of any genome, three parameters must be addressed carefully: a) the number of clones needed to screen, b) the copy number of the vector and c) the desired resolution of the map. The cosmid system seems to be the appropriate choice with a low copy number and a minimal number of clones to screen in order to generate a high resolution map because of its DNA insert capacity of 40-50 kb.

Different cloning systems are being developed by researchers involved in the Human Genome Project to overcome any of the deficiencies associated with YACs such as the co-cloning of noncontiguous DNA fragments in some clones. The BAC system is based on *E. coli* and its single-copy plasmid F factor which is capable of maintaining human genomic DNA fragments of >300 kb (Shizuya *et al.*, 1992). A low copy number cosmid vector based on the *E. coli* F factor replicon was developed for the Fosmid system (Kim *et al.*, 1992). The Fosmid system may be appropriate for cloning the rest of the genome of *S. solfataricus* P2 with its low copy number and ability to stably maintain DNA inserts of 40-50 kb.

ii) Host Systems

A proper host system is required to complement the cloning system employed because host restriction activities can effectively eliminate certain sequences from the library and the goal is to have all sequences represented

with equal probability. Most researchers involved in "Genome Projects" choose host systems that are deficient in their ability to cause recombination between homologous sequences such as repeated sequences. In the mapping project of *H. volcanii* DS2, two strains of *E. coli* that were recombination deficient were tested with ED8767 being the superior host as the DH5α clones had an obvious bias over-representing certain genomic regions (Charlebois *et al.*, 1989). It has been suggested that the deletion events observed with cosmids are RecA-independent and the use of a host strain deficient in other recombination pathways may help prevent cosmids from deleting their insert DNA (Ishiura *et al.*, 1990).

Ishiura *et al.* have shown that deletions have occurred in the insert portions between two short complete direct repeats in *recA* and *rec*$^+$ hosts (Ishiura *et al.*, 1990). There has been a report that homologous recombination mediated by the *recA* recombination system usually involves larger homologous sequences (Shen and Huang, 1986). The genome of *S. solfataricus* P2 contains many sites for recombination because of the presence of ISC elements and other short repeats. The *recA* recombination system may not be effective in eliminating the deletion events and it appears that other recombination systems are implicated. Another host system deficient in these mechanisms is required. Such a host has been engineered which is able to stably propagate recombinant cosmids without structural or functional alterations in these clones (Ishiura *et al.*,

1989). The drawbacks of this host are the nonhomogeneous growth, low yields of cosmid DNA and low transduction effieciency.

D) Future Work

Cloning of the whole genome of *S. solfataricus* P2 into cosmid does not seem possible. Other methods must be employed to supplement this library. Initially, it appeared that *Eag* I was the only candidate restriction enzyme that could be used for macrorestriction mapping but recent genomic digests by Ghislaine Allard with *Not* I have proven useful.

In order to overcome the toxic or incompatible nature of *S. solfataricus* DNA with the host system which prevents the cloning of certain regions, another vector system should be employed; specifically a phage-based one because the DNA gets cloned and host viability is not quite essential.

E) Conclusions

The dogma that genes, which are composed of DNA, code for structurally and biochemically important molecules such as proteins is entrenched in our pursuit to obtain the complete DNA sequence of model organisms. Initially, the study of an organism began at the level of the gene and now we have the tools that allow us to study an organism at the genome level. Genome projects allow us to study an organism's genetic make-up; providing information about function, organization and structure. These genome projects also allows us to employ and invent techniques to reach this goal. Cloning the entire genome provides

the means to determine the entire DNA complement of an organism with segments of the genome divided into manageable pieces essential for contig mapping and DNA sequencing.

The goal to generate a high resolution physical map was essential for the strategy of sequencing on a cosmid-by-cosmid basis for the *S. solfataricus* P2 Genome Project. The physical map of this genome would act like a road map displaying the exact location of each cosmid clone. With each cloned segment of the genome containing single genes and operons, it allowed for easy access to each region. This also facilitates the dissemination of clones among researchers who have interests in particular genes.

The unstable nature of *S. solfataricus* P2 DNA and its toxicity to the employed host system has prevented the cloning of its entire genome in cosmids, but there exists other methods such as macrorestriction mapping, phage-based cloning that will allow us to supplement the cosmid library and achieve our goals. Currently, this cosmid library is being supplemented by λ phage clones with PCR for the final few pieces for complete coverage of this genome so that DNA sequencing can be completed.

Appendix

Distribution of Cosmid Clones from the *Hind* III Library Among the Linked Contigs (As of May 9, 1996)

Contig 1

s13d3: s4g1, s14b1, s15c8, s15e11, s18a9, s21d9

s4b8: s1h9, s1h12, s2e7, s4e12, s4g12, s4f4, s14a4, s15h5, s17g1, s18d2, s18d9, s18f8, s20d4, s20e10, s22a9, s22b9, s22f5

s1a8: s1b9, s1c12, s1d3, s1e1, s1f10, s1h4, s13a9, s13a10, s13c1, s13c4, s13e1, s13g8, s13h5, s13h9, s14a3, s14b4, s14b8, s14c3, s14c9, s14e3, s14f11, s14h1, s15g1, s17a4, s17h4, s18a7, s18a10, s19e11, s19g3, s20b6, s20e3, s20e7, s21c1, s21c2, s21e3, s21e4, s21f7, s21g1, s22a7, s22c2, s22e8, s22h7, s23a3, s24a8, s24b4

s1e7: s1a6, s1a10, s1b3, s1b4, s1b7, s1b12, s1c4, s1c9, s1e2, s1e6, s1e10, s1f5, s1f6, s1f7, s1f8, s1f9, s1f11, s1g4, s1g5, s1g6, s1g7, s1g9, s1h1, s1h2, s1h3, s2a3, s2a10, s2b4, s2c8, s2c11, s2d1, s2d3, s2d7, s2f11, s2g4, s2g7, s2g9, s2h6, s2h9, s3a1, s3b3, s3b5, s3b9, s3b10, s3c6, s3c9, s3c11, s3e12, s3f4, s3h5, s4a9, s4d7, s4e8, s4f6, s4f12, s4g4, s4h1, s4h4, s4h5, s13a1, s13a3, s13a5, s13a8, s13b5, s13b9, s13b12, s13c9, s13c11, s13c12, s13d2, s13e9, s13e10, s13f5, s13f10, s13g7, s13h3, s13h4, s13h7, s14b3, s14b9, s14c2, s14c6, s14c12, s14d4, s14e1, s14e4, s14e5, s14e11, s14f5, s14g7, s14h3, s14h4, s14h7, s15a1,s15a2, s15a4, s15a5, s15a9, s15a10, s15a11, s15a12, s15b4, s15b6, s15b8, s15c2, s15c4, s15c11, s15d4, s15d9, s15d11, s15e1, s15e8, s15g11, s15h7, s16c3, s16c6, s16c9, s17a1, s17a3, s17a5, s17b1, s17b5, s17b8, s17b10, s17c1, s17d2, s17d4, s17d7, s17d8, s17e11, s17e12, s17f5, s17g12, s17h7, s17h8, s18a1, s18a2, s18a4, s18a5, s18b7, s18c1, s18c8, s18d1, s18d3, s18d5, s18e3, s18e5, s18e6, s18f6, s18f9, s18g4, s18g5, s18g8, s18g12, s18h2, s18h10, s19a2, s19a9, s19b5, s19b6, s19b7, s19b9, s19b12, s19c9, s19c12, s19d1, s19d7, s19e9, s19f1, s19f3, s19f6, s19f7, s19f8, s19f9, s19f10, s19g1, s19g7, s19g9, s19g11, s19h5, s19h8, s19h9, s19h10, s20a7, s20a11, s20b3, s20b7, s20b8, s20b9, s20c4, s20c5, s20c6, s20c9, s20c10, s20d1, s20d6, s20d10, s20f3, s20f11, s20g2, s20g9, s20h3, s20h7, s21a2, s21a9, s21b2, s21b3, s21b4, s21b5, s21b11, s21b12, s21c3, s21c6, s21c10, s21c11, s21d3, s21d4, s21d5, s21e7, s21f4, s21f9, s21f11, s21g9, s21h4, s21h6, s22a5, s22a6, s22a8, s22a11, s22b8, s22b11, s22c5, s22c11, s22d3, s22d4, s22d8, s22e9, s22g2, s22g5, s22g8, s22h1, s22h4, s22h11, s22h12, s23a1, s23a7, s23a8, s23a12, s24a3, s24a5, s24b10, s24b11

polymorphic clones:

s15e10u

s2d4, s22g7: s2d12, s2e1, s2g10, s3a10, s3b4, s3c4, s3d9, s3e5, s3f2, s3f8, s3f11, s3g2, s3g9, s4a3, s4b4, s4b6, s4c7, s4f7, s4f11, s13a7, s13b7, s13b8,s13d11, s13f7, s15d1, s15e2, s15e4, s15g5, s15g10, s15h10, s16c10, s17b2, s17c11, s17e4, s17f6, s18d12, s18e8, s19e6, s20b11, s20g6, s20h6, s21b1, s21b7, s21d2, s22d1

s13d4: s2a4, s3c12, s3d4, s13f8, s15a6, s15e7, s15f6, s18h8, s20c11

s15c1u

s20b5u

s22a12u

already characterized:

s2f3: s13f6

s3f7: s1b11, s1c3, s2a2, s2d11, s2e10, s3e6, s3g1, s4d2, s4e2, s4g5, s4h3, s13c2, s13h6, s15a7, s15b2, s15b11, s15c3, s15e9, s18a6, s18d7, s18f3, s18f7, s18f12, s18g1, s19c2, s19d5, s19e8, s19g5, s20c12, s20e5, s20f4, s20h5, s21f8, s22a1

Contig 2

s2g8: s21a7, s21c8

s1d5: s1b6, s1c6, s1g8, s2b9, s2c5, s2c6, s2e2, s2g5, s3h9, s4b3, s4c2, s4d9, s13c3, s13d1, s13f2, s13h2, s13h10, s14c4, s14f6, s14g1, s14g2, s14h6, s15b5, s15e5, s15f11, s17a9, s17b11, s17e8, s17f4, s18c4, s18d10, s18e9, s18g9, s18h3, s19a3, s19a11, s19b2, s19c1, s19c4, s19d3, s19d10, s19h1, s19h4, s20a2, s21a10, s21c4, s21e11, s21f10, s21g5, s21h10, s21h11, s22b4, s22c4, s22f2

s3f9: s1c7, s1d4, s1d9, s2a8, s2b3, s2b12, s2c1, s2c3, s2c12, s2f8, s2h12, s3a7, s3b12, s3d3, s3e9, s3f12, s4b2, s4c11, s4d11, s4d12, s4e11, s4g3, s4h6, s13f9, s14g6, s15b7, s15e3, s15h9, s17a11, s17c12, s17d3, s17d5, s17g5, s18a8, s18b5, s18b12, s18c12, s18h5, s19b3, s19c3, s19e4, s19e12, s19g8, s20b10, s20d5, s20e1, s20e6, s20h1, s21d1, s21d6, s21d12, s21e6, s22a4

s4c8, s2a7: s1a2, s1d6, s1h6, s2a6, s2f1, s2g1, s2h10, s3a5, s3b6, s3e3, s3e7, s3f10, s4a7, s4c6, s4c9, s4e4, s4e6, s4f1, s4f10, s4g9, s13b2, s13c7, s13d6, s13d12, s14d8, s14d12, s14e6, s14g12, s15b1, s15c9, s15f9, s15h4, s15h6, s17a10, s17d6, s17e2, s17e3, s18c2, s18d11, s18e10, s19c7, s19d2, s19h2, s20c3, s20f9, s20g10, s20h4, s21a8, s21e9, s21g2, s21g10, s22c12

s22g9: s1d10, s1e5, s1e8, s2f6, s3d5s3g3, s3h11, s4d10, s4g10, s14b11, s15d8, s15d10, s18b10, s18b11, s18c10, s18h12, s19c5, s20e4, s20f5, s21d11, s21f5, s21g11, s22d10, s22f3, s23a2

s18e4: s4c10

s4g11: s2e11, s13f1, s13g12, s14a6, s14a8, s14b12, s14g3, s14g11, s17e1, s18b1, s18d4, s19d8, s19f5, s20c1, s20g12, s20h2, s21a4, s21a12, s22b3, s22d5, s22e3, s22f1, s22h6, s24b2

s22e4: s15b12, s22b6

variant clone:

s2h3u

Contig 3

s1e4: s3a3, s3b11, s3e1, s3g5, s4f8, s14a9, s14c10, s17a2, s17c3, s20f2, s21c12, s21h3, s23a9, s24a6

s19c10: s21b9

s20c7: s1c8, s1h8, s2c10, s2d2, s2e8, s3d7, s3h7, s4d1, s13h8, s14f4, s15h11, s17f12, s18f2, s18h11, s19h3, , s21c5, s22d12, s24a9, s24b3

s1d8, s2d6: s1g10, s2b11, s2f9, s3g10, s3h1, s3h3, s4a11, s4d4, s4h11, s13b11, s13g4, s14a5, s14e8, s15f4, s15h1, s16c12, s17e10, s17g6, s17g7, s18a3, s18b6, s18d6, s18h4, s19a10, s19h6, s19h11, s20d8, s21b10, s21c7, s21g6, s21h1, s21h5, s22c9, s23a10

variant:

s13b6: s2h11, s3a12, s19g10, s20g4, s21d10, s22b1

Contig 4

s13a6, s19h12: s1a9, s2d10, s2f12, s3c5, s3f6, s4e7, s4e9, s14a10, s15d12, s15g2, s15g3, s18a12, s22h3

to place:

(s13b1: s18h9)

Contig 5

s13b4, s3g11: s13e3, s13e4, s13e5, s13f3, s13g3, s14h10, s15c12, s20d12, s20f6, s21e10, s24b9

Contig 6

s14b2: s20e2

s3g8: s1d11, s1f1, s2e3, s13e7, s13e11, s14e9, s14h5, s15f2, s17b3, s17b7, s18b8, s20b4, s21b8, s22e12, s24b5, s24b6

Contig 7

s3d2, s4b7: s1a4, s1a11, s2b10, s3a4, s3b2, s3c2, s3d10, s3g12, s4c3, s4d3, s4g8, s4h12, s13d8, s13f4, s14a7, s14f1, s14g9, s14h11, s15f3, s15f5, s15g4, s15g7, s15h3, s16c8, s17b4, s17c4, s17e9, s17g11, s17h3, s18e1, s18e11, s18g10, s19a6, s19e2, s20d11, s20f1, s21f1, s22f10, s22h5

variant:

s16c5: s20g3

Contig 8

s22f4u

s1e9: s2b2, s3h8, s4d6, s13f12, s14a12, s14c11, s14e12, s15h8, s16c1, s16c11, s17h1, s18c5, s19a8, s20a3, s21e2, s23a5

s19a4u

s14c8: s1c10, s1h10, s3c10, s3d12, s4f3, s14d1, s15e6, s16c7, s17e7, s17g3, s19h7, s20a5, s21e12, s22d11, s22g1, s24a7, s24b7

s4h9, s22e2: s1b5, s1h7, s2e6, s3d6, s4f5, s4h2, s14b5, s14d2, s14d6, s17a12, s17c9, s17f9, s18c3, s18e12, s18g6, s19a7, s19d12, s19e5, s20f7, s21a1, s21g8, s21g12, s22e5, s22f12, s22g10, s24a11, s24b12

to place:

(s2b7: s2e4, s2h7, s3h2, s13f11, s13g10, s14d10, s14h2, s14h9, s15b9, s15f8, s17d12, s17f7, s17h11, s17h12, s18e2, s20d2, s20e11, s20g5, s22f11)

(s18b3: s14g5)

Contig 9

s4a8: s1d7, s4c5, s14d3, s14d7, s15g9, s21f3

s2e12: s3h4, s21f2, s21h12

Contig 10

s13g6: s2g12, s3a11, s13h11, s14a2, s14e2, s19b8, s20f10, s22e10

s2a9: s1f4, s1h11, s13c6, s14a1, s18f1, s19c8, s20a8, s20b12, s20g1, s21a5, s21f6, s22b5, s22c6, s22e11

variant:

s20f8u

Contig 11

s21b6u

to place:

(s22b7u)

Contig 12

s3a2, s2g2: s1g11, s3d11, s3h12, s13g9, s13g11, s14c1, s14e10, s14f8, s17f10, s18g2, s22e1, s24a1

s3a6: s1a12, s4a2

to place:

(s1a7u)

(s19e10: s20h8)

(s13a11u)

(**s20a10**: s3b1, s3g4, s3g6, s13d9, s17d11, s17h10, s19e1, s20a12, s20b2, s20h10, s21a11, s22d2, s22g4)

variant:

s14b10: s1b8, s2h5, s20a9, s22e7, s23a11

Contig 13

s1c11, s15d5: s1d2, s1g12, s3f5, s4h7, s14b6, s14h8, s15a3, s15c10, s15d3, s15e12, s17g4, s18f10, s20a1, s20f12, s20h12, s21a3, s21a6, s22c3, s22d6, s22f8, s24a4

Contig 14

s18f4, s4a10: s4c4, s4c12, s15b10, s22d7, s24b8

to place:

(s14a11u)

Contig 15

s3b7: s1f2, s2h1, s15d2, s17c6, s17c10, s18h1, s19d11, s19f2, s22f9, s23a4, s23a6, s24a10

to place:

(s21e8u)

(s22c10u)

(**s1f3**: s2a1, s2g3, s13g2, s15g12, s18c9, s19g4, s20a6, s21h2)

variants:

s19a12: s3e4

s1g3: s3h6

s19g12u

Contig 16

s4b5, s3e2, s3g7: s2f4, s13h12, s14f10, s19c6, s20d3, s22b12, s22h10

Contig 17

s1e3, s13e8: s2b6, s2d5, s2f10, s13c5, s13d7, s13d10, s15a8, s21f12, s21h7, s22c1, s22g12

to place:

(s18f5u)

Contig 18

s13a4: s2c7, s14g4, s5d7, s20e12

to place:

(s3a8: s2h4, s15b3, s17d1)

(s22b10: s14c7)

Contig 19

s19b1: s4a5, s4a6, s17b6, s17b12, s20g7

Contig 20

s18h7: s15g8, s18e7

Contig 21

s4a1, s2c9, s1f12: s1a5, s1b2, s1e11, s2f5, s2h2, s2h8, s4d5, s13e2, s13h1, s14d9, s14f3, s14g8, s15c7, s15f10, s17a6, s18b4, s18b9, s18g11, s19e7, s20a4, s20c8, s20g11, s20h9, s21g7, s22e6, s22g11, s22h2, s22h8, s22h9

s13a2: s2b5, s2d8, s13b3, s22a2

to place:

(s22d9: s2c4, s3f3, s4b9, s13b10, s14d5, s14e7, s15c6, s15f1, s16c2, s17f3, s17h2, s19c11, s20d9, s20e9, s21e5, s21g4)

(s2g11u)

Contig 22

s3f1: s1g2, s19b11

s2d9: s2a12, s2e5, s13g5, s14d11, s14f2, s14f7, s16c4, s18b2, s18g3, s19a1, s19b4, s22b2, s22c7, s24a12

to place:

(s13e6u)

Contig 23

s20b1: s1e12, s2a11, s2b1, s2e9, s2f7, s3b8, s3e8, s13d5, s17f2, s18h6, s19e3, s19f4, s21c9, s22a3

to place:

(s15h12u)

Contig 24

s13e12: s13c8, s13c10, s17f11, s18c11, s21g3

Contig 25

s3a9: s3c3

Contig 26

s19b10u

Contig 27

s19g2u

Contig 28

s17g10u

Contig 29

s17h9u

Contig 30

s17f8: s2b8, s15c5, s17g2, s19a5, s19d9

Contig 31

s3h10: s14f9, s15d6, s17c8

s20c2u

References

Barns, S.M., Fundyga, R.E., Jeffries, M.W. and Pace, N.R. (1994) Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. USA, **91**, 1609-1613.

Berg, D.E. and Howe, M.M. (Eds): Mobile DNA. Washington, D.C.: American Society for Microbiology; 1989.

Birkenbihl, R.P. and Vielmetter, W. (1989) Cosmid-derived map of *E. coli* strain BHB2600 in comparison to the map of strain W3110. Nucl. Acids. Res., **17**(13), 5057-5069.

Bobovnikova, Y., Ng, W.-L., DasSarma, S. and Hackett, N.R. (1994) Restriction Mapping the Genome of *Halobacterium halobium* Strain NRC-1. System. Appl. Microbiol., **16**, 597-604.

Brock, T.D., Brock, K.M., Belly, R.T. and Weiss, R.L. (1972) *Sulfolobus*: A New Genus of Sulfur-Oxidizing Bacteria Living at Low pH and High Temperature. Arch. Microbiol., **84**, 54-68.

Brown, J.R. and Doolittle, W.F. (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. Proc. Natl. Acad. Sci. USA, **92**, 2441-2445.

Bukanov, N.O. and Berg, D.E. (1994) Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. Mol. Microbiol., **11**(3), 509-523.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts, K.M., Hurst, M.A., Kaine, B.P., Borodovsky, M., Klenk, H.-P., Fraser, C.M., Smith, H.O., Woese, C.R. and Venter, J.C. (1996) Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. Science, **273**, 1058-1073.

Burggraf, S., Stetter, K.O., Rouviere, P. and Woese, C.R. (1991) *Methanopyrus* kandleri: An Archaeal Methanogen Unrelated to all Other Known Methanogens. System. Appl. Microbiol. **14**, 346-351.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors. Science, **236**, 806-812.

Charlebois, R.L. and Doolittle, W.F. (1989) Transposable Elements and Genome Structure in Halobacteria. In *Mobile DNA*. Berg, D.E. and Howe, M.M. (eds.). Washington, D.C.: American Society for Microbiology., pp. 297-307.

Charlebois, R.L., Hofman, J.D., Schalkwyk, L.C., Lam, W.L. and Doolittle, W.F. (1989) Genome mapping in halobacteria. Can. J. Microbiol., **35**, 21-29.

Charlebois, R.L., Schalkwyk, L.C., Hofman, J.D. and Doolittle, W.F. (1991) Detailed Physical Map and Set of Overlapping Clones Covering the Genome of the Archaebacterium *Haloferax volcanii* DS2. J. Mol. Biol., **222**, 509-524.

Charlebois, R.L., Gaasterland, T., Ragan, M.A., Doolittle, W.F. and Sensen, C.W. (1996) The *Sulfolobus solfataricus* P2 genome project. FEBS Letters, **389**, 88-91.

Charlebois, R.L. (1993) Physical Mapping of Genomes Using the Landmark Strategy. In *The Second International Conference on Bioinformatics, Supercomputing and Complex Genome Analysis*. Lim, H.A, Fickett, J.W., Cantor, C.R. and Robbins, R.J. (eds.). Singapore: World Scientific., pp. 219-229.

Cohen, A., Lam, W.L., Charlebois, R.L., Doolittle, W.F. and Schalkwyk, L.C. (1992) Localizing genes on the map of the genome of *Haloferax volcanii*, one of the Archaea. Proc. Natl. Acad. Sci. USA, **89**, 1602-1606.

Cole, S.T. and Saint Girons, I. (1994) Bacterial genomics. FEMS Microbiol. Rev., **14**, 139-160.

Coulson, A., Sulston, J., Brenner, S. and Karn, J. (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA, **83**, 7821-7825.

Craig, A.G., Nizetic, D., Hoheisel, J.D., Zehetner, G. and Lehrach, H. (1990) Ordering of cosmid clones covering the Herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation. Nucl. Acids. Res., **18**(9), 2653-2660.

Danson, M.J, Hough, D.W. and Lunt, G.G. (Eds): The Archaebacteria: *Biochemistry and Biotechnology*. London: Portland Press; 1992.

DasSarma, S., RajBhandary, U.L. and Khorana, H.G. (1983) High-frequency spontaneous mutation in the bacterio-opsin gene in *Halobacterium halobium* is mediated by transposable elements. Proc. Natl. Acad. Sci. USA, **80**, 2201-2205.

Delong, E.F. (1992) Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA, **89**, 5685-5689.

Delong, E.F., Wu, K.Y., Prezelin, B.B. and Jovine, R.V.M. (1994) High abundance of Archaea in Antarctic marine picoplankton. Nature, **371**, 695-697.

Derkacheva, N.I. and Kagramanova, V.K. (1994) Archaebacterial Chromosomes (A Review). Mol. Biol., **28**(2), 178-184.

De Smet, K.A.L, Jamil, S. and Stoker, N.G. (1993) Tropist3: a cosmid vector for simplified mapping of both G+C-rich and A+T-rich genomic DNA. Gene, **136**, 215-219.

Fonstein, M. and Haselkorn, R. (1993) Chromosomal structure of *Rhodobacter capsulatus* strain SB1003: Cosmid encyclopedia and high-resolution physical and genetic map. Proc. Natl. Acad. Sci. USA, **90**, 2522-2526.

Fonstein, M. and Haselkorn, R. (1995) Physical Mapping of Bacterial Genomes. J. of Bact., **177**(12), 3361-3369.

Fonstein, M., Koshy, E.G., Nikolskaya, T., Mourachov, P. and Haselkorn, R. (1995) Refinement of the high-resolution physical and genetic map of *Rhodobacter capsulatus* and genome surveys using blots of the cosmid encyclopedia. EMBO J., **14**(8), 1827-1841.

Fox, G.E., Magrum, L.J., Balch, W.E., Wolfe, R.S. and Woese, C.R. (1977) Classification of methanogenic bacteria by 16S ribosomal RNA characterization. Proc. Natl. Acad. Sci. USA, **74**(10), 4537-4541.

Gogarten, J. P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Denda, K. and Yoshida, M. (1989) Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA, **86**, 6661-6665.

Green, E.D. and Olson, M.V. (1990) Chromosomal Region of the Cystic Fibrosis Gene in Yeast Artificial Chromosomes: A Model for Human Genome Mapping. Science, **250**, 94-98.

Grogan, D.W. (1989) Phenotypic Characterization of the Archaebacterial Genus *Sulfolobus*: Comparison of Five Wild-Type Strains. J. of Bact., **171**(12), 6710-6719.

Grogan, D.W. (1996) Exchange of Genetic Markers at Extremely High Temperatures in the Archaeon *Sulfolobus acidocaldarius*. J. of Bact., **178**(11), 3207-3211.

Gutierrez, M.C., Ventosa, A. and Ruiz-Berraquero, F. (1989) DNA-DNA homology studies among strains of *Haloferax* and other halobacteria. Curr. Microbiol., **18**, 253-256.

Hamilton, P.T. and Reeve, J. N. (1985) Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii*. Mol. Gen. Genet., **200**, 47-59

Hofman, J.D., Schalkwyk, L.C. and Doolittle, W.F. (1986) ISH51: a large, degenerate family of insertion sequence-like elements in the genome of the archaebacterium, *Halobacterium volcanii*. Nucl. Acids. Res., **14**(17), 6983-7000.

Ishiura, M., Hazumi, N., Koide, T., Uchida, T. and Okada, Y. (1989) A *recB recC sbcB rec J* Host Prevents *recA*-Independent Deletions in Recombinant Cosmid DNA Propagated in *Escherichia coli*. J. of Bact., **171**(2), 1068-1074.

Ishiura, M., Hazumi, N., Shinagawa, H., Nakata, A., Uchida, T. and Okada, Y. (1990) *RecA*-independent high-frequency deletion of recombinant cosmid DNA in *Escherichia coli*. J. of Gen. Microbiol., **136**, 69-79.

Iwabe, N., Kuma, K.-I., Hasegawa, M., Osawa, S. and Miyata, T. (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA, **86**, 9355-9359.

Jones, W.J., Nagle, D.A. and Whitman, W.B. (1987) Methanogens and the diversity of archaebacteria. Microbiol., Rev., **51**, 135-177.

Kargi, F. and Robinson, J.M. (1982) Removal of Sulfur Compounds from Coal by the Thermophilic Organism *Sulfolobus acidocaldarius*. Appl. Environ. Microbiol., **44**, 878-883.

Kates, M., Kushner, D.J. and Matheson, A.T. (Eds): The Biochemistry of Archaea (Archaebacteria). Amsterdam: Elsevier; 1993.

Keeling, P.J., Charlebois, R.L. and Doolittle, W.F. (1994) Archaebacterial genomes: eubacterial form and eukaryotic content. Curr. Opin. in Genet. and Develop., **4**, 816-822.

Keeling, P.J. and Doolittle, W.F. (1995) Archaea: Narrowing the gap between prokaryotes and eukaryotes. Proc. Natl. Acad. Sci. USA, **92**, 5761-5764.

Kim, U.-J., Shizuya, H., de Jong, P.J., Birren, B. and Simon, M.I. (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucl. Acids. Res., **20**(5), 1083-1085.

Klein, A. and Schnorr, M. (1984) Genome complexity of methanogenic bacteria. J. Bact., **158**, 628-631.

Kohara, Y., Akiyama, K. and Isono, K. (1987) The Physical Map of the Whole *E. coli* Chromosome: Application of a New Strategy for Rapid Analysis and Sorting of a Large Genomic Library. Cell, **50**, 495-508.

Kondo, S., Yamagishi, A. and Oshima, T. (1993) A Physical Map of the Sulfur-Dependent Archaebacterium *Sulfolobus acidocaldarius* 7 Chromosome. J. of Bact., **175**(5), 1532-1536.

Krawiec, S. and Riley, M. (1990) Organization of the Bacterial Chromosome. Microbiol. Rev., **54**(4), 502-539.

Lawson, F.S., Charlebois, R.L. and Dillon, J.-A. R. (1996) Phylogenetic analysis of carbamoylphosphate synthetase genes: complex evolutionary history includes an internal duplication within a gene which can root the tree of life. Mol. Biol. and Evol., **13**, 970-977.

Lindstrom, E.B., Wold, S., Kettaneh-Wold, N. and Saaf, S. (1993) Optimization of pyrite bioleaching using *Sulfolobus acidocaldarius*. Appl. Microbiol. Biotechnol., **38**, 702-707.

Little, P.F.R and Cross, S.H. (1985) A cosmid vector that facilitates restriction enzyme mapping. Proc. Natl. Acad. Sci. USA, **82**, 3159-3163.

Lloyd, R.G. and Brooks Low, K. (1996) Homologous Recombination. In *Escherichia coli and Salmonella Cellular and Molecular Biology, 2nd Edition, Volume 2.* Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (eds.). Washington, D.C.: American Society for Microbiology., pp. 2236-2255.

Lopez-Garcia, P., Abad, J.P., Smith, C., and Amils, R. (1992) Genomic organization of the halophilic archaeon *Haloferax mediterranei*: physical map of the chromosome. Nucl. Acids. Res., **20**(10), 2459-2464.

Lucier, T.S., Hu, P.-Q., Peterson, S.N., Song, X.-Y., Miller, L., Heitzman, K., Bott, K.F., Hutchison III, C.A. and Hu, P.-C. (1994) Construction of an ordered genomic library of *Mycoplasma genitalium*. Gene, **150**, 27-34.

Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Low, K., B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (Eds): *Escherichia coli and Salmonella* Cellular and Molecular Biology. 2nd Edition, Volume 2. Washington, D.C.: American Society for Microbiology; 1996.

Noll, K.M. (1989) Chromosome Map of the Thermophilic Archaebacterium *Thermococcus celer*. J. of Bact., **171**(12), 6720-6725.

Nolling, J., van Eeden, F.J.M. and de Vos, W.M. (1993) Distribution and characterization of plasmid-related sequences in the chromosomal DNA of different thermophilic *Methanobacterium strains*. Mol. Gen. Genet., **240**, 81-91.

Norris, P.R. (1992) Thermoacidophilic archaebacteria: potential applications. In The Archaebacteria: *Biochemistry and Biotechnology*. Danson, M.J., Hough, D.W. and Lunt, G.G. (eds.). London: Portland Press, pp. 171-180.

Olson, M.V., Dutchik, J.E., Graham, M.Y., Brodeur, G.M., Helms, C., Frank, M., MacCollin, M., Scheinman, R. and Frank, T. (1986) Random-clone strategy for genomic restriction mapping in yeast. Proc. Natl. Acad. Sci. USA, **83**, 7826-7830.

Peterson, S.N., Lucier, T., Heitzman, K., Smith, E.A., Bott, K.F., Hu, P.-C. and Hutchison III, C.A. (1995) Genetic Map of the *Mycoplasma genitalium* Chromosome. J. of Bact., **177**(11), 3199-3204.

Pfeifer, F., Weidinger, G. and Goebel, W. (1981) Genetic Variability in *Halobacterium halobium*. J. of Bact., **145**(1), 375-381.

Pfeifer, F., Friedman, J., Boyer, H.W. and Betach, M. (1984) Characterization of insertions affecting the expression of the bacterio-opsin gene in *Halobacterium halobium*. Nucl. Acids. Res., **12**(5), 2489-2497.

Pfeifer, F. and Betlach, M. (1985) Genome organization in *Halobacterium halobium*: A 70 kb island of more (AT) rich DNA in the chromosome. Mol. Gen. Genet., **198**, 449-455.

Pfeifer, F. (1986) Insertion Elements and Genome Organization of *Halobacterium halobium*. System. Appl. Microbiol., **7**, 36-40.

Pfeifer, F., Blaseio, U. and Ghahraman, P. (1988) Dynamic Plasmid Populations in *Halobacterium halobium*. J. of Bact., **170**(8), 3718-3724.

Pfeifer, F. and Blaseio, U. (1989) Insertion Elements and Deletion Formation in a Halophilic Archaebacterium. J. of Bact., **171**(9), 5135-5140.

Pfeifer, F., Blaseio, U. and Horne, M. (1989) Genome structure of *Halobacterium halobium*: plasmid dynamics in gas vacuole deficient mutants. Can. J. Microbiol., **35**, 96-100.

Ramirez, C., Kopke, A.K.E., Yang, D.-C., Boeckh, T. and Matheson, A.T. (1993) The structure, function and evolution of archaeal ribosomes. In *The Biochemistry of Archaea (Archaebacteria)*. Kates, M., Kushner, D.J. and Matheson, T. (eds.). Amsterdam : Elsevier. , pp. 439-466.

Roth, J.R., Benson, N., Galitski, T., Haach, K., Lawrence, J.G. and Miesel, L. (1996) Rearrangements of the Bacterial Chromosome: Formation and Applications. In *Escherichia coli and Salmonella Cellular and Molecular Biology, 2nd Edition, Volume 2*. Neidhardt, F.C., Curtiss III, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (eds.). Washington, D.C.: American Society for Microbiology., pp. 2256-2276.

Sapienza, C. and Doolittle, W.F. (1982) Unusual physical organization of the *Halobacterium* genome. Nature, **295**, 384-389.

Sapienza, C., Rose, M.R. and Doolittle, W.F. (1982) High-frequency genomic rearrangements involving archaebacterial repeat sequence elements. Nature, **299**, 182-185.

Schleper, C., Kubo, K. and Zillig, W. (1992) The particle SSV1 from the extremely thermophilic archaeon *Sulfolobus* is a virus: Demonstration of infectivity and of transfection with viral DNA. Proc. Natl. Acad. Sci. USA, **89**, 7645-7649.

Schleper, C., Roder, R., Singer, T. and Zillig, W. (1994) An insertion element of the extremely thermophilic archaeon *Sulfolobus solfataricus* transposes into the endogenous β-galactosidase gene. Mol. Gen. Genet., **243**, 91-96.

Schleper, C., Holz, I., Janekovic, D., Murphy, J. and Zillig, W. (1995) A Multicopy Plasmid of the Extremely Thermophilic Archaeon *Sulfolobus* Effects Its Transfer to Recipients by Mating. J. of Bact., **177**(15), 4417-4426.

Sensen, C.W., Klenk, H.-P., Singh, R.K., Allard, G., Chan, C.C-Y., Liu, Q.Y., Young, F., Schenk, M., Gaasterland, T., Doolittle, W.F., Ragan, M.A. and Charlebois, R.L. (1996) Organizational characteristics and information content of an archaeal genome: 156 kbp of sequence from *Sulfolobus solfataricus* P2. Mol. Microbiol., **22**(1), 175-191.

Sensen, C.W., Charlebois, R.L., Singh,R.K., Klenk, H.-P., Ragan, M.A. and Doolittle, W.F. Sequencing the genome of *Sulfolobus solfataricus* P2. In: (F.J. de Bruijn, J.R.Lupski, G. Weinstock, eds.) Bacterial Genomes: Physical Structure and Analysis, Chapman & Hall, New York, in press.

Shen, P. and Huang, H.V. (1986) Homologous Recombination in *Escherichia coli*: Dependence on Substrate Length and Homology. Genetics, **112**, 441-457.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA, **89**, 8794-8797.

Sitzmann, J. and Klein, A. (1991) Physical and genetic map of the *Methanococcus voltae* chromosome. Mol. Microbiol., **5**(2), 505-513.

Smith, C.L. and Condemine, G. (1990) New Approaches for Physical Mapping of Small Genomes. J. of Bact., **172**(3), 1167-1172.

Staley, J., Marvin, P., Pfennig, N. and Holt, J. (1989) In *Bergey's manual of systematic bacteriology*. Baltimore: Williams and Wilkins. vol. 3, pp. 2216-2228.

Sternberg, N. (1990) Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci. USA, **87**, 103-107.

Stetter, K.O., Huber, R., Blochl, E., Kurr, M., Eden, R.D., Fielder, M., Cash, H. and Vance, I. (1993) Hyperthermophilic archaea are thriving in deep North Sea and Alaskan oil reservoirs. Nature, **365**, 743-745.

Stettler, R. and Leisinger, T. (1992) Physical Map of the *Methanobacterium thermoautotrophicum* Marburg Chromosome. J. of Bact., **174**(22), 7227-7234.

St. Jean, A., Trieselmann, B.A. and Charlebois, R.L. (1994) Physical map and set of overlapping cosmid clones representing the genome of the archaeon *Halobacterium* sp. GRB. Nucl. Acids. Res., **22**(8), 1476-1483.

Symington, L.S., Morrison, P. and Kolodner, R. (1985) Intramolecular Recombination of Linear DNA Catalyzed by the *Escherichia coli RecE* Recombination System. J. Mol. Biol., **186**, 515-525.

Tabata, S., Higashitani, A., Takanami, M., Akiyama, K., Kohara, Y., Nishimura, Y., Nishimura, S. and Hirota, Y. (1989) Construction of an Ordered Cosmid Collection of the *Escherichia coli* K-12 W3110 Chromosome. J. of Bact., **171**(2), 1214-1218.

Wenzel, R. and Herrmann, R. (1988) Physical mapping of the *Mycoplasma pneumoniae* genome. Nucl. Acids. Res., **16**(17), 8323-8336.

Wenzel, R. and Herrmann, R. (1989) Cloning of the complete *Mycoplasma pneumoniae* genome. Nucl. Acids. Res., **17**(17), 7029-7043.

Whittaker, R.H. (1959) On the Broad Classification of Organisms. Q. Rev. Biol. **34**, 210-226.

Woese, C.R., Fox., G.E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B.J. and Stahl, D. (1975) Conservation of primary structure in 16S ribosomal RNA. Nature, **254**, 83-86.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. USA, **74**(11), 5088-5090.

Woese, C.R. (1987) Bacterial Evolution. Microbiol. Rev., **51**(2), 221-271.

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA, **87**, 4576-4579.

Yamagishi, A. and Oshima, T. (1990) Circular chromosomal DNA in the sulfur-dependent archaebacterium *Sulfolobus acidocaldarius*. Nucl. Acids. Res., **18**(5), 1133-1136.

Yuan, R. (1981) Structure and mechanism of multifunctional restriction endonucleases. Annu. Rev. Biochem., **50**, 285-319.

Zillig, W. (1981) Eukaryotic traits in archaebacteria. Ann. N.Y. Acad. Sci., **503**, 78-81.

Zillig, W., Holz, I., Janekovic, D., Schafer, W. and Reiter, W.D. (1983) The archaebacterium *Thermococcus celer* represents a novel genus within the thermophilic branch of the archaebacteria. Syst. Appl. Microbiol., **4**, 88-94.

Zillig, W., Kletzin, A., Schleper, C., Holz, I., Janekovic, D., Hain, J., Lanzendorfer, M. and Kristjansson, J.K. (1994) Screening for *Sulfolobales*, their Plasmids and their Viruses in Icelandic Solfataras. System. Appl. Microbiol., **16**, 609-628.