Haloarchaeal Comparative Genomics and the Local Context Model of Genomic Evolution

Andrew Louis St. Jean

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Ottawa-Carleton Institute of Biology

ISBN 0-612-20020-5

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ABSTRACT

Genomics is a rapidly expanding field of research that seeks to study the structure, function and evolution of an organism's genome. Genomic investigations were conducted on three species of haloarchaea, a monophyletic group of prokaryotes belonging to the kingdom Euryarchaeota of the domain Archaea that are adapted to high-salt environments. A physical and genetic map of the genome of *Halobacterium salinarum* GRB is described. This map and the previously published map of the genome of *Haloferax volcanii* DS2 were compared with the object of detecting any conservation in the order or spacing of homologous loci between the two genomes. A computer program—COMPAGEN—was developed to aid in the analysis of the data generated by this comparison. No map order conservation could be detected at the 15 kbp average resolution of this comparison between genomes estimated to have diverged 600 million years ago. A second comparison was performed between the chromosomes of *Haloferax volcanii* DS2 and *Haloferax mediterranei* ATCC 33500 (R-4). Extensive conservation was found between these two genomes which diverged approximately 80 million years ago showing only three rearrangements: two inversions and a transposition. Conclusions drawn from an analysis of the comparisons include: 1) that higher resolution is required to deal with distantly related genomes, likely making use of sequence data, and 2) that it is important to compare genomes that have diverged at different times if one wishes to investigate the dynamics of genomic evolution within a phylogenetic group. The local context model was developed in an effort to explain the pattern of conservation and divergence seen in these and other prokaryotic genome comparisons. This model states that since the expression of genes is affected by flanking genetic elements, genes will resist changing their position relative to one another so long as this change is likely to alter gene expression in a way

deleterious to the cell. The local context model thus provides a force promoting the conservation of genomic map order. The implications of this model for the evolution of the haloarchaea is discussed and future directions of prokaryotic genomics in general is explored.

RÉSUMÉ

La génomique est un champ de recherche qui se développe rapidement et cherche à étudier la structure, la fonction et l'évolution du génome d'un organisme. Des investigations génomiques ont été menées chez trois espèces de haloarchaea, un groupe monophylétique de procaryotes appartenant au règne Euryarchaeota du domaîne Archaea qui sont adapté à des environnements élevés en sel. Une carte physique et génétique du génome de *Halobacterium salinarum* GRB est décrite. Cette carte et la précédente publiée du génome de *Haloferax volcanii* DS2 ont été comparées dans le but de dépister une certaine conservation dans l'ordre ou dans la distance entre les loci homologues entre deux génomes. Un programme informatique -COMPAGEN- a été développé pour faciliter l'analyse des renseignements produits lors de cette comparaison. Aucune conservation dans l'ordre des cartes n'a pu être dépistée à une résolution moyenne de 15 kbp lors de cette comparaison entre ces génomes estimant avoir divergé il y a 600 millions d'années. Une seconde comparaison a été effectuée entre les chromosomes de *Haloferax volcanii* DS2 et *Haloferax mediterranei* ATCC 33500 (R-4). Une conservation étendue a été observée entre ces deux génomes ayant divergés il y a approximativement 80 millions d'années démontrant seulement trois réarrangements: deux inversions et une transposition. Les conclusions tirées de l'analyse de comparaisons comprennent: 1) qu'une résolution supérieure est requise lorsque des génomes de parentés éloignées sont étudiés, probablement par l'utilisation des renseignements des séquences, et 2) qu'il est important de comparer des génomes ayant divergés à différents temps si l'on veut enquêter sur la dynamique de l'évolution génomique à l'intérieur d'un groupe phylogénétique. Le modèle du contexte local a été développé dans le but d'expliquer les motifs de conservation et de divergence observés et autres comparaisons de génomes procaryotiques. Ce modèle

énonce que puisque l'expression des gènes est affectée par les éléments génétiques flanquant, les gènes vont résister à changer leur position relative à une autre puisque ce changement implique probablement de modifier l'expression des gènes de façon nuisible à la cellule. Le modèle du contexte local fourni alors une force soutenant la conservation de l'ordre dans une carte génomique. Les implications de ce modèle dans l'évolution des haloarchaea sont discutées et les directions futures de la génomique des procaryotes en général sont explorées.

ACKNOWLEDGEMENTS

Chapter 5 was made possible through a collaboration with Purificación López-Garcia and Ricardo Amils. The contribution of the program DERANGE II by David Sankoff greatly facilitated the analysis in chapter 6 and it is much appreciated. I must also thank Mathieu Blanchette for his patience in answering all my questions about the program and his willingness to modify it at my every request. Evan Weiher provided insightful discussion and guidance in the statistical analyses performed in chapter 6. Explaining statistics to a non-statistician is always a challenge and I appreciate the effort. Thanks also go to Anick De Moors for translating this poor monoglot's abstract into a résumé.

I would like to thank my advisory committee—L. Bonen, G. Drouin, and C. Wyndham—for their input throughout the course of my thesis. And, of course, I would like to thank my supervisor, Robert L. Charlebois. After four and a half years, I still think I made the right choice. Good job.

Last but not least, as the old saying goes, to all those people I have shared my life with here in Ottawa, both in and out of the lab, thank you for making every day an adventure.

The work described in this thesis was supported by the Natural Sciences and Engineering Research Council of Canada. Work performed in Madrid (portions of chapter 5) was supported by the Spanish Interministerial Commission for Science and Technology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ACR	**Ancient Conserved Region**
AP-PCR	**Arbitrarily Primed-Polymerase Chain Reaction**
ATCC	**American Type Culture Collection**
DLL	**Dynamic Link Library**
EMBL	**European Molecular Biology Laboratory**
EtBr	**Ethidium Bromide**
NCBI	**National Center for Biotechnology Information**
ODBC	**Open DataBase Connectivity**
ORF	**Open Reading Frame**
PFGE	**Pulsed-Field Gel Electrophoresis**
RAPD	**Random Amplified Polymorphic DNA**
REA	**Restriction Endonuclease Analysis**
REP	**Repetitive Extragenic Palindromic**
STL	**Standard Template Library**
TIGR	**The Institute for Genomic Research**

"Forty-two!", yelled Loonquawl. "Is that all you've got to show for seven and a half million years' work?"

"I checked it very thoroughly," said the computer, " and that quite definitely is the answer. I think the problem, to be quite honest with you, is that you've never actually known what the question is."

"But it was the Great Question! The Ultimate Question of Life, the Universe and Everything," howled Loonquawl.

"Yes," said Deep Thought with the air of one who suffers fools gladly, "but what actually *is* it?"

A slow stupefied silence crept over the men as they stared at the computer and then at each other.

"Well, you know, it's just Everything...everything..." offered Phouchg weakly.

"Exactly!" said Deep Thought. "So once you do know what the question actually is, you'll know what the answer means."

Douglas Adams
The Hitchhiker's Guide to the Galaxy

CHAPTER 1

Introduction

Molecular Biology and the Study of Genomes

There is no denying the impact molecular biology has had on our understanding of living things. This is due in part to the staggering range of applications to which it can be put. Molecular biology has contributed to the elucidation of everything from the mechanisms of viral infection of host cells and the distribution of antibiotic resistance determinants in bacterial populations, to the regulation of transcription and translation and the establishment of a phylogeny for the prokaryotic world. Another reason is that molecular biology is all about the direct manipulation of an organism's genetic material. Through recombinant DNA technologies, human beings are beginning to control the genetic future of organisms at a rate heretofore unheard of although the merit of this particular ability is still subject to vigorous debate. Combined with other disciplines such as biochemistry, cell biology, ecology, and population genetics, our ability to investigate the living world is enhanced even further; the potential of which has not yet been fully realized.

In most of its applications, molecular biology lends itself to the reductionist approach. The subject of investigation may be a gene, an operon, the distribution in a genome of a particular repeated DNA element, or the genes affected by a single regulatory signal. A genome is taken apart, as it were, and only a small bit is dealt with at any one time. The reductionist approach is an indispensable staple of the scientific method and a great deal can and is learned using it. However, the study of large, complex systems also

requires a broader view for their investigation to be complete. This is what the field of genomics aspires to address. As an outgrowth of molecular and cell biology, genomics uses the same techniques as these disciplines but with the object of investigating genomes as functional units. Whereas reductionist approaches are concerned with effects seen at the gene or operon level, genomics is concerned with how the forces that affect the functioning and evolution of these small genetic units combine to constrain and provide opportunities for the structure and functioning of the genome as a whole.

This thesis is concerned with the study of genomics as it applies to prokaryotic organisms. Here, genomics includes the most obvious activities of fingerprinting, mapping and sequencing of entire or major fractions of genomes. Also included are large scale gene expression studies, protein inventories, and computational analyses of map and sequence data such as the investigation of recombinational activity and the elucidation of gene families. Studies into nucleoid structure and organization and how these factors might affect gene expression complement the mainly sequence based investigations and round out genomics as a self-sustaining line of inquiry. This broad definition of genomics is necessary to encompass the two subjects of this thesis which are; 1) genomic organization and patterns of change within that group of prokaryotes known as the haloarchaea and 2) the development of the supercoil model of map stability, or more concisely the local context model, that seeks to explain in part the pattern of conserved and rearranged genome maps observed for prokaryotes generally.

As one might expect, the motivation for the work described in this thesis also comes in two parts. Certain members of the haloarchaea are known to be genetically unstable due to the presence of numerous insertion sequences distributed within their genomes, a topic which is more thoroughly discussed in the next section. The genomic instability of these

haloarchaea was inferred from the observed genetic instability and the instability of plasmid DNA though no direct evidence existed concerning chromosomes. The objectives of the genomic map described in chapter 3 and the two comparisons in chapters 5 and 6 were to further our understanding of the genomic organization of members of the haloarchaea and to elucidate patterns of genomic level change. Patterns of conservation or divergence found by these comparisons could then be used to draw conclusions about the evolutionary history of the haloarchaea as a whole.

The local context model was conceived in the course of the comparison work. It sprang from the observation that many comparisons between the chromosomal maps of different species or even genera of prokaryotes showed high degrees of conservation in the order of homologous genes with no adequate explanation given despite the existence of forces known to rearrange gene order. The local context model explains this observation by proposing that a force does exist that resists altering the position of a gene relative to its surroundings and that this force is the sensitivity of gene expression to changes in local context, principally DNA topology. As few theoretical models have yet been applied to comparative genomic data, this model provides a much needed framework within which to test hypotheses about the tempo and mode of prokaryotic genome evolution. Support for the model comes almost exclusively from the Bacteria since so much more work has been done on this group compared to the Archaea. This necessitates due caution when applying the model to the haloarchaeal genomic comparisons although, regardless of differences in specific mechanisms between the two groups, implications for the broad patterns of genomic change predicted by the model should still apply.

How this Thesis is Organized

Much of the work described in this thesis has been previously published. Each published article is represented in its own chapter and changed only to provide a consistent format although all references have been grouped together at the end of the thesis to minimize duplication. Because these chapters have been published over the course of three years, updates have been included in the chapters where appropriate. The complete author list is also given at the beginning of chapters representing published articles and my contribution to each is explicitly stated. The published work includes chapters 3 through 6 first describing the physical and genetic map of *Hb. salinarum* GRB (then known as *Halobacterium* sp. GRB), moving on to an explanation of the local context model of map stability in prokaryotes, and finishing with two comparison studies between haloarchaeal genomes. The order of these chapters is deliberate in that the later chapters make use of the results (the *Hb. salinarum* GRB map) or the concepts (the local context model) described in the previous chapters. To these four chapters have been added chapter 7 describing a database program used in the analysis found in chapter 6, a general introduction reviewing the state of the art in prokaryotic genomics (chapter 1), and a general discussion (chapter 8) bringing together the concepts and conclusions of the other chapters and making some suggestions about genomics as it is pursued today and its as yet unrealized potential. In some instances, direct reference is made in the text to one of the chapters representing a published work. Where this occurs, the normal reference is given followed by the chapter in square brackets as in the following example (Charlebois and St. Jean, 1995 [chapter 4]).

As a consequence of including published works essentially unaltered in this thesis, the materials and methods are scattered through four chapters and are intermingled with

techniques and procedures not actually conducted by me. To overcome this situation, a separate materials and methods section is included as chapter 2 that lists all the techniques, and only those techniques, executed by me. Specific protocols and recipes are included in Appendix A for ease of reference.

The Haloarchaea

Historical Perspective

This century has proved an eventful time for those who study prokaryotes. For much of this time, microbiology could justifiably be called the discipline that evolution forgot. The reason for this must surely be found in the difficulty inherent in performing analyses on organisms with almost no morphological characters or fossil record. The techniques available in the first half of this century allowed very little to be accomplished in the area of prokaryote classification. Although many physiological and biochemical characteristics were available for use in classification, there was no way to determine the polarity of a series of characters or to distinguish between primitively homologous and convergent systems (Kandler, 1985). The 1950's saw the introduction of numerical taxonomic methods and their application to the enormous biochemical diversity of prokaryotes. While the new techniques allowed a proliferation in the numbers of described species and genera, much less progress was made in ordering these diverse organisms into larger taxonomic units (Kandler, 1985). Electron microscopy was used to study prokaryotes from the late 1950's together with biochemistry, eventually leading to the proposal of Procaryotae as a formal kingdom with major divisions based on cell wall structure (Murray, 1968). Even so, through the 1960's and into the 1970's, prokaryotic phylogenetics was stagnant. This effectively prevented most evolutionary considerations from being applied to the study of

prokaryotes and severely limited the types of investigations that could be performed. Prokaryotes consisted of a multitude of species and genera of uncertain affiliation and the most that could be said for them was that they were more primitive than their likely descendants, the eukaryotes.

All this changed in the mid to late 1970's when molecular techniques were first applied to the problem of prokaryotic phylogeny. Suddenly, almost overnight, what had been denied microbiologists for so long was within their grasp; a testable, reliable phylogeny of the prokaryotic world. Finally, microbiology could apply evolutionary theory to the study of its subject and gain the foundation and framework it had lacked for so long. This revolution in microbiology as a field of scientific investigation, its immediate outcome, and the years leading up to it are admirably recounted in Kandler (1985) and Woese (1987).

Perhaps the most startling discovery to come out of the early molecular work was the realization that prokaryotes could be divided into two lineages, lineages as distinct from each other as either were from eukaryotes. In 1977 it was revealed that some prokaryotes belonging to a peculiar phenotypic group—methanogens—were quite different from most others according to rRNA oligonucleotide cataloging (Balch et al., 1977; Fox et al., 1977; Woese and Fox, 1977). This difference was great enough to suggest that these prokaryotes should not be grouped with the mainstream bacteria but separated into their own group. The following year saw examples of two other phenotypes join the ranks of the methanogens: the extreme halophile *Halobacterium* (Magrum et al., 1978) and the thermophiles *Sulfolobus* and *Thermoplasma* (Woese et al., 1978). This second group, separate from the bacteria, were called the archaebacteria in reference to the presumed primitive phenotypes exhibited by these organisms, while the remaining

prokaryotes were termed eubacteria (Woese and Fox, 1977; Woese et al., 1978). Although such a radical departure from the deeply held belief in the great dichotomy between prokaryotes and eukaryotes was not accepted unanimously (Steitz, 1978; Van Valen and Maiorana, 1980), subsequent work only confirmed the earlier findings (Woese and Gupta, 1981; Zillig et al., 1982; Woese et al., 1984). By the time Woese et al. (1990) proposed a new system for categorizing living things into three 'domains'—Bacteria (eubacteria), Archaea (archaebacteria), and Eucarya (eukaryotes)—at a taxonomic level above kingdom, the archaebacteria had become generally accepted as a distinct group (Doolittle and Daniels, 1985; Stackebrandt, 1985).

Since their first description as an independent group, evidence for the unique nature of the Archaea has grown. Archaea seem to be mosaics of prokaryotic and eukaryotic characters with a few characters that are uniquely their own thrown in for good measure. A partial list of these characters includes: i) 70S ribosomes possessing a mosaic of bacterial and eucaryal structural features as well as unique features (Stöffler and Stöffler-Meilicke, 1986; Kimura et al., 1989); ii) use of methionine rather than *N*-formyl-methionine in initiation of translation (Bayley and Morton, 1978); iii) a pattern of sensitivity and resistance to antibiotics that follows neither the bacterial or eukaryotic models (Elhardt and Böck, 1982; Altamura et al., 1988; Sanz et al., 1992); iv) a multisubunit RNA polymerase resembling that of eukaryotes (Langer et al., 1995); v) transcription initiation using homologs of eukaryotic TATA-binding proteins (Marsh et al., 1994; Rowlands et al., 1994); vi) tRNAs showing patterns of modified bases unlike Bacteria and Eucarya (McCloskey, 1986); vii) elongation factors susceptible to diphtheria toxin (Kessel and Klink, 1980); and viii) the use of ether-linked isoprenoid lipids in their membranes rather than ester-linked fatty-acid lipids (Langworthy, 1985). In addition to the

three main phenotypic groups described initially, evidence for the Archaea in pelagic marine environments has been discovered. This evidence comes in the form of sequences recovered by PCR from seawater samples using primers from regions of rDNA containing sequences characteristic of known Archaea. Archaeal sequences have been recovered from waters of the east and west coasts of North America (DeLong, 1992; Fuhrman et al., 1992), and coastal Antarctic waters (DeLong et al., 1994). Although no Archaea have been cultured from these environments as yet, the amount of recovered DNA found in these studies suggests that Archaea may form a significant proportion of the populations, as much as 34% in Antarctic waters (DeLong et al., 1994).

Although it has been argued that the Archaea are a diphyletic group on the basis of ribosome morphology and sequence studies (Lake et al., 1985; Lake, 1989; Rivera and Lake, 1996), most evidence points to the generally held belief of a monophyletic origin for the Archaea (Garrett et al., 1994). Iwabe et al. (1989) and Gogarten et al. (1989) were able to use anciently duplicated genes (elongation factors Tu and G and ATPase subunits α and β) to elucidate the root of the three domains and found that the Archaea are more closely related to Eucarya than to Bacteria. This finding too was disputed principally on the basis of the particular genes used for the test (Forterre et al., 1993). The results of Iwabe et al. (1989) and Gogarten et al. (1989) have since been supported by investigations using aminoacyl-tRNA synthetases (Brown and Doolittle, 1995). For now at least, the Archaea can be considered a monophyletic group outwardly resembling Bacteria but in fact more closely related to the eukaryotic lineage.

Phylogeny

The haloarchaea represent a monophyletic group within the domain Archaea consisting of prokaryotes that require high levels of salt in the environment (1.5 M-5.2 M) for survival and growth. They are strict aerobes or facultative anaerobes and can be found in such environments as the Dead Sea, salt lakes and salt flats, salted meat, certain brands of Soya and fish sauce and even within salt crystals extracted from salt mines (Norton, 1992). Additionally, two genera—*Natronobacterium* and *Natronococcus*—require basic pH for survival and are found in soda lakes (Tindall et al., 1984). They belong to the kingdom Euryarchaeota within the Archaea which also includes some thermophiles and marine species, and the methanogens (Fig. 1.1). The methanogens are the sister group of the haloarchaea (Woese, 1987) some of which are thermophilic (Jones et al., 1983; Huber et al., 1989), halophilic (Boone et al., 1986; Zhilina, 1986), or both. Fig. 1.1 also shows the second kingdom of the Archaea, the Crenarchaeota, which includes most sulfur-dependent extreme thermophiles and some marine species.

Until recently, there were seven recognized genera within the haloarchaea which are listed in Table 1.1. Two studies have independently attempted to reorganize one of these genera, *Halobacterium* (Kamekura and Dyall-Smith, 1995; McGenity and Grant, 1995) using evidence from 16S rRNA sequence data. This genus was one of the first to be described for the haloarchaea and has historically been the dumping ground for many species. Both groups find *Halobacterium* to be polyphyletic and recommend that it be divided into three separate lineages bringing the total number of genera up to ten according to Kamekura and Dyall-Smith and eight in McGenity and Grant. Fig. 1.2 reproduces the trees from these two studies indicating the proposed new genera. The trees from the two studies are consistent with one another and the difference in the number of

Fig. 1.1. The tree of life showing the domains Bacteria, Archaea and Eucarya and the phylogeny of the Archaea. The tree of life was rooted using phylogenies of duplicated genes found in all three domains. The phylogeny of the Archaea is based on 16S rDNA sequences. The specific branching order is taken from DeLong et al., 1994). The kingdom Crenarchaeota is made up of sulfur-dependent extreme thermophiles and marine species. The Euryarchaeota include extreme thermophiles, methanogens, extreme halophiles and marine species. Marine archaea are known from amplified rDNA sequences only and have not been cultured. West-coast archaeal sequences were sampled off the coast of Santa Barbara while East-coast sequences were sampled from Woods Hole (DeLong, 1992). For a more detailed phylogeny of the haloarchaea, see Fig. 1.2.

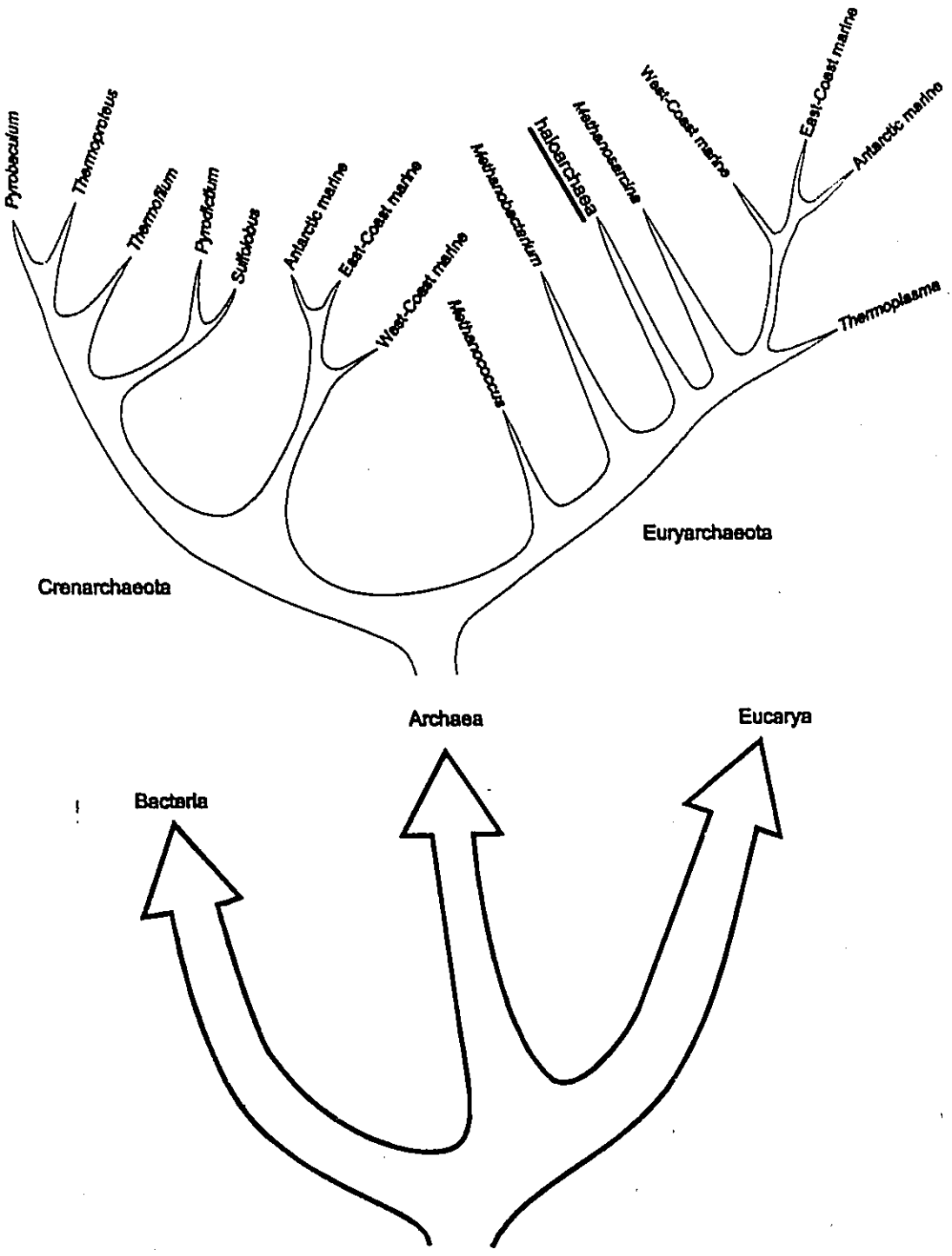
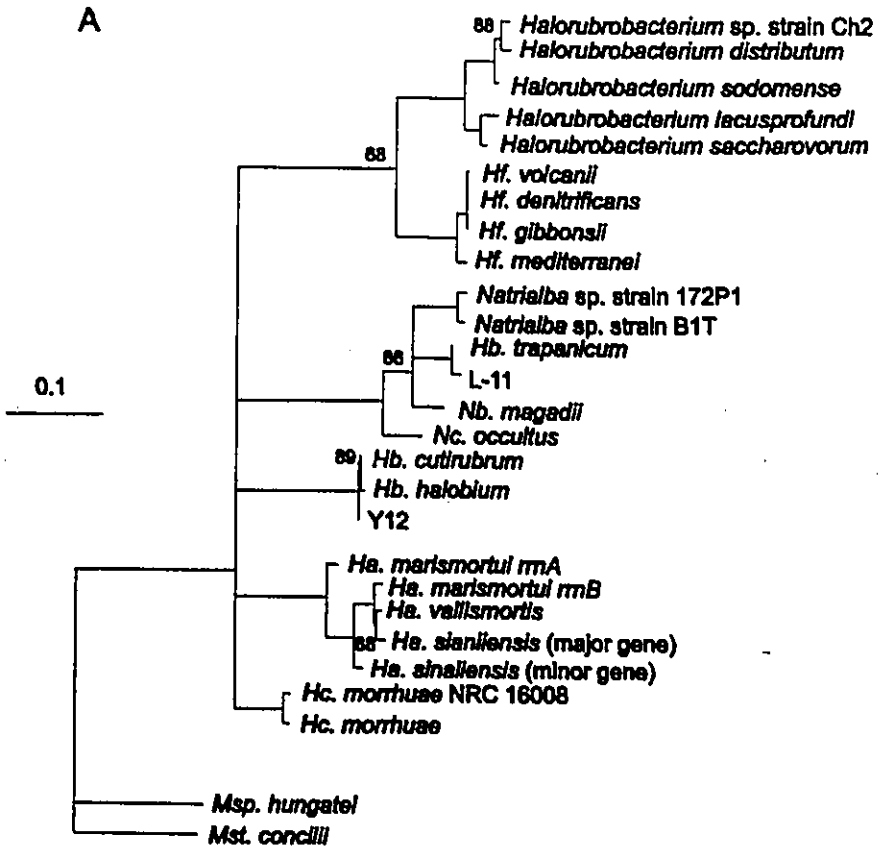

Pyrobaculum
Thermoproteus
Thermofilum
Pyrodictium
Sulfolobus
Antarctic marine
East-Coast marine
West-Coast marine
Methanococcus
Methanobacterium
haloarchaea
Methanosarcina
West-Coast marine
East-Coast marine
Antarctic marine
Thermoplasma
Crenarchaeota
Euryarchaeota
Archaea
Eucarya
Bacteria

Table 1.1. Currently recognized genera belonging to the haloarchaea and proposed changes to the genus *Halobacterium*.

Recognized Genera and Species	First Use of Genus Name	Kamekura and Dyall-Smith (1995)	McGenity and Grant (1995)
Haloarcula	Torreblanca et al., 1986		
Halobacterium salinarum	Elazari-Volcani, 1957	*Halobacterium*	*Halobacterium*
Halobacterium lacusprofundi *saccharovorum* *sodomense* *distributum* *trapanicum* NRC 34021	Elazari-Volcani, 1957	*Halorubrobacterium* N/A[a]	*Halorubrum* N/A[a]
Halobacterium trapanicum NCIMB 767 unnamed strain L-11	Elazari-Volcani, 1957	unnamed genus	
Halobaculum	Oren et al., 1995		
Halococcus	Schoop, 1935		
Haloferax	Torreblanca et al., 1986		
Natronobacterium	Tindall et al., 1984		
Natronococcus	Tindall et al., 1984		
unnamed strains B1T and 172P1		*Natrialba*	

[a] not applicable (not included in this study)

Fig. 1.2. Two alternative proposals for the restructuring of the genus *Halobacterium* and the phylogeny of the haloarchaea. Trees are derived from 16S rDNA sequences using distance-matrix methods. Numbers at nodes are bootstrap values expressed as percentages. Values of 100 are not shown. In A, instances where nodes had bootstrap values of less than 80%, the branches have been collapsed. Scale bars indicate 0.1 substitutions per site. The trees shown in A and B are consistent in their branching orders where this can be determined and differ only in the proposed genus names, *Halorubrum*, *Halorubrobacterium* and *Natrialba*. Species allocated to *Halorubrum* and *Halorubrobacterium* currently belong to *Halobacterium* while the two *Natrialba* strains in A have not previously been assigned to species. Genus names are as follows; Ha.) *Haloarcula*, Hb.) *Halobacterium*, Hc.) *Halococcus*, Hf.) *Haloferax*, Msp.) *Methanospirillum*, Mst.) *Methanosaeta*, Nb.) *Natronobacterium*, Nc.) *Natronococcus*. Note that *Hb. cutirubrum*, *Hb. halobium* and *Hb. salinarum* are considered synonyms for the same species. Trees are modified from; A) Kamekura and Dyall-Smith, 1995 and B) McGenity and Grant, 1995. According to the rules of priority, the genus names of Kamekura and Dyall-Smith have precedence.


A
88
Halorubrobacterium sp. strain Ch2
Halorubrobacterium distributum
Halorubrobacterium sodomense
Halorubrobacterium lacusprofundi
Halorubrobacterium saccharovorum
88
Hf. volcanii
Hf. denitrificans
Hf. gibbonsii
Hf. mediterranei
Natrialba sp. strain 172P1
Natrialba sp. strain B1T
Hb. trapanicum
88
L-11
0.1
Nb. magadii
Nc. occultus
89
Hb. cutirubrum
Hb. halobium
Y12
Ha. marismortui rrnA
Ha. marismortui rrnB
Ha. vallismortis
Ha. sianiensis (major gene)
Ha. sinaliensis (minor gene)
Hc. morrhuae NRC 16008
Hc. morrhuae
Msp. hungatei
Mst. conciili
B
Nc. occultus
172P1
62
Hb. trapanicum NCIMB 767
88
Nb. magadii
72
Halorubrum lacusprofundi
Halorubrum saccharovorum
Halorubrum trapanicum NRC 34021
Halorubrum sodomense
52
Hf. mediterranei
Hf. denitrificans
52
39
Hf. volcanii
96
Hf. gobbinsii
Hb. salinarum
Hc. morrhuae NRC 16008
Hc. morrhuae ATCC 17082
0.1
Ha. sinaliensis (minor gene)
Ha. sinaliensis (major gene)
99
Ha. marismortui rrnB
Ha. marismortui rrnA
Msp. hungatei

genera is a result of the more aggressive renaming scheme of Kamekura and Dyall-Smith (1995). Where the two naming schemes conflict—*Halorubrobacterium* (Kamekura and Dyall-Smith, 1995) and *Halorubrum* (McGenity and Grant, 1995)—*Halorubrobacterium* should be used if the new phylogeny is accepted since it appeared in print before *Halorubrum*. A feature of both trees to note is the ambiguous branching order between most of the genera. Kamekura and Dyall-Smith made no attempt to resolve the branching order between five of the genera while these same genera are separated by very short branch lengths with poor bootstrap values in the tree of McGenity and Grant. This suggests that these genera diverged from one another in a relatively short period of time apparently near the base of the haloarchaeal lineage.

Of particular concern is that the species *Halobacterium salinarium* remains within *Halobacterium*, and in fact seems to make up the only species remaining in this genus. A strain of this species, GRB, is the organism whose genomic map is presented in chapter 3 and it is also one of the genomes compared in chapter 6. As if the conflicting genus names were not confusing enough, a new species name for *Halobacterium salinarium* has recently been adopted (Ventosa and Oren, 1996). The specific epithet *salinarium* has been replaced by *salinarum* on the grounds that *salinarium* is grammatically incorrect. Ventosa and Oren (1996) also reiterates that the strains commonly referred to as *Halobacterium cutirubrum* and *Halobacterium halobium* show no properties to differentiate them from *salinarum* and should be considered strains of *Halobacterium salinarum*. The practical upshot of all this name calling is that *Halobacterium salinarium*, *Halobacterium halobium* and *Halobacterium cutirubrum* should all be considered synonyms of *Halobacterium salinarum*.

The three haloarchaeal species used in the comparisons, *Halobacterium salinarum*, *Haloferax mediterranei*, and *Haloferax volcanii*, possess a suite of similar and divergent traits. The most pervasive common theme between the three is, of course, their mechanism of adaptation to high salt environments. Like all described haloarchaea, all three maintain high levels of intracellular K^+ ions to balance the osmotic pressure of their medium (Grant and Larsen, 1989). All three are also chemoorganotrophic (Mullakhanbhai and Larsen, 1975; Rodriguez-Valera, 1983; Larsen and Grant, 1989). Differences include the ability of the two *Haloferax* species to grow on chemically defined medium while *Hb. salinarum* cannot and the ability of *Halobacterium* to grow anaerobically via a fermentative pathway while *Haloferax* is strictly aerobic (Torreblanca et al., 1986; Larsen and Grant, 1989). *Hb. salinarum* also possesses bacteriorhodopsin allowing it to use light energy to produce proton motive force (Larsen and Grant, 1989). However, it seems that this mechanism cannot supply all the energy needs of the cell and rather supplements other sources of energy in times of short supply or at low oxygen tensions (Grant and Ross, 1986). Another difference between the two species of *Haloferax* and *Hb. salinarum* are the specific proportions of salts each needs for growth. *Haloferax* species require a much higher magnesium content for optimum growth than does *Halobacterium* (Torreblanca et al., 1986). Likewise, *Halobacterium* grows well at NaCl concentrations that are nearly double the optimum for *Haloferax* (Torreblanca et al., 1986). Variations in the proportions of salts needed for optimum growth is quite common among haloarchaea, however. In this case, there exists enough overlap in the salt requirements of the three species that all can grow on a common medium (Mevarech and Werczberger, 1985), a phenomenon that extends to the genus *Haloarcula*.

While the ecological role of haloarchaea has been and is being investigated (Grant and Ross, 1986; Oren, 1991; Norton, 1992), a lack of information for the three species that are the subjects of this thesis means that a detailed comparison between their life histories is impossible. Given the common mechanism of adaptation to high salt and their similar energy metabolism, however, one may infer broadly similar niches for all three species while numerous differences in detail undoubtedly exist.

The Haloarchaea as Subjects for Genomic Studies

The haloarchaea have a reputation for genetic instability (Pfeifer et al., 1981b; Pfeifer and Blaseio, 1990). This reputation is mainly due to the extensive work done with some strains of *Hb. salinarum*. Some of the earliest described isolates belong to this species and have been of great interest to researchers because of their possession of bacteriorhodopsin, the light-driven proton pump unique to this genus. Some of these same isolates also harbour insertion sequences in great quantities, often numbering in the hundreds, belonging to over a dozen families (Sapienza and Doolittle, 1982). The activity of these insertion sequences is also easily identified due to their inactivation of pigment and gas vacuole genes, both of which have visible effects on colony morphology. In some strains, formerly identified as *Hb. halobium*, the frequency of inactivation of specific genes has been estimated to be as high as 10^{-2} per cell per generation (Weidinger et al., 1979; Pfeifer et al., 1981b). *Halobacterium salinarum* is not alone in possessing insertion sequences; *Haloferax volcanii* harbours them as well with at least one family being homologous between the two species (Pfeifer and Blaseio, 1990). While the insertion sequences aren't as numerous or as active in this genus, they still impact on the genome. A link between genetic instability and genomic instability has long been thought to exist for

the haloarchaea (Sapienza et al., 1982). The 150 kbp plasmid pHH1 found in certain strains of *Hb. salinarum* was observed to undergo constant rearrangements to its map such that obtaining a culture of cells all containing identical versions of the plasmid was deemed impossible (Pfeifer et al., 1989). It is easy to extrapolate the observed instability to the chromosome despite the lack of direct evidence to support this assumption. The comparative genomic work described here and elsewhere (López-García et al., 1993; Hackett et al., 1994) is helping to determine if such an assumption is justified.

Types of Genomic Studies

Advances in science are often driven by technology. This seems especially true for the field of genomics which relies heavily on technology to allow huge datasets to be acquired and manipulated. Computers are an obvious component but two key technologies that have contributed more specifically to genomics are pulsed-field gel electrophoresis (PFGE) and automated DNA sequencing. PFGE has allowed for the genomic characterization, through DNA fingerprinting and mapping, of a plethora of prokaryotic species, many of which lack the tools necessary for classical genetic studies. Researchers are no longer bound to a few, genetically amenable species but now have a virtual free reign to study whatever organism they like. This phylogenetic freedom has manifested itself in the explosion of physical maps now numbering well over one hundred that have been constructed mostly since 1990. Complementing PFGE is automated DNA sequencing which provides much finer genomic resolution—down to one nucleotide—with a concomitant increase in the amount of money, equipment and time involved. Even so, automated sequencing has made whole genome sequencing a reality and researchers have been eager to exploit this utterly new and exciting resource. As with all new

phenomena, however, it will be some time before the full potential of complete genome sequences is realized and exploited.

Genome Organization and Genomic Fingerprinting

The most basic level of genomic inquiry deals with the way the genetic material is divided into replicons in the cell and the physical structure of the DNA molecules involved. Exceptions to the long-held paradigm of one circular chromosome which may or may not be supplemented by one or more circular plasmids are now known to exist. Species of Bacteria with two unique chromosomes, linear chromosomes and linear plasmids (see Campbell, 1993; Hinnebusch and Tilly, 1993; Prozorov, 1995 for reviews) have been found often in the course of mapping their genomes. Such revelations provide little information in and of themselves other than demonstrating once again the pervasive diversity found within the Bacteria. They do, however, attract attention to such questions as how linear DNA remains stable within some bacteria, what role recombination between non-homologous chromosomes might play in their evolution, and whether nucleoid structure is significantly different for linear chromosomes as opposed to circular ones. Little work has so far been done to answer these and other questions with some notable exceptions (investigations into the terminal inverted repeats of the *Streptomyces ambofaciens* linear chromosome for example [Leblond et al., 1991; Leblond et al., 1996]). The reason for this is almost certainly the need for genetic tools such as transformation which, as outlined above, often do not exist for the subject organism. The result is that there will often be a lag between preliminary, often superficial studies, and more in-depth analyses designed to answer the questions the previous study raised. This pattern of investigation is apparent in many aspects of genomics today.

One area of genomics where depth of investigation is not an issue but usability, reproducibility, and speed are, is in the identification of bacterial strains in a clinical setting. Often called DNA fingerprinting, the use of molecular methods for this purpose has increased dramatically in recent years due to the invention of PFGE and PCR, Table 1.2. These methods, along with conventional agarose gel electrophoresis, allow for very rapid elucidation of the relationships between large numbers of isolates with low demands in materials and expertise, all important considerations in many clinical settings.

In all cases, fingerprinting is used to determine the degree of similarity between genomes, often involving organisms at the species or strain level. Electrophoretic technologies, either conventional agarose gels or PFGE, are used in most fingerprinting studies. PCR is also sometimes used in a protocol termed Random Amplified Polymorphic DNA (RAPD) analysis—also known as Arbitrarily-Primed PCR (AP-PCR)—which uses a number of arbitrary PCR primers to produce a pattern of PCR products on an agarose gel (Welsh and MacLeland, 1990; Williams et al., 1990). In both cases, it is the pattern of bands on a gel that are used to infer the phylogenetic relationship and distances between the strains under study. It is perhaps best that fairly closely related organisms are tested this way since it is known that restriction patterns (and by analogy, primer binding sites) can diverge significantly between two genomes while the genetic maps maintain a high degree of conservation (López-García et al., 1995 [chapter 5]). This could result in an overestimation in the differences separating two organisms, a serious problem when pathogens are involved.

The importance of molecular methods in DNA fingerprinting can be gauged from a recent paper setting out clear criteria for the classification of strains investigated using PFGE (Tenover et al., 1995). In it, the authors present a scale of relatedness ranging from

Table 1.2. Prokaryotes studied by DNA fingerprinting.

Organism	Fingerprinting Method	Reference
Brevibacterium lactofermentum *Brevibacterium linens* *Corynebacterium glutamicum*	PFGE	Correia et al., 1994
Chlorobium limicola	PFGE	Méndez-Alvarez et al., 1995
Enterobacter agglomerans	PFGE	Evguenieva-Hackenberg and Selenska-Pobell, 1995
Gardnerella vaginalis	REA[a]	Wu et al., 1996
Haloarcula spp. *Halobacterium salinarum* *Halococcus* spp. *Haloferax* spp. *Natronobacterium gregoryi* *Natronococcus occultus*	RAPD[b]	Martínez-Murcia et al., 1995
Helicobacter pylori	RAPD[b]	Vandamme et al., 1995
Lactobacillus plantarum	RAPD[b] REA[a]	Johansson et al., 1995a Johansson et al., 1995b
Prophyromonas gingivalis	RAPD[b]	Ménard and Mouton, 1995
Pseudomonas spp.	PFGE	Grothues and Tümmler, 1991
Salmonella typhimurium	PFGE	Liu and Sanderson, 1995b
Spiroplasma citri	PFGE	Ye et al., 1995
Streptococcus group A	PFGE REA[a]	Upton et al., 1996
Streptococcus thermophilus	PFGE	Boutrou et al., 1995

[a] restriction endonuclease analysis (conventional gel electrophoresis)
[b] random amplified polymorphic DNA

indistinguishable to unrelated for use when typing strains and stipulate the exact number of differences on a pulsed-field gel that corresponds to each category. Such analytical rigor is sadly lacking in other areas of genomics (see 'Comparisons' section below and chapter 6). Tenover et al. (1995) also includes an extensive list of Bacteria whose genomes have been analysed by PFGE.

DNA Mapping

No aspect of prokaryotic genomics has been as aggressively pursued by researchers as the physical mapping of chromosomes. Over 100 maps now exist of prokaryotes including proteobacteria, cyanobacteria, Gram-positive bacteria, spirochetes, crenarchaeotes and euryarchaeotes. Recent reviews (Cole and Saint Girons, 1994; Römling and Tümmler, 1994; Fonstein and Haselkorn, 1995) extensively cover the various methods used to construct these maps which fall into two broad categories; top-down and bottom-up strategies.

By far the more popular strategy is top-down mapping. Top-down strategies have the advantages of being relatively quick and easy to do (an important consideration when one wants to compare a number of genomes) and being applicable to a wide variety of organisms since genetic tools are often not necessary. The most widely used method of top-down mapping is PFGE. Cole and Saint Girons (1994) provide an extensive list of maps constructed using PFGE while Table 1.3 lists maps not included in their review. Often 1-dimensional PFGE is enough to construct a physical map but sometimes additional methods are required. Table 1.3 lists a number of maps constructed using 2-dimensional PFGE. PFGE mapping relies on the existence of restriction enzymes that will cut a genome into a small number of pieces (ideally less than 20) that are evenly

Table 1.3. Physical and genetic maps of prokaryotic chromosomes.

Organism	Mapping Method	Reference
Bacillus sp. C-125	PFGE	Sutherland et al., 1993
Bartonella bacilliformis	PFGE	Krueger et al., 1995
Borrelia garinii 20047 *Borrelia afzelii* VS461	PFGE (2-dimensional)	Ojaimi et al., 1994
Borrelia sp. 21 strains	PFGE	Casjens et al., 1995
Burkholderia cepacia ATCC 25416	PFGE (2-dimensional)	Rodley et al., 1995
Campylobacter upsaliensis ATCC 43954	PFGE	Bourke et al., 1995
Clostridium perfringens CPN50	PFGE	Katayama et al., 1995
Erwinia cartovora subsp. *atroseptica* 3-2	genetic, transduction experiments	Nikolaichik and Pesnyakevich, 1995
Halobacterium salinarum GRB	ordered clone (cosmid) library	St. Jean et al., 1994
Halobacterium salinarum NRC-1 S9	PFGE (2-dimensional)	Hackett et al., 1994
Haloferax mediterranei M2a M2b M4 M10	PFGE	López-García et al., 1993
Haloferax volcanii DS2	ordered clone (cosmid) library	Charlebois et al., 1991
Lactococcus lactis subsp. *cremoris* MG1363	PFGE	Le Bourgeois et al., 1995
Mycobacterium leprae	ordered clone (cosmid) library	Eiglmeier et al., 1993
Mycoplasma gallisepticum R ATCC 19610	PFGE	Tigges and Minion, 1994
Mycoplasma genitalium G37 (ATCC 33530)	ordered clone (cosmid) library	Lucier et al., 1994
Mycoplasma pneumoniae M129 B18	ordered clone (cosmid) library	Wenzel et al., 1992
Neisseria meningitidis B1940	PFGE (1- and 2-dimensional)	Gäher et al., 1996
Neisseria meningitidis Z2491	PFGE	Dempsey et al., 1995
Phytoplasma, Western X-Disease	PFGE	Firrao et al., 1996
Planctomyces limnophilus DSM 3776[T]	PFGE	Ward-Rainey, 1996
Pseudomonas aeruginosa C	PFGE (1- and 2-dimensional)	Schmidt et al., 1996
Pseudomonas fluorescens SBW25	PFGE (1- and 2-dimensional)	Rainey and Bailey, 1996
Rhizobium meliloti 1021	PFGE	Honeycutt et al., 1993
Rhodobacter capsulatus SB1003	ordered clone (cosmid) library	Fonstein et al., 1995

Table 1.3. Continued.

Organism	Mapping Method	Reference
Salmonella enteritidis SSU7998	PFGE	Liu et al., 1993c
Salmonella paratyphi A (ATCC 9150)	PFGE	Liu and Sanderson, 1995d
Salmonella paratyphi B	PFGE	Liu et al., 1994
Salmonella typhi Ty2	PFGE	Liu and Sanderson, 1995c
Serpulina hyodysenteriae B78^T	PFGE	Zuerner and Stanton, 1994
Spiroplasma melliferum BC-3	PFGE	Ye et al., 1994
Streptococcus thermophilus A054	PFGE	Roussel et al., 1994
Streptomyces ambofaciens DSM 40697 ATCC 15154 ETH 9427	PFGE	Leblond et al., 1996
Streptomyces coelicolor A3(2) M145	ordered clone (cosmid) library	Redenbach et al., 1996
Streptomyces griseus IFO3237	PFGE	Lezhava et al., 1995
Thermus thermophilus HB27	PFGE	Tabata et al., 1993 Tabata and Hoshino, 1996
Vibrio cholerae 569B	PFGE	Majumder et al., 1996

distributed over a specific size range so that they can be resolved on a gel. If the available restriction enzymes cut a genome into too many pieces or those pieces migrate together in a gel, additional resolving power may be needed in the form of 2-dimensional pulsed-field gels (Römling and Tümmler, 1991). In some cases, even this is not enough. Some haloarchaea have a mosaic structure to their genomes meaning that some regions have very different restriction enzyme cutting patterns than others (Pfeifer and Betlach, 1985; Charlebois et al., 1989). This can make the construction of an unambiguous map impossible using PFGE exclusively. It was necessary to confirm the maps for *Hb. salinarum* strains NRC-1 and S9 (Hackett et al., 1994), for example, using the bottom-up map of *Hb. salinarum* strain GRB (St. Jean et al., 1994; [chapter 3]).

The lack of suitable restriction enzymes for mapping purposes can be overcome by artificially inserting recognition sites into a genome. This is done using transposons that have been modified to carry the recognition sequence of a specific restriction enzyme. Recognition sequences for this enzyme may or may not occur naturally in the target genome although analysis of the results is aided in certain applications if the transposon is the only site the enzyme cuts at. Transposons carrying different restriction enzyme recognition sequences are inserted into the target genome allowing the frequency and distribution of sites to be controlled, easing the construction of a map. Of course, this procedure nullifies one advantage of PFGE mapping strategies in that a technique for introducing the transposon into the target organism must exist and the transposon itself must work once introduced. For these reasons, and the fact that the extra work involved in preparing the transposons often reduces the cost-effectivenes of top-down mapping, this technique has seen little use. Applications where it has been used include comparative mapping of strains of *E. coli* using a derivative of Tn10 (Bloch et al., 1994; Rode et al.,

1995) and a study on the genomic organization of *E. coli*, *Brucella melitensis* and *Agrobacterium tumefaciens* using a derivative of Tn5 (Jumas-Bilak et al., 1995).

A rather specialized variation of standard PFGE mapping involves the use of the recently discovered restriction enzyme I-*Ceu*I (Gauthier et al., 1991; Marshall and Lemieux, 1991). This enzyme is one of the so-called homing endonucleases and is encoded by a group I intron found in the chloroplast 23S ribosomal RNA gene of *Chlamydomonas eugametos*. This enzyme recognizes a 26 bp sequence that is found only in the 23S ribosomal RNA genes of many bacteria (proteobacteria and Gram-positive bacteria), mitochondria and chloroplasts (Liu et al., 1993a). Because of this enzyme's specificity, it has been used extensively by Sanderson and colleagues in comparative mapping of *E. coli* and various strains of *Salmonella* (Liu et al., 1993a; Liu et al., 1993b; Liu et al., 1993c; Liu et al., 1994; Liu and Sanderson, 1995a; Liu and Sanderson, 1995c; Liu and Sanderson, 1995d) as well as in fingerprinting studies (Liu and Sanderson, 1995b). I-*Ceu*I has also been confirmed to cut in all ten ribosomal RNA (*rrn*) operons of *Bacillus subtilis* 168 and the three *rrn* operons of *Rhizobium meliloti* 1021 and nowhere else in either genome (Honeycutt et al., 1993; Toda et al., 1995). Claims for the usefulness of this enzyme for very quick and easy comparisons between strains is well founded but only in certain cases. The utility of I-*Ceu*I is doubtful when applied to genomes with only one or two *rrn* operons and it cannot be applied to the Archaea since it does not cut within the *rrn* operons of members of this domain.

The problem of unsuitable recognition sequence frequency and distribution comes into play only with certain organisms however and the major drawback of PFGE maps and top-down strategies in general is the low resolution of the physical maps generated. This carries over into the genetic map which is often created by localizing previously cloned

genes to specific restriction fragments by hybridization. Top-down map studies are often reduced to reporting that mapped genes were localized to a certain region on the genome and that the specific order of these genes is unknown. While the ability to map a number of related genomes quickly—21 strains of *Borrelia* sp. in one study (Casjens et al., 1994)—allows certain questions about the evolutionary history of a lineage to be addressed, the superficial nature of these maps precludes most in-depth analyses.

Bottom-up mapping strategies provide much higher resolution than top-down mapping with the cost of greater initial effort in their construction. This cost has greatly biased the research community against bottom-up methods; of the 93 maps compiled by Fonstein and Haselkorn (1995), eighty were constructed by PFGE and only 13 by what they term gene encyclopedias (a bottom-up strategy). Bottom-up genome maps are constructed by creating ordered clone libraries of the genomes. In most cases, lambda or cosmid vectors are used as these provide insert sizes (20-45 kb) large enough to keep the total number of clones in the library down to a reasonable number while still being relatively easy to work with. Yeast artificial chromosomes (YAC's) which have seen extensive use in eukaryotic genome projects have been used only on the unusually large genome of *Myxococcus xanthus* (Kuspa et al., 1989) as well as *B. subtilis* (Azevedo et al., 1993; Serror et al., 1993) which clones poorly using *E. coli* vectors. Although they have not been used for such a purpose as yet, Cole and Saint Girons (1994) point out that P1 vectors and bacterial artificial chromosomes (BAC's) may find an application in prokaryotic genome mapping. Vectors with large insert sizes have the advantage of reducing the total number of clones in a library needed to cover the target genome but they have the disadvantage of reducing the resolution of the resulting map. This reduction in resolution can be overcome by making restriction maps of each clone but this greatly

increases the amount of work necessary and restriction mapping can become very complicated as the size of the clone increases. Therefore, a compromise must be struck between library size and map resolution which considers the uses to which the resulting map will be put.

While in some cases top-down maps can approach the resolution of bottom-up maps when techniques such as 2-dimensional gels are used (Hackett et al., 1994), top-down maps do not produce anything comparable to the clone library of bottom-up maps. Once a bottom-up map is completed, the clone library becomes a valuable resource not only as a continuing source of genomic DNA but also for any application that seeks to locate genetic loci on the genome. This is illustrated by the number of studies that have used the cosmid clone libraries of *Hf. volcanii* DS2 (Charlebois et al., 1991) and *Hb. salinarum* GRB (St. Jean et al., 1994 [chapter 3]) since their construction (Trieselmann and Charlebois, 1992; Hackett et al., 1994; López-García et al., 1995 [chapter 5]; Ferrer et al., 1996; St. Jean and Charlebois [chapter 6]) to say nothing of the value of the lambda clone library of *E. coli* described in Kohara et al. (1987).

Compared to physical mapping methods, genetic mapping in prokaryotes has fallen out of favour among researchers. The reasons for this are not hard to understand; the amount of effort involved is at least comparable to a bottom-up physical map, genetic maps often include inaccuracies in the ordering of closely linked loci, and an efficient system for transferring DNA between cells is needed before a genetic map can be made. These limitations were accepted when the genetic maps of such genomes as *E. coli* (Taylor and Thoman, 1964), *S. typhimurium* (Sanderson and Demerec, 1965) and *B. subtilis* (Henner and Hoch, 1980) were first constructed since there were no alternatives, but even for these genomes physical maps have since supplemented the genetic maps as

sources of map information (Kohara et al., 1987; Smith et al., 1987; Liu and Sanderson, 1992; Wong and McClelland, 1992; Itaya, 1993; Liu et al., 1993b). A number of other genetic maps also exist but most work seems now to be concerned with integrating them with more recently derived physical maps (Hopwood et al., 1993; Pattee, 1993; Vary, 1993; Welker, 1993). This trend is unlikely to reverse since once a physical map is made of a genome, it is a simple matter to construct a genetic map by hybridizing probes of specific loci to the genome either in the form of a clone library for bottom-up maps or digested DNA in a pulsed-field gel for top-down maps. A variation combining both genetic and physical mapping techniques has been applied to the bottom-up map of *Hf. volcanii*. Biosynthetic genes were mapped by complementing auxotrophic strains with DNA derived the cosmid library of *Hf. volcanii*'s genome (Conover and Doolittle, 1990). A total of 139 auxotrophs involving 14 amino acids and three nucleic acids were mapped to 23 chromosomal loci using this method (Cohen et al., 1992). Physical maps have another advantage over purely genetic maps in that whole classes of loci—such as tRNA's or a family of insertion sequences—can be screened for and all or most loci found in a single hybridization.

Large Scale Gene Expression Studies

While genome fingerprinting and mapping are concerned with the structure of the genome, gene expression studies start to look at its functioning. Investigations into the controls of expression of individual genes cannot legitimately be included in genomics which applies more to the genome as a whole but some expression studies do fulfill the aims of genomics, specifically investigations into the response of prokaryotes to changes in their environment. In the past such studies have consisted of total protein extracted

from cells grown under different conditions and separated on 2-dimensional electrophoresis gels. Differences in the protein pattern on the gels imply changes in gene expression in response to the changed growth condition. Running previously purified proteins under the same conditions allows the identification of these proteins in cell lysate preparations but otherwise identification of the various proteins is problematic and in no case is there an easy way to map or clone a target protein once identified. Even so, this method has been used successfully on a number of organisms to identify their patterns of gene expression (Daniels et al., 1984; Steck et al., 1993; Giometti et al., 1995; Hartke et al., 1995; Laurent-Winter et al., 1995).

A recent innovation makes use of the clone libraries associated with bottom-up physical maps discussed in the previous section. By extracting total RNA from cells and hybridizing this RNA to Southern or dot blots of a genomic clone library, a pattern of gene expression can be determined for a genome under whatever growth conditions desired. This system has the advantage of mapping all expressed genes as a matter of course and by definition, any gene giving a signal is already cloned. This method has been used in three studies on two prokaryotes. Trieselmann and Charlebois (1992) examined heat shock and growth in rich and chemically defined media during exponential phase in *Hf. volcanii*, Ferrer et al. (1996) used the same organism grown in three concentrations of salt and Chuang et al. (1993) investigated *E. coli* grown under nutrient starvation, heat shock, salt shock, anaerobic conditions, and in the mouse gut, as well as four strains carrying mutations showing pleiotropic effects. This sort of study can target whole classes of genes which may be of interest which can then be pursued with relative ease by Northern blot analysis or DNA sequencing.

Identifying genes of interest is not the only use for this method. As its inclusion in genomics implies, the pattern of expressed genes can itself be useful. The megaplasmids present in *Hf. volcanii* have no known function even though they collectively make up 29% of the genome. The studies of Trieselmann and Charlebois (1992) and Ferrer et al. (1996) indeed show that most of the plasmids are transcriptionally quiescent under the conditions tested but that some loci are active. Are these loci simply maintaining the plasmids within the cell or do they actually benefit the cell in some way? Having these loci already cloned into cosmids (Charlebois et al., 1991) would make it a relatively simple matter to find out.

DNA sequencing and Bioinformatics

The most celebrated branch of genomics is, of course, the sequencing of complete genomes. Whatever one's opinion about the scientific merit of sequencing the genome of an organism, it is hard to deny the appeal the prospect has on the imagination. An indication of this appeal is given by the fact that announcements proclaiming the completion of the sequencing projects of both *Haemophilus influenzae* and *Mycoplasma genitalium* appeared in daily newspapers. This is a notable feat for any story involving bacteria that doesn't deal with the outbreak of a pathogen or the curing of a disease.

The first prokaryotic genome sequencing project was begun in the late 1980's with none other than *E. coli* as the target genome. At the time, this project was started almost as an afterthought to the much more publicly engrossing Human Genome Project. Because of this 'little-brother' image, and because of somewhat naïve ideas about the difficulties inherent in large-scale DNA sequencing, the then conservative estimates for the completion of the *E. coli* genome have all come and gone with a significant portion of the

genome still unsequenced. Despite this delay, the consensus in the research community still favours the sequencing of genomes as indicated by the continued support of the *E. coli* project (recounted in Danchin [1995]). Indeed, advances in technology, methodology, informatics, and organization have brought the possibility of genome sequencing within the grasp of more and more researchers and they have responded with a deluge of new projects.

The most well known organization dedicated to the sequencing of prokaryotic genomes is The Institute for Genomic Research, TIGR. To date, this is the only body to have published complete prokaryotic genomes in the form of *Haemophilus influenzae* (Fleischmann et al., 1995), *Mycoplasma genitalium* (Fraser et al., 1995), and *Methanococcus jannaschii* (Bult et al., 1996) and is by far the most productive group, being able to complete a moderately sized bacterial genome in less than a year. TIGR by no means has a monopoly on genome sequencing and Table 1.4 provides a list of ongoing prokaryotic sequencing projects. References for these projects are given where possible but because in many cases no information is published until the completion of the project, citations may not be available. In addition to the listed projects, there is a trend for some companies to retain as intellectual property the genomes they sequence. This has been the case for two strains of *Helicobacter pylori* sequenced by Glaxo and Genome Therapeutics and *Staphylococcus aureus* sequenced by Human Genome Sciences Inc. (HGSI) which is associated with TIGR. A great deal has been made of how scientific progress will be hurt by withholding genome sequences from the public domain. Given the amount of data now being generated, however, researchers would probably be better served thinking about what they can do with what they have rather than what they can't do with what they don't have.

Table 1.4. Prokaryotic genome sequencing projects in progress.

Organism	Company or Group Leader(s)	Reference
Archaeoglobus fulgidus	TIGR	
Bacillus subtilis	European and Japanese consortium	Kunst et al., 1995; Ogasawara et al., 1995
Deinococcus radiodurans	TIGR	
Escherichia coli	F. Blattner and T. Horiuchi	Burland et al., 1995
Methanobacterium thermoautotrophicum	Genome Therapeutics	Smith et al., 1995
Mycobacterium leprae	S. Cole; Genome Therapeutics	Honoré et al., 1993; Smith et al., 1995
Mycobacterium tuberculosis	Genome Therapeutics	Smith et al., 1995
Mycoplasma capricolum	W. Gilbert	Bork et al., 1995
Mycoplasma pneumoniae	R. Herrmann	
Pyrobaculum aerophilum	J. Miller and M. Simon	
Pyrococcus furiosus	F. Rabb and B. Weiss	Rabb et al., 1995
Rhodobacter capsulatus	R. Haselkorn	
Rhodobacter sphaeroides	S. Kaplan and G. Weinstock	
Rickettsia prowazekii	C. Kurland	
Sulfolobus solfataricus	R. Charlebois, M. Ragan, and W.F. Doolittle	Sensen et al., 1996
Synechocystis strain PCC6803	S. Tabata	Kaneko et al., 1995a Kaneko et al., 1995b
Ureaplasma sp.	Schlessinger	

What to do with the prodigious amounts of sequence data, in fact, has the potential to be one of the most interesting aspects of genomics but this potential has yet to be fully realized. One of the points that is brought up time and again when discussing whole genome sequencing is the novel situation of possessing a complete 'blueprint' of a cell. 'For the first time in history', so the story goes, 'we will have at our fingertips all the instructions necessary to make a cell. Such a windfall will deepen our understanding of how living things work in ways piecemeal approaches simply can't achieve'. To this end, a great deal of work has been devoted to so-called gene inventories; computer aided identification of open reading frames (ORFs) and searches of sequence databases that try to assign them functions by identifying similarities with better characterized genes from the same or other organisms. This was the major theme of both genome sequence papers so far published (Fleischmann et al., 1995; Fraser et al., 1995) as well as a several additional works (Gonnet et al., 1992; Green et al., 1993; Borodovsky et al., 1994; Bork et al., 1995; Casari et al., 1995; Fsihi and Cole, 1995; Koonin et al., 1995; Ouzounis et al., 1995). NCBI, EMBL and TIGR have all been particularly active in this field and have software available at their world wide web sites for the identification and/or assignment of ORFs. This sort of investigation is important for annotating newly sequenced DNA and for identifying candidate loci for more intensive study. The latter pursuit seems in many cases to have been forgotten, however, as the identification of ORFs has become an end in itself. This can only go on for so long before the inferences made to identify the functions of ORFs become too tenuous and recourse to direct observations need to be made. This is especially true given the fact that ORFs with no assigned function make up 43% of all identified ORFs in *H. influenzae* and 32% in *Mycoplasma genitalium* (Fraser et al., 1995). Unfortunately, given the ease with which similarity searches can be done and the time and

effort involved in the biochemical characterization of any gene product, the only evidence for a gene's function in many cases will be 'it looks a lot like this one' for some time to come.

Gene inventories are only one use for genome sequences. The organization of genes in the genome and their origins and patterns of evolution are other aspects beginning toreceive more attention. Using the genome sequence data available for *E. coli*, Eyre-Walker (1995) found a positive correlation between intergenic spacing and the expression level of the downstream gene. Labedan and Riley (1995a, 1995b) have used database similarity searches to investigate the origins of *E. coli* genes. They have found that the majority of *E. coli* genes can be grouped into gene families and propose that extensive gene duplication and divergence has occurred to provide *E. coli* with its current gene complement. This method was also used to identify genes that may have been introduced into *E. coli* via horizontal transfer.

The large numbers of sequenced genes also allows statistical methods to be brought to bear. *E. coli* genes have been divided into two classes according to their codon usage and level of expression (Gouy and Gautier, 1982; Blake and Hinds, 1984). Using the much larger data set available due to the *E. coli* sequencing project, Médigue et al. (1991) were able to divide *E. coli* genes into three categories according to codon usage. These categories corresponded to 1) highly expressed genes, 2) rarely expressed or genes expressed at low levels and 3) genes from mobile genetic elements, genes coding for cell surface components and genes related to the fidelity of DNA replication. More recently, the AT bias of late replicating DNA was investigated in *E. coli* (Deschavanne and Filipski, 1995). Weakly expressed genes were examined from around the *E. coli* genome and it was found that genes nearer the terminus of replication did indeed have a consistently higher

mol% A+T. This AT bias was attributed to different DNA repair mechanisms being used in genes nearer the origin (homologous repair between sister chromosomes) and the terminus (excision repair which preferentially incorporates adenine). The distribution of genes around the *E. coli* chromosome has been investigated (Williamson et al., 1993). Genes were found to occur in clusters evenly spaced around the chromosome, an arrangement which could not be reproduced by randomized trials. Based on their results, the authors suggest that enteric bacterial chromosomes evolved by the accretion of smaller replicons. While these studies do not absolutely require the large amounts of data generated by genome sequencing projects, they are greatly facilitated by them. It can be seen, however, that in most cases sequence data are not enough and that additional information in the form of expression studies, characterization of biochemical processes (such as DNA repair) or codon usage biases due to tRNA abundance is needed. Perhaps this is why so many of these studies continue to be conducted on *E. coli* while so many other genomes are being mapped and sequenced.

Comparison Studies

Probably the most common use genomics data is put to is in comparison studies. Comparisons as they are presently conducted can provide information on the number and types of rearrangements that have occurred between genomes and differences in their gene inventories. Genome fingerprints by their very nature are limited to investigating the numbers of rearrangements between genomes and as discussed previously, are used almost exclusively to determine phylogenies for closely related organisms. Large scale sequence-level comparisons are just beginning to be done as more sequence data from a wider range of genomes become available and as the technical difficulties in coping with the prodigious

amounts of information are overcome (for example see Kunisawa, 1995). Mapping studies for now provide just the right amount of information to allow comparisons to be done with relative ease while still supplying worthwhile results to make them especially attractive for comparisons.

Most mapping studies are conducted with an eye for comparing sets of maps, usually between strains or species but sometimes between different genera. The usual course for such comparisons is that genetic markers are placed by hybridization onto a physical map constructed by a top-down or bottom-up strategy. The order of the genetic markers on each map is then compared and conclusions drawn about the types and frequency of rearrangements that have occurred since the genomes diverged from one another. The quality of these comparisons varies widely with the resolution of the maps being used and the numbers of genetic markers placed on the maps. In some cases, enough information is gained that specific conclusions about the evolutionary history or the organization of the genomes under study can be gained. Comparisons of ten strains of *Clostridium perfringens* led to the conclusion that serological variation and differences in pathogenicity may be due entirely to differences in extrachromosomal elements (Canard et al., 1992). High-resolution bottom-up maps of three strains of *Rhodobacter capsulatus* revealed a mosaic pattern of sequence divergence in that regions with highly conserved restriction maps were interspersed with highly polymorphic regions (Nikolskaya et al., 1995). This pattern of conservation and divergence is mirrored by the three strains of *Hb. salinarum* compared in Hackett et al. (1994) where restriction maps for all three strains are conserved except in two regions of the chromosome. Much work has been done by K.E. Sanderson and colleagues on various strains of *Salmonella*. The mapping of five strains of *Salmonella* (*typhimurium*, *paratyphi* A, *paratyphi* B, *enteritidis*, and *typhi*)

allowed the unusual organization of *S. typhi*'s chromosome relative to the other strains to be identified (Liu and Sanderson, 1995a). A tentative series of recombination events that led to the rearranged chromosomal map of *S. typhi* was even proposed.

Often, the only conclusions that are drawn are purely qualitative, either the genomes show conservation in their gene orders or they do not. The utility of such conclusions is doubtful when one considers that there are no standards by which to measure the similarity of two or more genomes as have been proposed for genome fingerprinting studies (Tenover et al., 1995). The varied quality of maps probably negates the possibility of formalizing such broadly applicable standards.

In many cases, genomic map comparisons suffer from a lack of resolution. The use of top-down maps and the availability of limited numbers of genetic markers contributes to this problem but the difficulty in analyzing even moderate numbers of markers is also a significant factor. This difficulty is manageable when closely related genomes that have not undergone many rearrangements are compared. As phylogenetic distance increases, and as sequence data begin to be used which will make available many more genetic loci for each comparison, it will become necessary to use computer aided combinatorial methods to make sense of the data. Work in this area has already begun with computer algorithms being developed that can determine the shortest distance (measured in number of rearrangements) separating two genomes (Bafna and Pevzner, 1995; Blanchette et al., 1996; Kececioglu and Sankoff, 1995). The comparison described in chapter 6 makes use of one of these programs and the problems of genomic map comparisons and some possible solutions are more fully discussed there.

The Nucleoid

No discussion of prokaryotic genomics would be complete without looking at the nucleoid. When concentrating on map, sequence and protein expression data, it is easy to forget that the prokaryotic genome exists as a 3-dimensional structure with a specific organization influencing, and influenced by, the functional needs of the cell. From light and electron microscopy revealing nucleoid morphology, to studies of DNA binding proteins and regulation of DNA supercoiling by topoisomerases, such investigations provide insights not obtainable from staring at strings of letters no matter how long they might be.

Organization of the Prokaryotic Chromosome

To any organism, the problem of packing its genetic material in a confined space while allowing accessibility for gene expression is of paramount importance. The nucleoid of *E. coli* and various species of *Bacillus* have been particularly well studied in this regard (for a recent review see Robinow and Kellenberger, 1994). Chromosomal DNA is confined to an area of the cell termed the nucleoid. The nucleoid is not a static structure but is constantly changing shape. It is made up of a central core of ribosome-free DNA from which numerous projections of DNA (called excrescences) extend into the cytoplasm. The core nucleoid is free of ribosomes while proteins such as RNA polymerase, HU, topoisomerase I (Ryter and Chang, 1975; Dürrenberger et al., 1988) and H-NS (Dürrenberger et al., 1991) as well as single-stranded DNA (Hobot et al., 1987) have been found to occur at the periphery of the nucleoid, implying that the core DNA is sequestered and not metabolically active. It is thought that the excrescences specifically are the site of transcription, an idea supported by the fact that they disappear if the cell is treated with many compounds that interrupt protein synthesis (Robinow and Kellenberger,

1994). The nucleoid can often become attached to the inner membrane of the cell through these excrescences. Integral membrane proteins and proteins to be exported can insert into the membrane as they are being translated forming a continuous link between the nucleoid and membrane through the transcription/translation apparatus (Ma et al., 1994). The nucleoid also is attached to the outer membrane at *oriC* soon after chromosome replication is initiated as a normal part of cell cycle regulation and segregation of daughter chromosomes after replication (Worcel and Burgi, 1974; Hendrickson et al., 1982).

The topology of the chromosome is tightly regulated in *E. coli* through the complementary actions of topoisomerase I and topoisomerase II (gyrase). Gyrase introduces negative supercoils into the DNA in an ATP-dependent manner while topoisomerase I can relax only negatively supercoiled DNA. Normally, the chromosome is negatively supercoiled with an average linking number of -0.06 (Bates and Maxwell, 1993). Promoters controlling the *gyrA* and *gyrB* genes coding for the two subunits of gyrase are more active when they are relaxed while the *topA* (topoisomerase I) gene promoter is more active when it is supercoiled (Menzel and Gellert, 1983; Tse-Dinh, 1985). The activity of the proteins is also affected by DNA topology. As one might guess, gyrase binds more efficiently to relaxed DNA while topoisomerase I binds preferentially to negatively supercoiled DNA (Wang, 1971; Sugino and Cozzarelli, 1980). While the actions of these two topoisomerases can maintain the global level of supercoiling needed by the nucleoid, the process of transcription can overwhelm this regulation leading to altered local supercoil levels (Rahmouni and Wells, 1992). Two additional topoisomerases exist in *E. coli* as well. Topoisomerase III possesses a decatenating activity and may be involved in decatenation of daughter chromosomes (Digate and Marians, 1988; 1989). Topoisomerase IV shows sequence homology to gyrase at the amino acid level but relaxes

both positively and negatively supercoiled DNA; it cannot introduce supercoils (Kato et al., 1990; 1992). This enzyme is bound to the inner membrane and is involved in the segregation and possibly decatenation of daughter chromosomes (Adams et al., 1992). Although overexpression of topoisomerase IV can compensate for loss of topoisomerase I activity (Dorman et al., 1989; Free and Dorman, 1994; McNairn et al., 1995), the functions of topoisomerase III and IV appear to be rather specialized and are probably not involved in the regulation of global supercoiling under normal circumstances.

The *E. coli* chromosome is organized into approximately 50 loops or domains of about 100 kbp each (Sinden and Pettijohn, 1981). These loops are independently supercoiled in that relaxation of one domain through breaks in the DNA strand will not relax others (Drlica, 1987). This independence is thought to be mediated by gyrase binding to the base of each loop at repetitive extragenic palindromic (REP) sequences or toposites (Yang and Ames, 1990; Condemine and Smith, 1990). Experiments using the gyrase inhibitor oxolinic acid have indicated that 50 active gyrase sites exist per cell (Snyder and Drlica, 1979); a number which matches nicely with the estimated number of domains. Independent experiments in *E. coli* and *S. typhimurium* indicate that these domains are not regulated to different supercoil levels (Miller and Simons, 1993; Pavitt and Higgins, 1993). While only 50 gyrase sites seem to be active per cell at any given time, there are far more potential binding sites for gyrase on the *E. coli* chromosome. REP sequences are estimated to make up 0.5% of the chromosome, distributed in hundreds of copies (Yang and Ames, 1990). This opens the possibility that different gyrase binding sites are active at different times and that the domains are not static, although no direct evidence exists supporting or refuting this hypothesis.

In contrast to eukaryotic chromosomes, bacterial chromosomes are to a large extent free of proteins (the core nucleoid). Nevertheless, small DNA-binding proteins have been found in bacteria, notably in *E. coli* (for reviews see Drlica and Rouvière-Yaniv, 1987; Schmid, 1990). Four of the best characterized of these proteins are HU, H-NS (H1), IHF, and FIS. These proteins have been given the moniker histone-like proteins mainly because of their small size, relative abundance in the cell and association with the nucleoid. This link to the eukaryotic histones seems to be for the most part unfounded, however, since similarities between the two groups end there. HU, the most abundant of these proteins within the cell, has been shown to compact DNA only in vitro (Broyles and Pettijohn, 1986) and it is unknown whether it does so in vivo while X-ray crystalography of HU derived from *Bacillus* has elucidated a 3-dimensional structure completely unlike that of eukaryotic histones (Tanaka et al., 1984). H-NS plays a major role in the *E. coli* cell as a transcriptional regulator (Göransson et al., 1990; Ueguchi and Mizuno, 1993; Zhang et al., 1996) which also influences genetic recombination (Higgins et al., 1990) while IHF and FIS are involved in site-specific recombination reactions (Johnson et al., 1986; Friedman, 1988). Mutants lacking HU or H-NS show pleiotropic phenotypes indicating multiple functions for these proteins.

No convincing eukaryote-like nucleosomes have been observed in the *E. coli* chromosome. A study by Eickbush and Moudrianakis (1978) purporting to show the typical 'beads-on-a-string' nucleosomes has since been reinterpreted to be an artifact of the fixation procedure (Kellenberger and Arnorld-Schulz-Gahmen, 1992). To replace nucleosomes, it has been proposed that *E. coli* maintains the bulk of its DNA in what are called compactosomes (Kellenberger, 1990; Kellenberger and Arnorld-Schulz-Gahmen, 1992). A compactosome is simply DNA that is to some degree condensed due to its being

supercoiled. Compactosomes do not require the presence of proteins and thus the supercoils would not be constrained as they would in DNA intimately associated with protein. All this fits well with the absence of protein in the core of the nucleoid and the unconstrained nature of half the supercoils present in the *E. coli* chromosome. A further problem is encountered when trying to package this protein-free DNA in that the negative charges on the phosphate backbone must be neutralized. In the Eukarya, this is done by the positively charged histones. In Bacteria, it has been suggested that the same function may be performed by polyamines such as spermidine and putrescine as well as Mg^{2+} present within the core nucleoid (Kellenberger, 1990).

These disparate observations allow a general overview of the workings of the nucleoid to be developed. The bulk of chromosomal DNA is held in a ribosome-free region of the cell called the core nucleoid; it is not metabolically active and for the most part is not in contact with the cytoplasm. This DNA is condensed into compactosomes that involve no protein interactions with the DNA and depend on the action of topoisomerases to maintain their topology. All the action occurs at the surface of the nucleoid where various DNA-binding proteins and transcription apparati have been localized. Excrescences projecting into the cytoplasm are continuously forming and melting back into the core nucleoid bringing different genes and operons to the surface to be expressed. Some of these excrescences connect the nucleoid to the inner membrane of the cell through the products of transcription and translation of integral membrane proteins. All the while, topoisomerase I and gyrase work to maintain the level of supercoiling that is an absolute requirement for the proper functioning of the nucleoid.

Much less is known about the archaeal nucleoid compared to its bacterial counterpart. Enough is known, however, that parallels can be drawn between the Archaea

and both Bacteria and Eukarya. The main thrust of nucleoid studies in the Archaea has been in the areas of DNA-binding proteins and certain topoisomerases of thermophilic Archaea. In the latter case, a reverse gyrase activity has been found in many genera of the Crenarchaeota and a subset of the thermophilic Euryarchaeota (Bouthier de la Tour, 1990). Reverse gyrase introduces positive supercoils into DNA in an ATP-dependent manner in vitro (Kikuchi and Asai, 1984; Forterre et al., 1985) as opposed to negative supercoils as gyrase does. The enzyme responsible for this activity has been characterized only in the Crenarchaeota so it is unknown whether any other similarities between the two kingdoms exist. It is unknown whether these Archaea maintain their genomes in a positively supercoiled state although the isolation of positively supercoiled SSV1 DNA (a virus-like particle from *Sulfolobus* sp. B12) suggests that this might be the case, at least for the Crenarchaeota (Nadal et al., 1986). Suggestions for the purpose of reverse gyrase include maintaining the stability of DNA at high temperatures (Kikuchi and Asai, 1984; Kovalsky et al., 1990), counteracting the effects of other topoisomerases, and for converting cruciform or Z-DNA back to B-DNA (Collin et al., 1988). Reverse gyrase activity has also been found in *Thermotoga* spp., a member of the Bacteria (Bouthier de la Tour, 1991) but since no genes for this activity have yet been cloned from Bacteria, it is unknown whether this activity in the two domains has a common origin. If homologous genes are eventually found and they function in thermostability, this would tend to support the idea of a thermophilic habit for the common ancestor of all life.

The study of DNA-binding proteins in Archaea encompasses both major divisions of the group, the Crenarchaeota and the Euryarchaeota, focusing on the sulfur-dependent thermophiles and the methanogens. As with the Bacteria, the DNA-binding proteins from Archaea have been labeled histone-like proteins although in this case the name seems to be

more applicable. These proteins can be divided into four groups; the MC1 family from the Methanosarcinaceae, the HMf family from the Methanobacteriales, HTa from *Thermoplasma acidophilum* and a family of proteins from members of the genus *Sulfolobus* (Grayling et al., 1994). The archaeal histone-like proteins provide a bridge between the Bacteria and Eukarya. HTa shows weak homology to HU—they share certain amino-acids known to be involved in the interaction between HU and DNA (Drlica and Rouvière-Yaniv, 1987)—while the HMf family proteins show sequence homology to all eukaryotic histones but especially histone H2a (Sandman et al. 1990; Grayling et al., 1994). The relationship between HMf proteins and eukaryotic histones has recently been bolstered by an NMR study of HMfB from *Methanothermus fervidus* (Starich et al., 1996). The 3-dimensional analysis of this protein revealed significant similarities in secondary, tertiary and quaternary structures with those of eukaryotic core histones.

The interactions of histone-like proteins with DNA seems to be significantly different between Bacteria and Archaea. In *T. acidophilum* (Searcy and Stein, 1980) and *Hb. salinarum* (Takayanagi, et al., 1992), electron microscopy of nucleoids released from cells showed nucleosome-like structures. These nucleosome-like structures were smaller than eukaryotic nucleosomes and could appear at irregular intervals along the DNA strand. Treatments with proteinase K and DNase I confirmed that the haloarchaeal nucleosome-like structures were not artifacts. In *Hb. salinarum*, only a portion of the DNA was found to be associated with these nucleosome-like structures, the rest being protein-free. The relative amounts of protein-free and protein-bound DNA changed depending on the growth phase of the cells; protein-free DNA predominating in log phase and protein-bound DNA making up almost the entire nucleoid in late stationary phase (Takayanagi et al., 1992). The histone-like proteins from *Sulfolobus* species have also

been investigated and been found to form specific structures with DNA in vitro (Lurz et al., 1986). These structures are morphologically dissimilar to those found in *Hb. salinarum* and *T. acidophilum*. This finding suggests that a fundamental difference might exist in how the Crenarchaeota and Euryarchaeota structure their nucleoids.

A study by Ronimus and Musgrave (1995) looking at the DNA-binding properties of various archaeal histone-like proteins tends to support this conclusion. Proteins were isolated from three euryarchaeotes and *Sulfolobus solfataricus*, a crenarchaeote. Whereas the protein profiles seen on SDS-PAGE were consistent for all the Euryarchaeota, they were quite different from that given by *S. solfataricus*. The effects on the mobility of DNA through electrophoretic mobility shift assays (EMSAs) also revealed consistent results between the euryarchaeotes and differences from the crenarchaeote, *S. solfataricus*. The differences observed are unlikely to be due to the thermophilic habit of *S. solfataricus* as the euryarchaeotes tested—two species of *Pyrococcus* and *Thermococcus celer*—are also thermophiles.

Many more gaps exist in our knowledge of archaeal nucleoids compared to their bacterial counterparts. For example, it is unknown whether the archaeal chromosome is divided into domains as bacterial chromosomes are. Gyrase has been isolated from Archaea and there is even some evidence to suggest that the *gyrA* and *gyrB* genes of haloarchaea are regulated by changes in DNA topology in the same way those of *E. coli* are (Holmes et al., 1991). However, sequences analogous to REP's have not been found in Archaea making comparisons of the roles of gyrase in nucleoid organization between Bacteria and Archaea tenuous at best. Changes to global nucleoid supercoiling have also been much better studied in the Bacteria compared to the Archaea, an area that will be addressed in the next section. From what is known of archaeal nucleoids, it appears that a

fundamental difference emerged between the Crenarchaeota and Euryarchaeota in organization and/or functioning soon after divergence from the lineage that would lead to the Eukarya. This also makes inferences about the Archaea using data from the Bacteria problematic since changes that caused the two kingdoms of Archaea to diverge almost certainly caused at least one kingdom to diverge from the Bacteria by the same degree.

Supercoiling and Gene Expression

As stated in the previous section, DNA supercoiling is tightly regulated in the bacterial chromosome by the complementary actions of topoisomerase I and gyrase such that, on average, genomic DNA is underwound with a linking number of -0.06 in *E. coli*. The torsional stress induced by the negative supercoiling of DNA serves multiple functions: helping to compact DNA in the nucleoid, facilitating such cellular activities as transcription by contributing free energy to the denaturation of DNA by RNA polymerase, and affecting gene expression by influencing the activity of promoters. A great deal of work has been done to elucidate the role of DNA supercoiling on gene expression. Different methods have been used to try to overcome the inherent difficulties in measuring DNA topology in vivo including reporter genes fused to supercoil-sensitive promoters, reporter plasmids, topoisomerase activity levels from cell extracts and DNA migration through sucrose density gradients. Although these methods have provided many insights into the workings of supercoiling in the bacterial cell, some only indirectly measure the topology of the DNA of interest (reporter plasmids) while others have been used in systems with artificially induced levels of supercoiling (systems that make use of topoisomerase mutants or topoisomerase inhibitors). This necessitates caution when interpreting the results of such experiments.

It is fairly well established that the activity of many genes is affected by their level of supercoiling, both in the Bacteria (Drlica et al., 1990) and to a lesser extent, the Archaea (Holmes et al., 1991; Yang et al., 1996). It is not hard to understand why this might be the case. Any protein binding to DNA will make contact at multiple sites along the DNA strand which may depend on specific sequences or on more general criteria such as helical pitch. A change in DNA supercoiling in either direction will change the spacing and rotational orientation between sites on the DNA strand by changing the twist of the DNA. The latter effect could serve to push two binding sites that were originally on the same side of the DNA helix out of alignment. Changes in spacing and rotational orientation would almost certainly affect the binding constants of proteins involved in the expression of genes whether they be repressors or activators binding to a promoter or RNA polymerase itself. With the sensitivity of gene expression to supercoiling readily accounted for, the question becomes: does DNA supercoiling change enough in the normal course of a cell's life to affect gene expression and if so, does the cell take advantage of this effect? In both cases, the answer appears to be yes.

DNA supercoiling, while tightly controlled in Bacteria such as *E. coli*, is not maintained at the same level under all environmental conditions. An early study showed that supercoil levels change in a consistent manner due to altered oxygen availability (Yamamoto and Droffner, 1985). From their investigation, Yamamoto and Droffner proposed that a changing environment—in this case a shift from aerobic to anaerobic growth or vice versa—would cause a change in global supercoiling of the genome providing a signal to alter the expression of large numbers of genes. Since then, further studies have supported the link between shifts in oxygen availability and DNA supercoiling (Hsieh et al., 1991a) and extended it to include other environmental conditions including

medium osmolarity (Hsieh et al., 1991b), pH (Karem and Foster, 1993), nutrient status (Balke and Gralla, 1987) and thermoregulation (Rohde et al., 1994). Meanwhile, other work has supported the idea that global supercoiling acts as a regulator of gene expression in response to changes in the environment (Higgins et al., 1990; Steck et al., 1993). The study of Steck et al. (1993) is especially important in this regard as it made use of topoisomerase mutants that caused deviations in supercoil levels of no more than 20%, comparable to the changes in supercoiling seen in wild type cells. This is unlike many of the other studies that either used plasmids to infer changes in chromosomal DNA or used topoisomerase mutants or inhibitors to produce exaggerated changes in supercoiling. It is unknown how a change in the environment triggers a change in DNA supercoiling although it has been suggested that gyrase activity is the mechanism by which this change is effected (Dorman, 1995). For Bacteria, gyrase is the only enzyme known to possess the ability to introduce negative supercoils into DNA making it an attractive target for a regulatory pathway. Gyrase also requires ATP to produce negative supercoils and is affected by the [ATP]/[ADP] ratio within a cell (Drlica, 1992) and it has been found that the [ATP]/[ADP] ratio changes in concert with changes in DNA supercoiling in response to altered oxygen availability (Hsieh et al., 1991a) and osmolarity (Hsieh et al., 1991b). More recent work has strengthened the idea of a causal link between [ATP]/[ADP] ratios and DNA supercoiling by using an experimental procedure that altered cellular phosphorylation potential in three different ways while minimizing possible side effects from such processes as transcription (Van Workum et al., 1996). Again, gyrase was proposed as the factor mediating the coupling between changes in [ATP]/[ADP] ratio and changes in supercoil levels.

As with global regulation, the relationship between gene expression and local changes in supercoiling has been investigated. RNA polymerase is known to alter the topology of the DNA it is transcribing as it migrates along the strand (Gamper and Hearst, 1982). To explain results obtained using reporter plasmids and topoisomerase mutants, the twin-supercoiled-domain model was first proposed by Liu and Wang (1987). This model predicts that transcription induces positive supercoils downstream of RNA polymerase while negative supercoils are induced upstream and that these topological changes could affect the transcription of neighboring genes. It is assumed that the transcription/translation ensemble does not rotate significantly around the DNA because of the significant viscous drag produced by RNA polymerase, the elongating RNA and attached ribosomes but that the DNA rotates as it is pulled through RNA polymerase. The supercoils so induced could be relieved either by the diffusion of the torsional stress along the DNA strand or by the complementary actions of topoisomerase I and gyrase. Neighboring transcriptional units can amplify the effect if they are convergently transcribed by causing a region of extremely high positively supercoiled DNA to form between. Follow up work supported the model (Wu et al., 1988) although, as in the original study, topoisomerase mutants were used that altered the supercoiling of the cell's genome to abnormal levels. Investigations by Rahmouni and Wells (1989; 1992) using wild-type strains also found that transcription affected expression of neighboring genes under certain conditions proving that such an effect at least has the potential of occurring in nature. Since then, the model has been modified such that only certain classes of genes are found to affect a neighbor's expression. It was found that genes coding for integral inner membrane proteins could strongly affect the expression of a neighbor, genes for periplasmic proteins would produce only a moderate effect and genes for cytosolic

proteins produced a negligible effect (Cook et al., 1992; Ma et al., 1994). The different properties of the different classes of proteins have been explained by the anchoring of the transcription/translation ensemble to the inner membrane via the nascent protein in those genes coding for inner membrane and periplasmic proteins (Ma et al., 1994).

All these studies have been done with plasmids and there is no direct evidence that changes in DNA topology caused by transcription alter gene expression on the chromosome. There is some indirect evidence, however. It has been found that the expression of the photosynthetic genes of *Rhodobacter capsulatus* is halted by inhibition of gyrase (Zhu and Hearst, 1988) suggesting that changes in the level of supercoiling—specifically the relaxation of DNA—represses these genes. Direct measurements of supercoiling proved that this was not the case and that supercoiling did not change during derepression (Cook et al., 1989). As derepression of the photosynthetic genes leads to high levels of expression, it was postulated that gyrase is required to relieve positive supercoiling building up between the various operons of the gene cluster. As many of the gene products involved are membrane-bound, this idea is consistent with the findings of Ma et al. (1994). Another example is the type-1 fimbrial inversion system of *E. coli*. Results using topoisomerase mutants suggest that relief of negative supercoils by topoisomerase I is necessary for the proper functioning of the inversion system (Dorman, 1995). It is thought that the nucleoid protein IHF constrains supercoils when the system is in the Phase-OFF orientation preventing them from diffusing along the DNA strand and requiring the topoisomerase to relieve the tension.

If local supercoiling effects as described by the twin-supercoiled-domain model operate on the chromosome, then they have the potential for exerting a significant influence on gene expression. Genes involved in transport make up a significant fraction

(221 or 13%) of the identified genes in *E. coli* (Riley, 1993) and this can almost certainly be extended to most prokaryotes. The large number of potential anchor points together with the organization of the chromosome into topologically isolated domains may make maintenance of local supercoiling on the chromosome very important since these topological constraints would tend to impede the diffusion of torsional stress along the DNA strand. These effects may be less significant for plasmids depending on their size (small plasmids are likely not divided into independent domains), gene content and density (the absence of genes encoding membrane proteins eliminates potential anchor points while low gene density allows for greater diffusion of torsional stress along the plasmid), and method of replication (plasmids that do not bind to the membrane during replication avoid this topological constraint). For the chromosome then—and perhaps to a lesser degree plasmids—the actions of topoisomerases may play a role in conditioning local supercoil levels for proper gene expression. The abundance of gyrase in the cell would allow this enzyme to be used in such a manner (Thornton et al., 1994). The still undetermined role of local DNA topology on chromosomal gene expression should also serve as a note of caution to those linking altered gene expression to environmentally induced changes in global supercoiling. Examples exist of loci—such as *tonB* (Dorman et al., 1988; Y ung and Postle, 1994) and *proU* (Higgins et al., 1988; Ramirez and Villarejo, 1991) of *E. coli*—whose expression was said to be regulated at least in part by global changes in supercoiling, only to have this claim brought under suspicion after investigation using more exacting methods. In these cases, it may be that supercoiling does indeed effect expression but that this reflects a need for functional topoisomerases to relieve transcriptionally induced torsional stress rather than regulation dependent on global changes in supercoil levels.

Genomic Rearrangements

Genetic recombination is a pervasive phenomenon in biology that is intimately linked with genomics. Most genomic studies seek to address the mechanisms of recombination either by examining the results through comparative fingerprinting, mapping or sequencing; or by examining the organization of the genetic material through studies of the nucleoid. Since chapters 4, 5, and 6 of this thesis deal directly with genomic rearrangements, a discussion of the forces promoting such rearrangements seems in order. For the purposes of this discussion, I define genetic recombination as the processes or mechanisms that bring about changes to the arrangement of an organism's genome while genomic rearrangements refer to the specific changes brought about by these processes.

Homologous and Non-homologous Recombination

Genetic recombination has been divided into two broad categories, homologous recombination and non-homologous (also called illegitimate) recombination. Of the two, homologous recombination has been more intensively investigated and the mechanisms are more thoroughly understood than those for illegitimate recombination. Homologous recombination is defined as being recombinational mechanisms that require extensive regions of sequence similarity to operate although the term 'extensive' can be rather variable; as little as 30 bp in *E. coli* (Gonda and Radding, 1983) and 70 bp in *B. subtilis* (Khasanov et al., 1992) for example. Homologous recombination in Bacteria is also dependent on a functioning *recA* gene. The product of this gene is involved in the regulation of many genes that deal with recombination, mutagenesis, cell division and DNA repair (including the SOS regulon) as well as mediating strand exchange between regions of DNA with similar sequence (Mahajan, 1988; Radding, 1988). Mutations in this

gene abolish homologous recombination in *B. subtilis* (Dubnau, 1993) while in *E. coli*, only the RecE pathway which is encoded on a defective rac prophage can function (Mahajan, 1988). Homologs of *E. coli*'s *recA* gene have been found in every branch of the Bacteria (Karlin et al., 1995) testifying to its ancient origin and implying similar mechanisms operate throughout the domain. Of the various functions performed by homologous recombination in Bacteria, the most important appear to be DNA repair and the recombining of exogenous DNA with the host genome. This latter function is especially important in those organisms that are naturally transformable including *Bacillus* and *Neisseria* but also plays a role in other Bacteria where exogenous DNA can be introduced into the cell through transduction or conjugation. While pathways of homologous recombination are capable of catalyzing certain reactions that lead to DNA rearrangements—amplifications and inversions (Mahajan et al., 1984; Ennis et al., 1987)—their major impact on the evolution of bacterial genomes seems to lie in processes that do not result in genomic rearrangements. DNA repair by its very nature is an attempt by the cell to preserve the sequence of its DNA and only when the repair mechanisms 'get it wrong' are things changed. Even so, the resultant changes are often point mutations with no rearrangement of the DNA. This is often the case with the incorporation of exogenous DNA into the genome. Studies of *E. coli* have found that many genes are mosaics made up of segments of DNA from different strains (Milkman and McKane Bridges, 1990; Maynard Smith et al., 1991; Guttman and Dykhuizen, 1994; Milkman and McKane, 1995). This is despite the fact that *E. coli* populations are generally considered to be clonal (Milkman and McKane Bridges, 1990; Nelson et al., 1991; Nelson and Selander, 1992). Maynard Smith et al. (1993) used multilocus enzyme electrophoresis to examine seven genera of Bacteria and found that the same mechanisms of allelic

replacement were operating although to very different degrees. These are important functions in and of themselves and this emphasis on processes that do not normally involve genomic rearrangements probably means that the consequences of homologous recombination on the evolution of genomes are quite different from those of illegitimate recombination.

Nevertheless, homologous recombination between repeated sequences can have a significant effect on the organization and functioning of genomes. Anderson and Roth (1981) found that spontaneous tandem duplications occurred in *S. typhimurium* between *rrn* operons with a frequency as high as 3% of the population. They suggested that such duplications could provide a selective advantage under conditions conducive to high growth rates by altering the pattern of gene expression. This mirrors the process by which DNA segments can be amplified after an initial duplication with the effect of increasing the amount of some gene product. The numerous chromosome maps of various *Salmonella* strains produced by K.E. Sanderson and colleagues have identified that some strains (*S. typhi* and *S. paratyphi* A) have suffered rearrangements between *rrn* operons (Liu and Sanderson, 1995a; Liu and Sanderson, 1995d). Such a rearrangement between *rrn* operons was also found upon comparing the chromosomes of *Hf. volcanii* and *Hf. mediterranei* (López-García et al., 1995 [chapter 5]).

Illegitimate recombination is defined as recombination events that require only very short stretches of DNA with similar sequence to operate (Albertini et al., 1982; Marvo et al., 1983; Whoriskey et al., 1987) or no sequence similarity at all (Farabaugh et al., 1978; Shimizu et al., 1995) and that do not involve such homologous recombination functions as *recA* (Franklin, 1967). Illegitimate recombination can be further defined by excluding the requirement for any factors encoded by mobile genetic elements or site specific

recombination systems (Allgood and Silhavy, 1988). This type of recombination has been linked to deletions, duplications, inversions and replicon fusions on and between bacteriophages, plasmids, cosmids and the chromosomes of different Bacteria including *B. subtilis*, *E. coli*, and *Spiroplasma citri* (Marvo et al., 1983; Ishiura et al., 1990; Chédin et al., 1994; Yamaguchi et al., 1995; Marais et al., 1996). While illegitimate recombination normally occurs less often and more sporadically than homologous recombination, it can be stimulated by conditions that damage DNA such as irradiation by UV light (Yamaguchi et al., 1995). The label 'illegitimate' reveals the attitude held by many towards this type of recombination; that it is an unintentional consequence of other recombinational processes occurring in the cell. Even so, because of the spectrum of changes illegitimate recombination can potentially catalyze, its impact on genomes over evolutionary time may be quite profound.

Mutations in a variety of genes such as DNA polymerase I (Coukell and Yanofsky, 1970), exonuclease I (Allgood and Silhavy, 1991), topoisomerase III (Whoriskey et al., 1991), H-NS (Lejeune and Danchin, 1990) and DNA gyrase (Ikeda et al., 1982) have been found to affect the frequency of illegitimate recombination suggesting that a number of mechanisms are involved. As mentioned, short stretches of sequence similarity—less than 20 bp and as short as 4 bp—in either direct or inverted orientation often form the endpoints of various illegitimate recombination reactions. Investigations by Ishiura et al. (1989; 1990) on cosmid vectors grown in *E. coli* found that deletions could occur within the cosmid insert using endpoints with sequences similar to the Chi sequence of *E. coli*, a second 8 bp sequence, or strings of adenines, cytosines, guanines or thymines never more than ten nucleotides long. Tandem repeats of certain trinucleotides were also found at or near all of the deletion endpoints looked at. Albertini et al. (1982) found that introducing

mismatches into the repeats used as deletion endpoints could reduce the frequency of recombination at those repeats although recombination was not totally abolished.

Tandem duplications can be catalyzed by illegitimate recombination (Allgood and Silhavy, 1988) and homologous recombination through the RecBCD and RecF pathways (Mahajan et al., 1984). One study investigating duplications involving the *lacI* gene of *E. coli* found that all the duplications examined had endpoints in repeated sequences less than 15 bp long (Whoriskey et al., 1987). A second study also found duplications of the *lacI* gene but no repeats could be found to indicate the endpoints implying that no repeated DNA may have been necessary to generate these duplications (Schaaper et al., 1986). Tandem duplications are of special interest because once they occur, homologous recombination can amplify the repeated DNA. In some cases, the amplification can alter gene expression in a way advantageous to the cell. Two such examples are known to exist in the haloarchaea, HMG coenzyme-A reductase and dihydrofolate reductase. Both enzymes can be amplified in *Hf. volcanii* to produce resistant phenotypes to specific drugs (Rosenshine et al., 1987; Lam and Doolittle, 1992). In neither case have the endpoints of the amplifications been examined so the specific mechanisms by which they occur are unknown.

Inversions and cointegrate formation are two other types of rearrangements that can occur by both homologous and illegitimate recombination. Site-specific inversion systems aside, inversions can have an impact on prokaryotic genomes by reversing the orientation of large numbers of genes as well as introducing two novel joints. Evidence for evolutionarily stable inversions exist between *E. coli* and *S. typhimurium*, in certain other strains of *Salmonella* and within *Hb. salinarum* and *Lactococcus lactis* (Riley and Krawiec, 1987; Liu et al., 1993c; Hackett et al., 1994; Le Bourgeois et al., 1995).

Multiple, nested inversions provide an avenue for the eventual loss of observable conservation between two genomes without the need for exogenous DNA or loss of genetic material. It is known, however, that the chromosome is constrained as to the positions inversion endpoints can have due, it is thought, to nucleoid structure (Mahan et al., 1990; Krug et al., 1994). As well, in *E. coli*, many inversions result in reduced growth rates compared to wild type cells providing a selective disadvantage (Hill and Gray, 1988). This may explain why more inversions are not found when comparing chromosomal maps although these studies did not deal with the occurrence of small (< 10 kbp) inversions. The integration of two replicons by illegitimate recombination has been studied mostly in the context of fusions of λ and pBR322. As with duplications and deletion events short stretches of sequence similarity are often found at the sites of cointegration (King et al., 1982). In some cases additional deletions and duplications have been found adjacent to the cointegration endpoints which may indicate that more than one mechanism is involved (Marvo et al., 1983).

Numerous mechanisms have been proposed to account for the myriad types of rearrangements observed. The effects of mutations in DNA polymerase I and the common presence of direct repeats at deletion and duplication endpoints led to the idea that the polymerase can slip between repeated sequences on the template (Brunier et al., 1989) either skipping the DNA between the repeats or replicating it more than once. The polymerase may even switch templates moving from one strand of DNA to another and back again (Ghosal and Saedler, 1979; Schaaper et al., 1986). Both these phenomena are related in that in each case the polymerase enzyme releases the template DNA it is replicating and binds somewhere else either on the same or a different DNA strand. For this reason both processes are referred to as template-switching. To explain the influence

of gyrase on rearrangements, Ikeda et al. (1982) proposed the subunit exchange model. Gyrase functions as a tetramer to introduce and reseal a double stranded break in DNA. In the model, once a double stranded break is made, one subunit of the gyrase-DNA complex can be exchanged with one subunit of another complex. This exchange is presumed to include one end of each DNA break so that when the DNA is ligated, two novel joints are made. This would introduce a deletion if both gyrase-DNA complexes are on the same replicon or replace a segment of DNA if two such exchanges occurred between two replicons.

In some cases, no repeated sequences of any length are found at rearrangement endpoints. The role of palindromic DNA was elaborated by Glickman and Ripley (1984) as a way to explain this. Palindromic DNA (also DNA forming imperfect palindromes called quasipalindromic) can form cruciform or hairpin structures in DNA. These hairpins can bring sequences that are physically separated on the DNA strand into close juxtaposition. A replication fork could then move through the region and if the hairpin is not denatured, template-switching will result and the DNA making up the hairpin will not be replicated, forming a deletion. Large palindromes in DNA are in fact unstable and are lost very quickly when introduced into either plasmid or chromosomal DNA (Collins, 1980).

Insertion Sequences

Insertion sequences are a special case in genetic recombination as they can be involved in homologous and illegitimate recombination as well as provide their own mechanisms for recombining DNA. The pervasiveness of mobile genetic elements in both prokaryotes and eukaryotes and their unique mode of propagation have ensured their

intensive study. They have been found in most prokaryotic lineages, can occur on chromosomes and plasmids and can in some cases provide the functions necessary for the transfer of themselves and host DNA between cells. Some are even capable of gathering useful genes and integrating them into themselves (hence their name integrons) through a site-specific recombination system while also providing a promoter for the genes' expression (Hall and Stokes, 1993). Of special relevance here is the fact that some haloarchaea— *Hf. volcanii* and some strains of *Hb. salinarum*—are loaded with insertion sequences. For the purposes of this thesis, only a brief overview of insertion sequences will be given in order to illustrate the various ways they can impact on prokaryotic genomes. For a comprehensive review of prokaryotic insertion sequences see Galas and Chandler (1989), Charlebois and Doolittle (1989) and other chapters in the same volume.

The definitive way in which insertion sequences can cause rearrangements in DNA is of course through replicative or non-replicative transposition within or between replicons. In this way insertion sequences form two novel DNA joints at the site of integration which are characterized by a short target site duplication whose length varies with the specific type of insertion sequence. Target sites can be sequence-specific or not, implying that in some cases other features of the DNA such as its topology are more important for recognition. By inserting themselves into novel positions in a host genome, insertion sequences can disrupt the proper functioning of genes (an effect extensively studied in bacteriorhodopsin and gas vacuole mutants of *Halobacterium* [DasSarma, 1989; Pfeifer and Blaseio, 1990]), disturb the local transcriptional context of genes which may alter their expression (see the previous section on the nucleoid and chapter 4), and in some cases provide promoters that can directly influence the expression of downstream genes.

Insertion sequences can also excise from DNA. This process often requires host factors to accomplish, with some insertion sequences of *E. coli* being completely dependent on the host for excision (Allgood and Silhavy, 1988). Precise excision results in the removal of the insertion sequence and a restoration of the original DNA sequence at the target site. Imprecise excision through illegitimate recombination also occurs which can lead to the loss of host sequences flanking the insertion sequence resulting in a deletion.

Apart from the transpositional activity of insertion sequences, they can also be the source of rearrangements through homologous recombination. It has been seen that rearrangements can occur between repeated sequences in prokaryotes, specifically between *rrn* operons, and multiple copies of insertion sequences are no exception (Craig and Kleckner, 1987; Umeda and Ohtsubu, 1990). It is through the homologous recombination between copies of insertion sequences on the *E. coli* chromosome and the F plasmid that Hfr strains are generated (Umeda and Ohtsubu, 1989). The resulting cointegrates are then able to transfer chromosomal DNA between cells at a high frequency. Alternatively, imprecise excision of the cointegrate can give rise to an F' plasmid carrying host sequences. In either scenario, host chromosomal DNA can be transferred to a new cell providing an avenue, along with defective phages carrying host DNA, for the allelic variation found in various bacterial lineages (Maynard Smith et al., 1993).

Chapter 2

Materials and Methods

In preparation for DNA extraction, *Hf. volcanii* was grown in 10 mL or 12 mL volumes of low salt (LS) medium, grown at 37°C and shaken at 90-100 rpm in a reciprocating incubator. *Hb. salinarum* was grown under the same conditions but in high salt (HS) medium. In both cases, cultures grew to saturation in 2-3 days. For maintaining cultures for short periods of time, cells were plated onto either LS or HS medium containing 1.5% agar, incubated at 37°C for 2-3 days, and stored in bags at room temperature. For longer term storage, strains were grown in the appropriate medium to saturation, divided into 1 mL aliquots and mixed with 500 μL of 50% glycerol, and frozen at -80°C. Recipes for LS and HS media are given in Appendix A.

Cosmid libraries of genomic DNA had previously been prepared for the *Hb. salinarum* GRB mapping project (chapter 3). For the completion of this project and the two genomic comparisons described, it was necessary to isolate cosmid DNA from the GRB library and the cosmid library of *Hf. volcanii* DS2 (Charlebois et al., 1991). Cosmids were maintained in *E. coli* ED8767 grown in YT medium (Messing, 1983) with 30 μg/mL kanamycin sulfate. For long term storage, cells were frozen at -80°C as described for haloarchaeal strains except that YT medium was used. For extraction of cosmid DNA, cells were plated onto YT + kanamycin plates and grown overnight at 37°C. A single colony for each cosmid was picked and grown in a 12 mL YT + kanamycin culture at 37°C in a reciprocating incubator at 90-100 rpm overnight. The alkaline extraction

protocol used for the isolation of cosmid DNA is given in Appendix A. After extraction, cosmid DNA was routinely checked by digestion with the cloning enzyme and comparing the pattern of bands on an agarose gel to the pattern obtained when the cosmid was originally isolated.

Total genomic DNA was isolated from *Hb. salinarum* GRB and *Hf. volcanii* DS2 for the GRB map and genomic comparisons (chapters 3, 5, and 6). Cells were grown in 4 mL cultures of either LS or HS medium at 37°C for 2-3 days at 90-100 rpm. One mL aliquots of the cultures were transferred into 2 mL polypropylene tubes for the extraction procedure (Appendix A). Starting with smaller volumes for this protocol is preferred because centrifugation times can be unacceptably long for volumes of only 10 mL.

Digests of cosmid and total haloarchaeal DNA were routinely performed in 10 µL volumes, either in 0.5 mL polypropylene tubes or in 96 well microtiter plates. Cosmid digests used 2 units of enzyme and 400-500 ng of DNA. Total haloarchaeal DNA digests increased the amount of enzyme to 4 units. In both cases, reduced amounts of *Bam*HI were used (0.5-1 unit) to prevent star activity produced by this enzyme. Digestions were carried out at the appropriate temperature in a water bath for digests in tubes or an incubator for digests in microtiter plates. DNA for hybridization probes was digested for one hour while DNA to be run on agarose gels was digested four hours.

DNA fragments isolated from agarose gels were used to probe Southern and dot blots for the *Hb. salinarum* GRB map (chapter 3) and the GRB-*Hf. volcanii* DS2 genomic comparison (chapters 6). DNA fragments used for these purposes were invariably from cosmids belonging to the genome libraries of either *Hb. salinarum* or *Hf. volcanii*. For the GRB map, cosmids were run on FMC SeaPlaque GTG agarose gels. DNA frangments were cut from the gels, stored at 4°C, and melted in a water bath just prior to use.

Digested cosmid DNA for the genomic comparison was run on agarose gels ranging between 0.9% to 1.5% depending on the size of the DNA fragment to be isolated in TAE buffer. DNA was recovered using the GeneClean kit (Bio 101), the protocol for which is listed in Appendix A. Isolated DNA was stored at -20°C.

Total DNA was extracted from *Hb. salinarum* GRB to be used for PFGE in the construction of the GRB map (chapter 3). Total DNA was extracted according to the protocol given in Appendix A. Digestion of DNA in agarose plugs was done by incubating the plugs in restriction enzyme buffer for 20 minutes on ice. The buffer was removed and replaced with fresh buffer and 20 units of enzyme. Plugs were incubated on ice for 30 minutes and then at the enzyme's optimum temperature for 3 hours. PFGE was performed using a Tyler Research Instruments MB-10 Megabase electrophoresis system. Gels were run at 170 volts for 20-24 hours at 11°C with switching times of 15 seconds to 30 seconds depending on the size range of the DNA to be resolved. Gels were 1% agarose in TBE buffer. DNA markers for PFGE gels were Promega's Δ39 λ ladder.

Southern blots of cosmid and total haloarchaeal DNA—including PFGE DNA—were prepared for the completion of the GRB map (chapter 3) and the two comparisons (chapters 5 and 6). Cosmid DNA was digested as described above. In most cases, the cloning enzyme—either *Bam*HI or *Mlu*I—was included in the digest to produce a vector band of 5.4 kbp in size. Digests were electrophoresed in 0.9% agarose gels for 18-22 hours at 1.5 volts/cm. Gels were then stained with ethidium bromide (EtBr) and visualized on an ultra violet light box. A lane for haloarchaeal genomic DNA cut with the same enzyme as the cosmid DNA was often included as a positive control for hybridizations where appropriate. Markers of λ DNA digested with *Bst*EII, *Xba*I, and *Xho*I served as negative controls. Southern blots were prepared using GeneScreen nylon membrane

(Dupont) and a Tyler Research Instruments VT-20 vacuum transfer unit as described in Appendix A.

Dot blots of cosmid libraries were prepared during the construction of the *Hb. salinarum* GRB map (chapter 3) and the comparison between the genomes of GRB and *Hf. volcanii* DS2 (chapter 6). Fifty ng (genomic comparison) or 100 ng (GRB map) of each cosmid was mixed with NaOH to provide a final concentration of 0.4 M NaOH. Three uL aliquots were then spotted onto GeneScreen nylon membranes. Equi-molar amounts of λ and Lorist M DNA (the cosmid vector [Charlebois et al., 1989]) were included on the dot blots as negative controls. The membranes were then rinsed in 2X SSC and allowed to air dry. Finally, the membranes were irradiated in an ultra violet light box for 5 minutes and stored in plastic bags when not in use.

Hybridizations of haloarchaeal DNA to Southern (chapters 3 and 5) and dot blots (chapter 6) were performed as per the protocol in Appendix A. Membranes to be used more than once were stripped between each hybridization. Membranes were incubated in stripping solution (5 g SDS, 0.46 mL 1M Na_2HPO_4, 4.06 mL 1M NaH_2PO_4, 500 mL formamide, 480 mL distilled water) at 70°C for one hour in a Tyler Research Instruments HI-16000 hybridization incubator, rinsed with 2X SSC and exposed to X-ray film at -80°C with an intensifying screen for at least one day to assess the effectiveness of the stripping. After stripping and exposure to film, membranes were stored in plastic bags at room temperature.

CHAPTER 3

Physical map and set of overlapping cosmid clones representing the genome of the archaeon *Halobacterium* sp. GRB

This chapter is published as:

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

Abstract

We have constructed a complete, five-enzyme restriction map of the genome of the archaeon *Halobacterium* sp. GRB, based on a set of 84 overlapping cosmid clones. Fewer than 30 kbp, in three gaps, remain uncloned. The genome consists of five replicons: a chromosome (2038 kbp) and four plasmids (305, 90 , 37, and 1.8 kbp). The genome of *Halobacterium* sp. GRB is similar in style to other halobacterial genomes by being partitioned among multiple replicons and by being mosaic in terms of nucleotide composition. It is unlike other halobacterial genomes, however, in lacking multicopy families of insertion sequences.

My Contribution

I prepared Southern blots of cosmid DNA run on regular agarose gels and genomic DNA run on pulsed-field gels. By hybridizing these blots with cosmid DNA and gel isolated-fragments, I completed and confirmed the contig map that was already at an

advanced stage. I then prepared Southerns of the minimal cosmid set and performed more hybridizations on these. In this way, I placed most of the genes currently identified on the GRB map. I also performed all 38 hybridizations done in the search for repeated sequences in the GRB genome. I assisted in the editing of this manuscript which was written by R.L. Charlebois.

Introduction

Halobacterium salinarium (*Hb. halobium*, *Hb. cutirubrum*), an extremely halophilic archaeon, has been a model organism for the study of instability in prokaryotic genomes (Pfeifer, 1988; DasSarma, 1989; Charlebois and Doolittle, 1989). Insertion elements are numerous and active (Sapienza and Doolittle, 1982;Sapienza et al., 1982), leading to mutation in phenotypic markers at frequencies of up to 10^{-2} per cell per generation (Weidinger et al., 1979; Pfeifer et al., 1981b). Moreover, recombination between repeated copies of the elements occurs so frequently in some strains that a consistent and reproducible map of the genome might not exist (Sapienza et al., 1982; Pfeifer, 1988; Pfeifer et al., 1989). Important forces must therefore be acting to limit damage to the genome and to maintain some measure of genetic integrity.

The disruptive nature of frequent transpositions and rearrangements seems to be partly mollified by the preference of insertion elements for regions of the genome such as plasmids and compositionally distinct islands inserted within the chromosomal DNA (Pfeifer et al., 1982; Ebert and Goebel, 1985; Pfeifer and Betlach, 1985). These subgenomic compartments are of different nucleotide composition than the bulk of the DNA, and they form a satellite fraction termed FII when DNA is separated according to G+C content (Joshi et al., 1963; Moore and McCarthy, 1969; Pfeifer et al., 1982). FII

DNA is of lower mol% G+C (58%) than is main fraction (FI) DNA (68% G+C), and it represents about 10-30% of the total *Hb. salinarium* genome, depending on the strain. Insertion element activity is not confined to these regions of FII DNA, but it is concentrated there (Pfeifer et al., 1983; Pfeifer, 1988).

Analysis of FII DNA from *Haloferax volcanii*, a more genetically stable extreme halophile, revealed that FII has markedly different restriction enzyme site frequencies compared with FI DNA, probably reflecting different oligonucleotide composition, and suggesting different evolutionary origins (Charlebois et al., 1991). As in *Halobacterium*, insertion elements in *Hf. volcanii* favour FII DNA and are abundant there (Hofman et al., 1986; Cohen et al., 1992; Schalkwyk et al., 1993).

Several hypotheses have been put forth to explain the existence of FII DNA and to justify the preference of insertion elements for it (Pfeifer, 1988; Charlebois and Doolittle, 1989; Charlebois et al., 1991; Schalkwyk et al., 1993). The target sequence of the elements may in general be A+T-rich (Pfeifer, 1988; Charlebois and Doolittle, 1989), thus increasing the chances that the elements will transpose within the less G+C-rich FII DNA fraction. Since halobacterial insertion elements themselves tend to be A+T-rich in sequence (Charlebois and Doolittle, 1989; DasSarma, 1989), it is also possible that FII DNA is nothing but an accumulation of decaying insertion elements, built one upon the other. If, on the other hand, there tends to be no biased target specificity, insertion elements might still accumulate in FII DNA if it is inherently gene-poor, because of selection against insertion within (gene-rich) FI DNA (Pfeifer et al., 1983). (However, an analysis of gas vesicle gene disruption by insertion elements [Pfeifer et al., 1989] showed that the plasmid-encoded p-*vac* [in FII DNA] was inactivated more than a thousand times more frequently than the chromosomally-encoded c-*vac* [in FI DNA]. Thus what controls

the rate of insertional inactivation is the context, not the genetic function.) Finally, it is possible that FII DNA plasmids may have been imported from another unrelated organism by some form of mating, infection, or transformation process. In support of this latter hypothesis are the observations that gas vesicles and ferredoxin are remarkably similar between halobacteria and cyanobacteria (Walker et al., 1984; Pfeifer et al., 1993), and that the gas vesicle genes often map to plasmid DNA (Weidinger et al., 1979; Simon, 1978; DasSarma et al., 1987). If halobacterial insertion elements for some reason found this foreign DNA attractive, and if halobacteria have the ability to transfer DNA among themselves horizontally (as is suggested by the results of Mevarech and coworkers [Mevarech and Werczberger, 1985; Rosenshine et al., 1989; M. Mevarech, pers. comm.]), the spread of FII—accumulating insertion elements with each visit to a new strain—would seem inevitable.

Whatever the scenario, it is clear that *Hb. salinarium* is plagued by numerous insertion elements which have important consequences on the evolution of its genome. In order to study the origin of this situation further, we decided to examine the structure of the genome of a genetically stable close relative of *Hb. salinarium*, *Halobacterium* sp. GRB. This strain is known to possess none of the characterized halobacterial insertion elements (Ebert et al., 1986), but otherwise it is a good representative of the genus, containing purple membrane (bacteriorhodopsin) and gas vesicles (Ebert et al., 1984). With its genomic stability and a lack of restriction systems, the strain shows promise for genetic work (Soppa and Oesterhelt, 1989). One barrier to its usefulness—an incomplete resistance to a halocin that it produces—has already been overcome with the finding of a mutant deficient in halocin production (A. St. Jean, M.R. Rajab, and R.L. Charlebois, unpublished).

In this paper, we present a 257-site restriction map of the 2.47-Mbp *Halobacterium* sp. GRB genome constructed from maps of 84 overlapping cosmid clones, supplemented with pulsed-field gel information. We elucidate the structure of the extensive satellite DNA regions in this genome, and we demonstrate an unusual impoverishment of repeated sequences within them.

Materials and Methods

Strains and media

Halobacterium sp. GRB was obtained from W. Goebel (Universität Würzburg) via W.F. Doolittle (Dalhousie University). It was grown in a medium previously described for *Halobacterium halobium* (Cline and Doolittle, 1987). *Escherichia coli* ED8767 was grown in YT medium (Messing, 1983). Cosmid-bearing clones were grown in YT supplemented with 30 mg/L kanamycin sulfate.

Bam*HI and* Mlu*I cosmid libraries*

Total genomic DNA was isolated from a 1-L culture of *Hb.* sp. GRB in the late-log phase of growth. Cell pellets were resuspended in a total of 4 mL of medium salts solution. To the pooled resuspension in a 1-L flask, 40 mL of lysis buffer (50 mM Tris•Cl [pH 7.6]; 50 mM EDTA, 1% SDS, 0.15 mg/mL proteinase K) was added and the mixture allowed to incubate at 55°C for 2 h. Two gentle phenol extractions and two phenol:chloroform (1:1) extractions were followed by ethanol precipitation. The pellet was washed twice with ethanol, then allowed to dry partially before dissolving in TE (10 mM Tris•Cl [pH 7.6], 1 mM EDTA).

Genomic DNA partially-digested with *Bam*HI or with *Mlu*I, then treated with alkaline phosphatase, was ligated to appropriate arms of the Lorist M cosmid vector (Charlebois et al., 1989), then packaged with Stratagene's Gigapack II-XL packaging extract following Stratagene's instructions. *E. coli* ED8767 served as the host.

Cosmid DNA extraction

Colonies were picked from the *Bam*HI library and grown in 3 mL of YT-kanamycin broth. A 0.5-mL portion of each of the 2112 cultures (clone designations G3A1-G24H12) was frozen (-80°C) with 15% v/v glycerol.

Cosmids were isolated by an alkaline extraction procedure, adapted for processing 192 samples at a time in Titertek tubes (Biorad), 1-mL capacity tubes in an 8 X 12 array. Larger DNA stocks of selected cosmid clones from either the *Bam*HI or the *Mlu*I library were prepared from 10-15 mL cultures in regular microcentrifuge tubes. Cultures were always inoculated with fresh colonies.

Pulsed-field gel electrophoresis

Hb. sp. GRB was embedded in agarose by mixing a cell suspension with an equal volume of 1% FMC SeaPlaque GTG agarose, 1 M NaCl at 55°C, then pouring the mixture onto a cast acrylic plate. Once hardened, plugs were prepared by slicing with a glass coverslip. The agarose-embedded cells were lysed at 55°C in 1 M Tris base, 0.5 M EDTA, 1% N-lauroylsarcosine, then soaked in TE, 1 M NaCl on ice with four changes of buffer, then again twice with TE. Plugs were equilibrated in restriction enzyme buffer on ice, then after replacing the buffer, enzyme was added and allowed to diffuse into the plug on ice, before incubation for three hours at the enzyme's optimal temperature. Pulsed-field

gels were run on the Tyler Research Instruments' (Edmonton, Alberta) MB-10 Megabase electrophoresis system. Gels were 1% agarose in TBE buffer, and ran at 11°C for 20-24 h at 170V with constant switching of 15 or 30 s, depending on the size range to be resolved. For size markers, we used Promega's lambda D39 ladder.

Blot preparation

Gels containing digests of cosmid or genomic DNA were transferred to GeneScreen membrane (DuPont) with a Tyler VT-20 vacuum transfer unit, following Tyler's protocols.

The *Mlu*I library was used for chromosome walking. Cosmid DNA was isolated from forty-one clones (designated G25A1-G25D5) and examined on a gel in order to assess the library before proceeding with the preparation of colony blots for chromosome walking. 760 colonies (designated G26A6 to G30N9) were patched onto YT-kanamycin plates in six replicas using toothpicks, 152 colonies per plate. Once grown, cells from five of the replicas were transferred to Colony/Plaque Screen membrane (DuPont) according to the manufacturer's booklet using the autoclaving option for fixing and denaturing the DNA. We also prepared DNA dot blots from clones G25A1-G25D5, by spotting about 100 ng of cosmid DNA in 0.4M NaOH onto GeneScreen membrane, then rinsing the blots in 2X SSC (0.3M NaCl, 0.035M trisodium citrate) before use.

Before the first hybridization, all blots except colony blots were irradiated with ultraviolet light to fix the DNA onto the membranes.

Hybridization conditions

Hybridizations and washes were performed in a Tyler HI-12000 hybridization

incubator. Blots were prehybridized for about 2 hours at 38-40°C in 1 M NaCl, 50 mM Tris•Cl [pH7.6], 5% w/v SDS, 50% v/v formamide, and 50 µg/mL sheared and denatured herring sperm DNA. Probe was prepared by the random priming method, on entire restriction-digested cosmids, or on melted slices of FMC SeaPlaque GTG agarose gel containing DNA fragments. Hybridization was done for 1-2 hr at 38-40°C. Blots were then rinsed in 2X SSC, washed in 2X SSC for 30 minutes at room temperature, then washed in 2X SSC, 1% SDS for 60 minutes at 65-70°C. Washes were carried out in the hybridization incubator.

Blots were stripped for reuse in 25 mM (for sodium) sodium phosphate [pH7.2], 0.5% w/v SDS, and 50% v/v formamide, for 60-90 minutes at 65-70°C.

Results

The landmark strategy for physical mapping of genomes

The landmark strategy for finding cosmid clone overlaps has been described previously (Charlebois et al., 1989; Charlebois et al., 1991; Charlebois, 1993). First, a cosmid library is constructed from genomic DNA partially digested with a restriction enzyme with an average site frequency of about 5 kbp. Subsequent digestion of a cosmid with the cloning enzyme thus yields about ten fragments, one of which is the cosmid vector. This digest is run alongside a double digest using the cloning enzyme and another, less frequently cutting enzyme. A restriction landmark is identified as a cloning enzyme fragment which is cut into distinctive subfragments by the infrequently-cutting enzyme. Since the infrequent cutter is chosen to have between 50-150 sites in the genome, each of its sites will be located within a cloning-enzyme fragment of a different size, or at least at a different position within different cloning-enzyme fragments of the same size. When two

cosmid clones contain the same landmark (e.g. a 5.2 kbp *Bam*HI fragment cut by *Hin*dIII into subfragments of 4.1 and 1.1 kbp), their overlap is virtually proven (Lander and Waterman, 1988). The sensitivity of overlap detection is a function of the density of landmarks, and thus can be increased by examining more double digests for each clone. A compromise has to be reached where the advantage of analysing more clones balances the extra time and cost involved.

Mapping the genome of Halobacterium sp. GRB

We constructed a *Bam*HI partial-digest cosmid library of *Hb.* sp. GRB DNA from which to sort clones by landmark analysis. Here, instead of examining ten double digests initially as in the *Hf. volcanii* mapping project (Charlebois et al., 1989; Charlebois et al., 1991), we first compared the *Bam*HI single digests with only two double digests—*Bam*HI/*Eco*RI and *Bam*HI/*Hin*dIII—before embarking on the more thorough analysis. Using this two-step approach (Charlebois, 1993), we were able to screen a large library of 2112 clones or 40 genome equivalents, with statistical (though not biological [Charlebois, 1993]) assurance of covering 100% of the genome (Lander and Waterman, 1988).

*Eco*RI cuts the genome of *Hb.* sp. GRB rather frequently (Å150 sites), giving a high density of landmarks. *Hin*dIII landmarks are half as common, but are more distinctive. Together, these two sets of landmarks were able to eliminate 90% of the redundancy in the library, by allowing clones to be sorted around 69 different distinctive landmarks. (An additional group of unrelated clones had no landmarks for either *Eco*RI or *Hin*dIII.) None of this initial landmark analysis required accurate sizing of fragments or use of a computer. The landmarks are distinctive enough to match cosmid clones almost unambiguously by eye.

We digitized the fragment sizes (Charlebois et al., 1989) from the smaller set of about 200 clones, to link up those cosmids sharing landmarks too obscure to be detected by eye, and to verify our previous assignments. Additional landmarking digests (*Bam*HI alone, and together with *Afl*II, *Ase*I, *Dra*I, *Ssp*I, and *Xho*I) were performed and analysed (Charlebois et al., 1989) to link the contigs further, resulting at this point in ten contigs plus three closed circles (the plasmids pGRB305, pGRB90 and pGRB37). No cosmids were free of landmarks.

The initial effort to link the ten contigs was done by hybridizing cosmid DNA from each of the contig ends to *Bam*HI digests of the collection of end cosmids. This only identified one overlap not detected by landmark analysis. Nine uncloned regions (gaps) thus remained.

Next, we prepared a second cosmid library for chromosome walking, using the cloning enzyme *Mlu*I. We hybridized colony blots with gel-isolated fragments from contig ends or from regions of the genome uniquely represented in our *Bam*HI library. Putative gap-filling, contig-extending, or redundancy-generating clones were confirmed by restriction mapping, landmark analysis, or by Southern hybridization of restriction digests. After this step, three chromosomal contigs remained.

Cosmid clones representing the minimal set, and some others, were mapped by single and double digests for *Afl*II, *Ase*I, *Dra*I, *Hin*dIII and *Ssp*I. Small fragments were resolved in 1.8% agarose gels. When the map was insufficient to set the exact size of the overlap between clones (we mapped the sites for a subset of the landmark enzymes),

overlaps were determined by hybridization to *Bam*HI digests, by *Bam*HI partial-digest mapping, or by mapping sites for other enzymes (like *Bgl*II, *Eco*RI and *Xho*I).

We observed widely varying colony sizes in our cosmid clone libraries, undoubtedly reflecting *E. coli*'s distaste for certain halobacterial sequences. Since we could only pick colonies of visible size, those clones more severely inhibiting the growth of *E. coli* were missed. Therefore, we doubted that a continued hunt for the three missing pieces of the genome, as cosmid clones, would be successful. We thus turned to pulsed-field gels to link the contigs and to estimate the size of the gaps. Total DNA, or a gel-isolated fragment, from cosmid clones representing both ends of each contig were hybridized to blots of pulsed-field gels (as in Charlebois et al., 1991) containing digests (*Afl*II, *Ase*I, *Dra*I, *Hin*dIII, and *Ssp*I) of *Hb.* sp. GRB genomic DNA. The six contig ends paired to give an unambiguous map. Gap sizes were estimated by subtracting from the hybridizing fragment sizes, the distances to those sites from each end of the linked contigs.

Finally, we verified that all regions of the genome were covered by independent, redundant clones. Seven areas for which redundancy was insufficient were verified by finding the predicted hybridization pattern on blots of pulsed-field gels (as in Charlebois et al., 1991). This assured us that our map was correct at the coarse level. It was also necessary to worry about smaller rearrangements—usually deletion—within cosmid clones caused by the selective disadvantage of some sequences. Though some cosmids did exhibit signs of instability, producing substoichiometric bands in gels of restriction digests, we minimized the chance for rearrangement by growing cultures promptly and in small volumes. The high degree of redundancy in our library permitted us to identify the odd rearranged cosmid and to exclude it from our map.

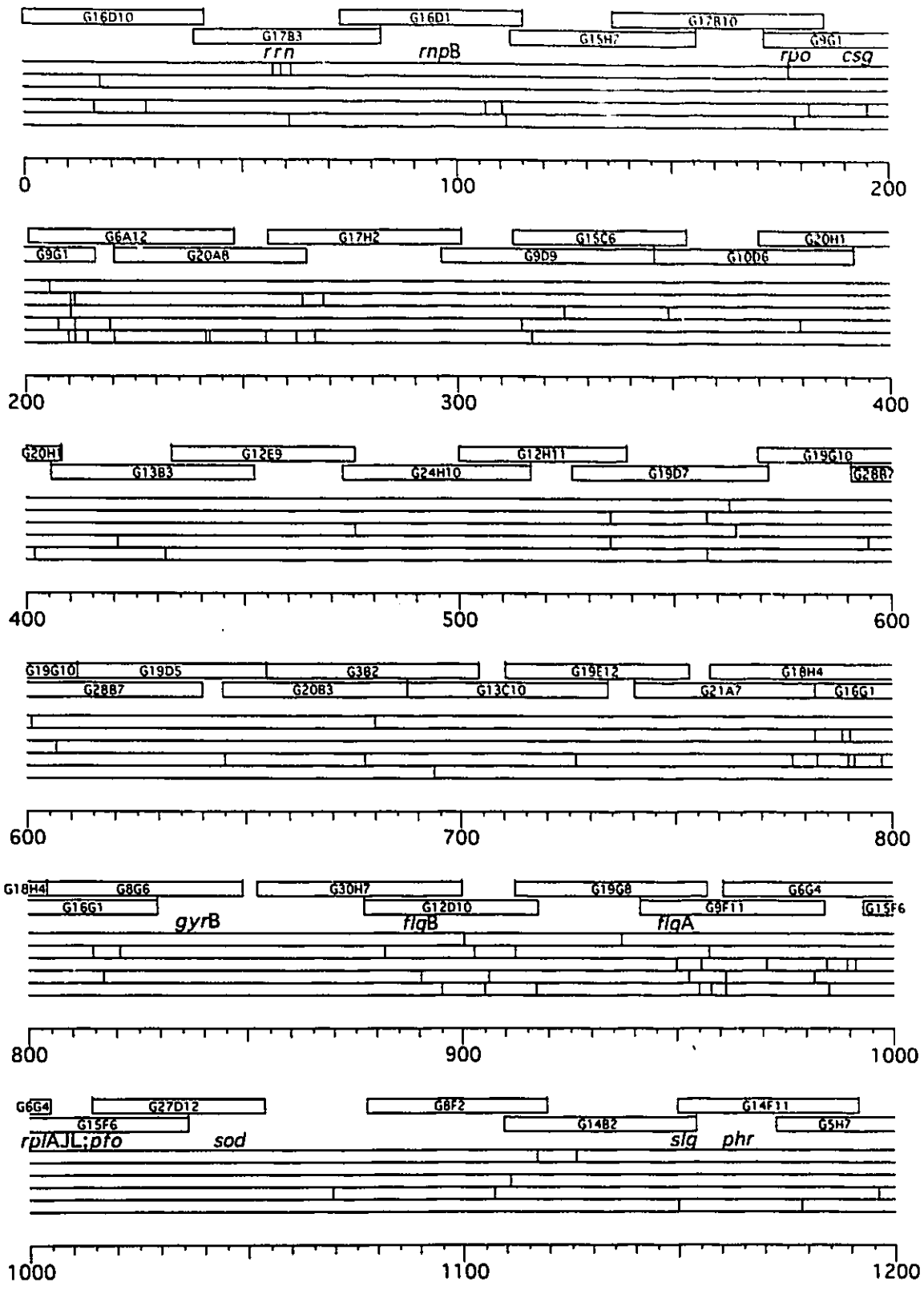
The structure of the Halobacterium sp. GRB genome

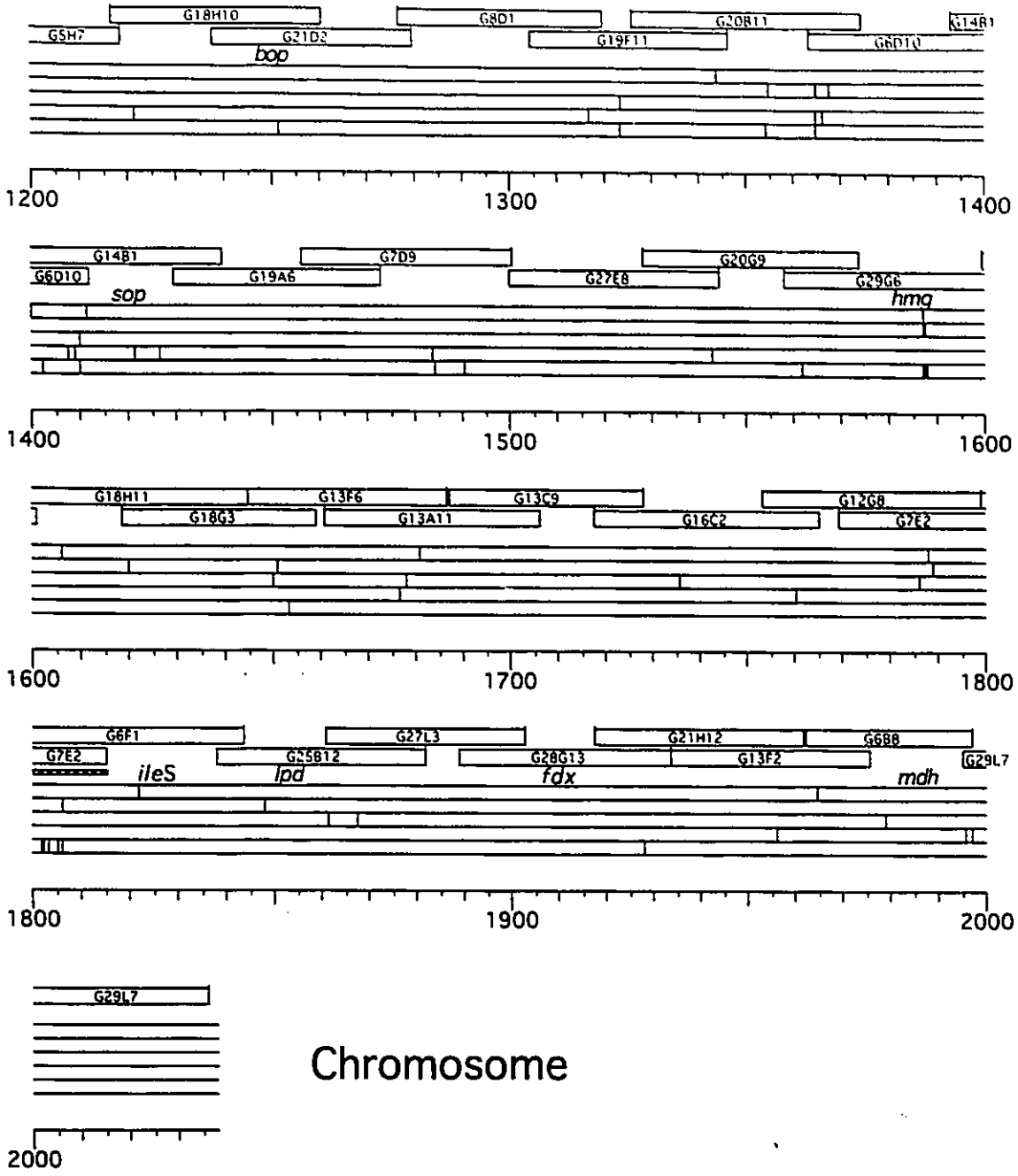
Using the landmark strategy, we produced a complete physical map of the *Halobacterium* sp. GRB genome (Fig. 3.1). The map is based on a set of 84 overlapping cosmid clones, representing 98.8% of the genome, supplemented in three places with information obtained from hybridizations to blots of pulsed-field gels. On this map, we placed 257 restriction sites (summarized in Table 3.1) and eighteen previously-cloned protein-coding genes, the single rRNA operon, and the gene encoding RNaseP RNA (Table 3.2).

The genome of *Hb.* sp. GRB is partitioned into five replicons: a chromosome of 2038 kbp, and four plasmids of 305, 90, 37, and 1.8 kbp. The chromosome of *Hb.* sp. GRB is of similar size to that from *Hb. salinarium* (Ng and DasSarma, 1993), smaller than the 2.9 Mbp chromosome present in *Haloferax* spp. (Charlebois et al., 1991; López-García et al., 1992). As in many other halobacteria (Ebert et al., 1984; Gutiérrez et al., 1986; Charlebois et al., 1991; Pfeifer et al., 1981a), plasmids are responsible for an important proportion of the genomic map, in this case 17.6%.

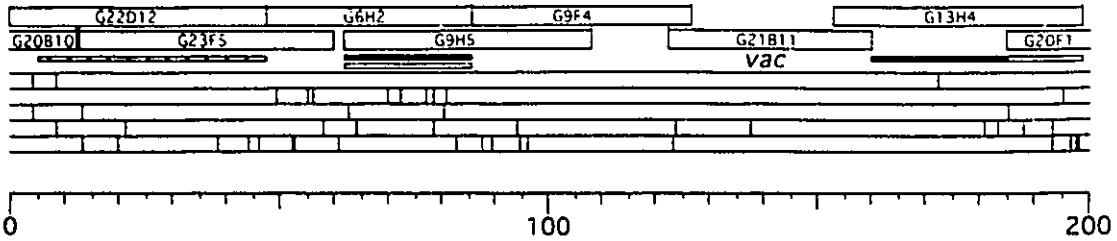
In 1984, Ebert *et al.* characterized extrachromosomal DNA populations in various new isolates of halobacteria (Ebert et al., 1984). From within *Hb.* sp. GRB, four species of cccDNA (covalently-closed circular DNA) were identified: 1.6 kbp, 35 kbp, 65 kbp and an unsized but large "minor cccDNA". The small (actually 1.8 kbp) high-copy number plasmid has since been cloned and characterized (Hackett et al., 1990) and shows remarkable heterogeneity in sequence (Kagramanova et al., 1989; Akhmanova et al., 1993).

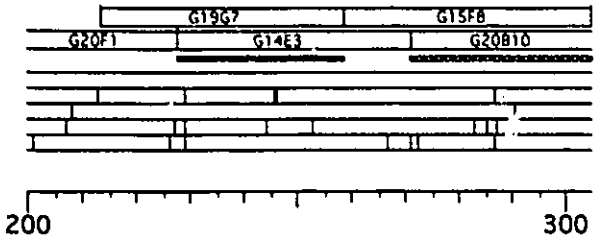
Fig. 3.1. Physical map of the *Halobacterium* sp. GRB genome with genes and repeated sequences. The staves indicate the positions of restriction sites for (from top to bottom) *Afl*II, *Ase*I, *Dra*I, *Hin*dIII, and *Ssp*I. The cosmid clones are shown above as boxes; distance in kilobase pairs is indicated below. Bars above the staves show the locations of the five cross-hybridizing regions in the genome. Resolution of the cloned portions of the map is better than 0.5 kbp and the accuracy of intersite distances is better than 2%. Sizes of uncloned gaps in the chromosome (3 kbp at pos. 849-852 kbp; 24 kbp at pos. 1054-1078 kbp; and 2 kbp at pos. 2036-0 kbp) are accurate to within about 10 kbp. The origin of the map (pos. 0) was chosen arbitrarily.


G16D10
G16D1
G17B10
G17B3
G15H7
G9G1
rrn
rnpB
rpo
csg
0
100
200
G6A12
G17H2
G15C6
G20H1
G9G1
G20A8
G9D9
G10D6
200
300
400
G20H1
G12E9
G12H11
G19G10
G13B3
G24H10
G19D7
G28B7
400
500
600
G19G10
G19D5
G3B2
G19E12
G18H4
G28B7
G20B3
G13C10
G21A7
G16G1
600
700
800
G18H4
G8G6
G30H7
G19G8
G6G4
G16G1
G12D10
G9F11
G15F6
gyrB
flgB
flgA
800
900
1000
G6G4
G27D12
G8F2
G14F11
G15F6
G14B2
G5H7
rplAJL;pfo
sod
slg
phr
1000
1100
1200


G18H10
G8D1
G20B11
G14B1
G5H7
G21D2
G19F11
G6D10
bop
1200
1300
1400
G14B1
G7D9
G20G9
G6D10
G19A6
G27E8
G29G6
sop
hmq
1400
1500
1600
G18H11
G13F6
G13C9
G12G8
G18G3
G13A11
G16C2
G7E2
1600
1700
1800
G6F1
G27L3
G21H12
G6B8
G7E2
G25B12
G28G13
G13F2
G29L7
ileS
lpd
fdx
mdh
1800
1900
2000
G29L7
2000

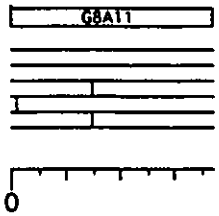
Chromosome



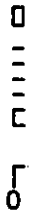


pGRB305

pGRB90



pGRB37



pGRB1

Table 3.1. Restriction enzyme site statistics for each of the replicons making up the *Halobacterium* sp. GRB genome. Shown are the number (#) of restriction sites for each of the mapped enzymes, with mean fragment sizes in kbp. The five-enzyme site density in the two FII plasmids pGRB305 and pGRB90 is significantly ($p<<.0005$) higher than all 395-kbp windows within the chromosome, based on a χ^2 test(Churchill et al., 1990).

	chromosome 2038 kbp	pGRB305 305 kbp	pGRB90 90 kbp	pGRB37 37 kbp	pGRB1 1.8 kbp	genome 2472 kbp
*Afl*II #	21	3	1	0	0	25
mean	97.0	101.6	90.2	-	-	98.9
*Ase*I #	27	14	4	0	0	45
mean	75.5	21.8	22.6	-	-	54.9
*Dra*I #	22	7	2	1	0	32
mean	92.6	43.5	45.1	37.3	-	77.3
*Hin*dIII #	46	20	9	1	0	76
mean	44.3	15.2	10.0	37.3	-	32.5
*Ssp*I #	46	25	6	1	1	79
mean	44.3	12.2	15.0	37.3	1.8	31.3
Total #	162	69	22	3	1	257
mean	12.6	4.4	4.1	12.4	1.8	9.6

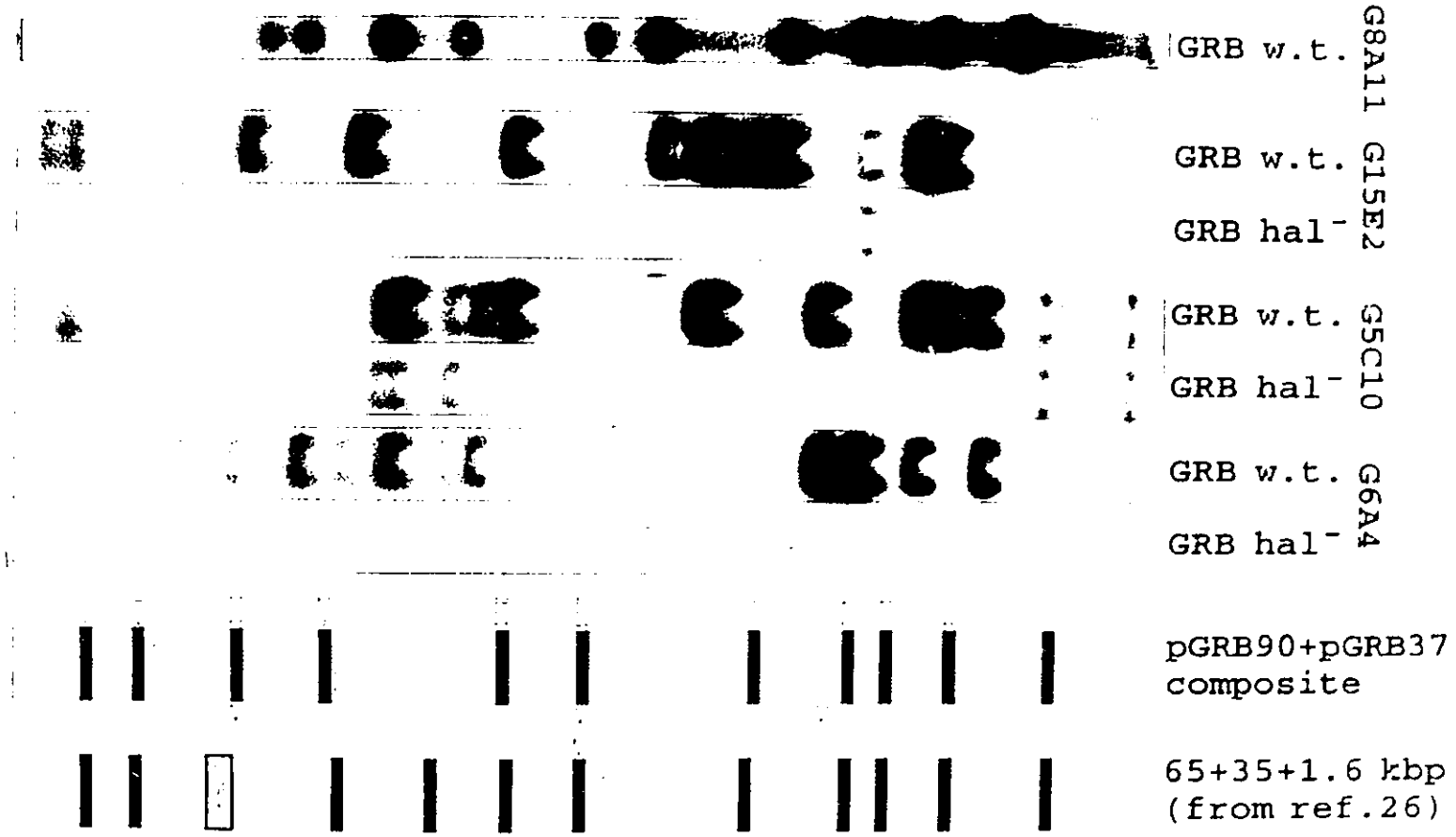
Table 3.2. Genes mapped, and their source. *Ha.=Haloarcula*; *Hb.=Halobacterium*; *Hf.=Haloferax*.

Gene/ locus	Identity	Location (kbp)	Cosmid clones	Probe source/reference
bop	bacterio-opsin	1238-1261	G18H10/ G21D2	*Hb. halobium* (Betlach et al., 1983)
csg	cell-surface glycoprotein	185- 201	G9G1	*Hb. halobium* (Lechner and Sumper, 1987)
fdx	ferredoxin	1905-1934	G28G13	*Hb. salinarium* (Pfeifer et al., 1993)
*flg*A	flagellin operon A	942- 957	G19G8/ G9F11	*Hb. halobium* (Gerl and Sumper, 1988)
*flg*B	flagellin operon B	877- 900	G30H7/ G12D10	*Hb. halobium* (Gerl and Sumper, 1988)
*gyr*B	gyrase subunit B	830- 849	G8G6	*Hf.* sp. Aa2.2 (Holmes et al., 1991)
hmg	HMG-CoA reductase	1574-1600	G29G6	*Hf. volcanii* (Lam and Doolittle, 1992)
*ile*S	isoleucyl-tRNA synthetase	1815-1838	G6F1	*Hf. volcanii* (C.J. Daniels, pers. comm.)
lpd	dihydrolipoamide dehydrogenase	1852-1882	G25B12	*Hf. volcanii* (Vettakkorumakankav and Stevenson, 1992)
mdh	malate dehydrogenase	1976-1995	G6B8	*Ha. marismortui* (Cendrin et al., 1993)
pfo	pyruvate:ferredoxin oxidoreductase	1005-1036	G15F6	*Hb. halobium* (Plaga et al.. 1992)
phr	photolyase	1154-1183	G14F11	*Hb. halobium* (Takao et al., 1989)
*rnp*B	RNase P RNA	83- 112	G16D1	*Hf. volcanii* (Nieuwlandt et al., 1991) and *Hb. cutirubrum* (C.J. Daniels pers. comm.)
*rpl*AJL	ribosomal proteins L1, L10, L11	1005-1036	G15F6	*Hb. halobium* (Itoh, 1988)
rpo	RNA polymerase operon	171- 185	G17B10/ G9G1	*Hb. halobium* (Leffers et al., 1989)
rrn	ribosomal RNA operon	41- 73	G17B3	*Hf. volcanii* (Charlebois et al., 1989)
slg	*sod*-like gene	1150-1154	G14B2/ G14F11	*Hb.* sp. GRB (Joshi and Dennis, 1993)
sod	superoxide dismutase	1036-1054	G27D12	*Hb. cutirubrum* (Joshi and Dennis, 1993; May et al., 1989)
sop	sensory opsin	1412-1440	G14B1	*Hb. halobium* (Blanck et al., 1989)
vac	gas vacuole	pGRB305 126-153	G21B11	*Hb. halobium* (Horne et al., 1988)

We were faced with two discrepancies regarding the plasmid complement of the *Hb.* sp. GRB genome, relative to this published work. The first was a minor one, likely reflecting differences in sizing methods, for the 35 kbp plasmid (Ebert et al., 1984) which we size at 37 kbp. The difference between the size estimate of 65 kbp for the next largest plasmid (Ebert et al., 1984) and our estimate of 90 kbp, however, concerned us more. We felt it important to verify that the plasmids had not changed in the years between *Hb.* sp. GRB's original isolation and our own genomic mapping work. Therefore, since *Pst*I fragment patterns of the 35- and the 65-kbp plasmids were available (Ebert et al., 1984), we digested total genomic DNA from *Hb.* sp. GRB and hybridized it with cosmids representing pGRB37 and pGRB90 to determine correspondence (Fig. 3.2). Fortunately, we could control for the hybridization signals caused by pGRB90's repeated sequences (see below) by comparing the *Hb.* sp. GRB genomic hybridization pattern with that from the halocin non-producing mutant of *Hb.* sp. GRB, which lacks pGRB90. (Direct *Pst*I digestion of the cosmid DNAs would not have given us a proper list of *Pst*I fragments because the inserts do not have *Pst*I ends.)

It is clear that pGRB37 and the "35 kbp" plasmid (Ebert et al., 1984) are the same (Fig. 3.2). The two differences between pGRB37's *Pst*I pattern and the *Pst*I bands assigned to the 35 kbp plasmid in Ebert et al. (1984) are easily explained. The 1.1-kbp *Pst*I doublet arising from pGRB37 comigrates with the intense band corresponding to uncut supercoiled pGRB1. The plasmid pGRB1 is also responsible for the second difference, the observed absence of a 1.8-kbp *Pst*I fragment in pGRB37. If one runs uncut *Hb.* sp. GRB genomic DNA on a 1% agarose gel, pGRB1 is seen as a major band at about position 1.1 kbp, and a minor but nevertheless bright band at about position 1.8 kbp (data not shown).

Fig. 3.2. Identification of pGRB37 and pGRB90 as being equivalent to the previously described '35 kbp' and '65 kbp' plasmids (Ebert et al., 1984), respectively. Southern blots of *Pst*I-cut genomic DNA from *Hb.* sp. GRB were hybridized with the pGRB37 cosmid G8A11, and with the three pGRB90 cosmids, G15E2, G5C10 and G6A4. Bands present, at the same intensity, in both the 'GRB w.t.' and 'GRB hal$^-$' lanes belong to pGRB305 and result from repeated sequences. We digitized the fragment patterns, and after adjusting their scales (the G8A11 and pGRB90 gels did not run the same distance), we created the pGRB90+pGRB37 composite. Black bands represent pGRB37 fragments, and thinner gray bands represent pGRB90 fragments. Some pGRB90 fragments are obscured by comigrating bands derived from the higher-copy number pGRB37. Next, we digitized the plasmid bands from the *Hb.* sp. GRB lane from Fig. 2 of Ebert et al. (1984), and scaled it to match our composite. See text for interpretation.


G8A11
G15E2
G5C10
G6A4
GRB w.t.
GRB w.t.
GRB hal⁻
GRB w.t.
GRB hal⁻
GRB w.t.
GRB hal⁻
pGRB90+pGRB37
composite
65+35+1.6 kbp
(from ref.26)

The 1.8 kbp band thus does not belong to the 35 kbp plasmid. Not surprisingly, then, pGRB37 is the 35 kbp plasmid from Ebert et al. (1984).

We are also convinced that no differences exist between *Hb.* sp. GRB's "65 kbp" plasmid (Ebert et al., 1984) and pGRB90. The faint *Pst*I fragments conservatively assigned to the 65 kbp plasmid in Fig. 2 of Ebert et al. (1984) are a subset of our tabulation of *Pst*I fragments from pGRB90. If one examines Ebert et al.'s data carefully, it is possible, in retrospect, to find each of the remaining bands above 1.5 kbp—without any others not present in pGRB90. Thus the evidence is fully in favour of the equivalence of pGRB90 and the 65 kbp plasmid. This point is of concern because halobacterial plasmids are prone to rapid change, yet we hope that *Hb.* sp. GRB can become the paradigm for genetic stability in *Halobacterium*.

The genome is a compositional mosaic

The physical map of the *Halobacterium* sp. GRB genome reveals regions of different restriction site density (Fig. 3.1 and Table 3.1). Although small site clusters of unknown significance are scattered around the chromosome, two large site clusters take the form of pGRB90 and (most of) pGRB305. We hypothesize that the relatively site-rich DNA represents FII DNA (not tested directly here), an inference based on three lines of evidence. First, there is a correspondence between FII and site-rich DNA in *Hf. volcanii* (Schalkwyk et al., 1993). Secondly, *Hb. salinarium* FII DNA is relatively richer in *Eco*RI and *Hin*dIII sites than is FI DNA (Pfeifer et al., 1983; Pfeifer and Betlach, 1985); we find that pGRB305 and pGRB90 have 3-4 times (respectively) the density of *Hin*dIII sites and 4-6 times (respectively) the density of *Eco*RI sites (Table 3.1, and data not shown) relative to the chromosome. Finally, Ebert et al. (1984) determined that the 65 kbp plasmid and

the minor-cccDNA were FII. We have shown that the "65-kbp" plasmid is indeed pGRB90 (see above), and we assume that the minor-cccDNA is pGRB305. Minor cccDNA is a very low copy number species present in halobacterial cccDNA preparations, and arises either from rearrangement of a major plasmid or of the chromosome, or from the inability to obtain a good yield of plasmid DNA which is very large (Pfeifer et al., 1983; Ebert et al., 1984). It was believed that *Hb.* sp. GRB's minor-cccDNA was chromosomally derived, since it was present in both the chromosomal DNA as well as in the cccDNA fractions (Ebert et al., 1984). We believe it more likely that pGRB305 is simply difficult to obtain as a cccDNA because of its size.

Repeated sequences in the Halobacterium sp. GRB genome

Halobacterium sp. GRB does not contain any of the characterized halobacterial insertion sequences (Ebert et al., 1986), the likely explanation for the strain's relative genetic stability. Our map of the *Hb.* sp. GRB genome, however, shows a considerable amount of FII DNA, which in other halobacteria seems to have the principal function of housing insertion elements. Although rare, we have observed the occasional sectored colony of *Hb.* sp. GRB (unpublished), a characteristic feature of transpositional bursts. We were therefore interested in cataloguing the repeated sequences which might be present within the *Hb.* sp. GRB genome, and our set of overlapping cosmid clones allowed us to find them easily.

Since insertion elements in halobacteria are known to be common in FII and in plasmids (Pfeifer et al., 1982; Pfeifer and Betlach, 1985; Cohen et al., 1992; Schalkwyk et al., 1993), we used every cosmid representing *Hb.* sp. GRB plasmid DNA as a hybridization probe against the entire minimal set of cosmids. From all of the 432 kbp of

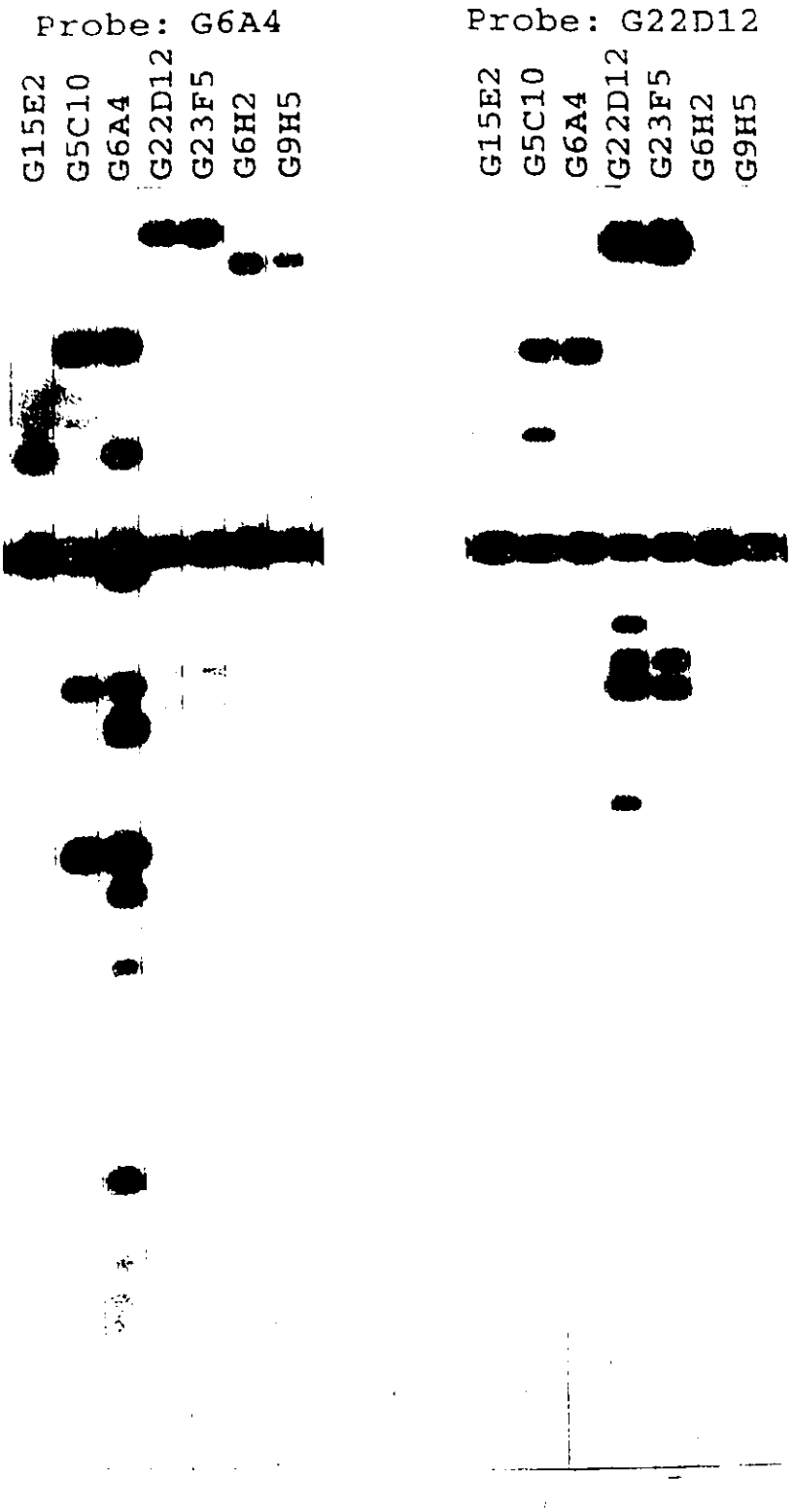
plasmid DNA, we found only five pairs of cross-hybridizing sequences (Figs. 3.1 & 3.3), which may or may not include insertion elements. Hybridization of 22 non-redundant cosmids representing a total of over 40% of the chromosomal DNA revealed no repeated sequences other than the one matching a copy on pGRB305, although in our gene-mapping effort, we later found the two known gene duplications, involving *flgA*/*flgB* (Gerl and Sumper, 1988) and *sodI*/*sIg* (Joshi and Dennis, 1993). Our hybridization stringency was high enough to avoid the problems of background hybridization common in this 65-70 mol% G+C DNA, but as a control, we hybridized the *Hf. volcanii* repeat-containing cosmid clone E11 (Schalkwyk et al., 1993) to blots of the *Hf. volcanii* cosmid clone collection and found numerous hybridization signals (results not shown).

Extensive homology is shared between regions of pGRB305 and pGRB90 (Fig. 3.3), although the restriction maps are not recognizably similar. This suggests that the homology may be broken up into smaller units, perhaps in the form of insertion elements. Even so, our systematic cross hybridization experiments have shown that any insertion element family present in *Hb.* sp. GRB must be of very low copy number. This is in stark contrast to the "typical" halobacterial situation, where various clans of insertion elements abound.

Genetic map

Although we have not yet mapped many genetic loci, all twenty map to site-poor DNA, and all but one (the genes for gas vesicle production) map to the chromosome (Fig. 3.1). It is interesting that the gas vesicle locus (*vac*), known to be within the FI fraction of *Hb.* sp. GRB's DNA (Horne et al., 1988), lies within a site-poor pocket of pGRB305.

Fig. 3.3. Cross hybridization between G6A4 of pGRB90 and G22D12 of pGRB305, involving several restriction fragments. Southern blots of *Bam*HI cosmid clone digests (a subset of which are shown in this figure; see Fig. 3.1 for map positions) were hybridized with a ^{32}P-labeled cosmid clone. This procedure identifies homologies between cosmids, due either to overlap or to the presence of repeated sequences. (The 5.4-kbp hybridization signal present in each lane marks the Lorist M vector fragment.) All five *Bam*HI fragments from G22D12 hybridize with the G6A4 probe, and conversely, four of G6A4's twelve *Bam*HI fragments, as well as an additional fragment from the adjoining pGRB90 cosmid G5C10, hybridize with the G22D12 probe. The four other cross-hybridizing regions within the *Hb.* sp. GRB genome (Fig. 3.1) involve single restriction fragments; an example can be seen in the G6A4 panel, showing hybridization to the 17.5 kbp *Bam*HI fragment from G6H2/G9H5.


Probe: G6A4
G15E2
G5C10
G6A4
G22D12
G23F5
G6H2
G9H5
Probe: G22D12
G15E2
G5C10
G6A4
G22D12
G23F5
G6H2
G9H5

This *vac* operon is homologous to the c-*vac* of *Hb. salinarium*, believed to map to the chromosome (Horne et al., 1988) (*Hb. salinarium* does, however, have megaplasmids [Ng and DasSarma, 1993] and it will be interesting to see if c-*vac* actually turns out to map to one of them rather than to the chromosome.)

The gross partititioning of genes in *Hb.* sp. GRB is similar to that observed for *Hf. volcanii* (Charlebois et al., 1991; Cohen et al., 1992), which is to say that genes are scattered around the chromosome. Only eight of the mapped genetic loci correspond to loci on the current map of *Hf. volcanii*, however, making a detailed comparison of gene order difficult at this time. We are presently working on increasing the resolution of this multipoint comparison. Early results suggest extensive rearrangement of the chromosomal structure.

Discussion

The landmark strategy is an efficient method for cloning and mapping small genomes. Refinements in our implementation of this approach, presented here, made the mapping of *Halobacterium* sp. GRB's genome easier and more efficient than it had been for *Haloferax volcanii*. Beginning with forty genome equivalents of cosmid clones instead of three (Charlebois et al., 1989) reduced the time spent in later chromosome walking and it facilitated verification of the contigs by providing ample levels of redundancy. Chromosome walking into a cosmid clone library constructed with a different restriction enzyme avoided the representational bias of the original library, and allowed us to clone the rest of the genome, apart from a few small regions which are probably unclonable as cosmids.

With mapping methodology simplified, one can better focus on the reason for mapping: to gain information on structure, function and evolution. The structure of the genome of *Hb.* sp. GRB is not much different in style than are the genomes of other halobacteria. Indeed, it is now possible to make some general but firm conclusions about their construction. A single 2-3 Mbp circular chromosome of FI DNA contains the vast majority of important genes, and may be interrupted occasionally by islands of FII. In addition, the genome contains plasmids of various size and of various composition—some mosaic—containing few known genes. FII DNA is rich in repetitive DNA, much of which can be identified as insertion sequences. Finally, genetic stability—at least as measured through mutation of pigment genes—seems correlated with the load of insertion elements borne by the genome.

The true value of the *Halobacterium* sp. GRB genome mapping project lies in the collection of resources generated, which can be used to test hypotheses about the origins and evolution of the halobacterial genome. *Hb. salinarium* is rich in "selfish DNA", consisting of numerous, perhaps competing, families of insertion elements parasitizing the genome. Its close relative, *Hb.* sp. GRB, has a genome which either has escaped invasion by the more aggressive insertion elements or which has been cured of them. The *Hb.* sp. GRB genome may thus be considered primal, or at least healthy. Besides possibly serving as a control strain in such studies as testing the invasiveness of halobacterial insertion elements, or for characterizing relatively undisturbed FII DNA, it will be most valuable in comparative genomics. Conservation of gene order between *Hb. salinarium* and *Hb.* sp. GRB would provide a strong indication of the existence of forces which may be preserving genomic structure; lack of conservation might testify to the contrary. Few such stabilizing forces are currently known, and it is unclear how effective these may be in counteracting

the many well-known destabilizing forces. Our map, and our ordered clone collection, make such studies feasible.

Acknowledgements

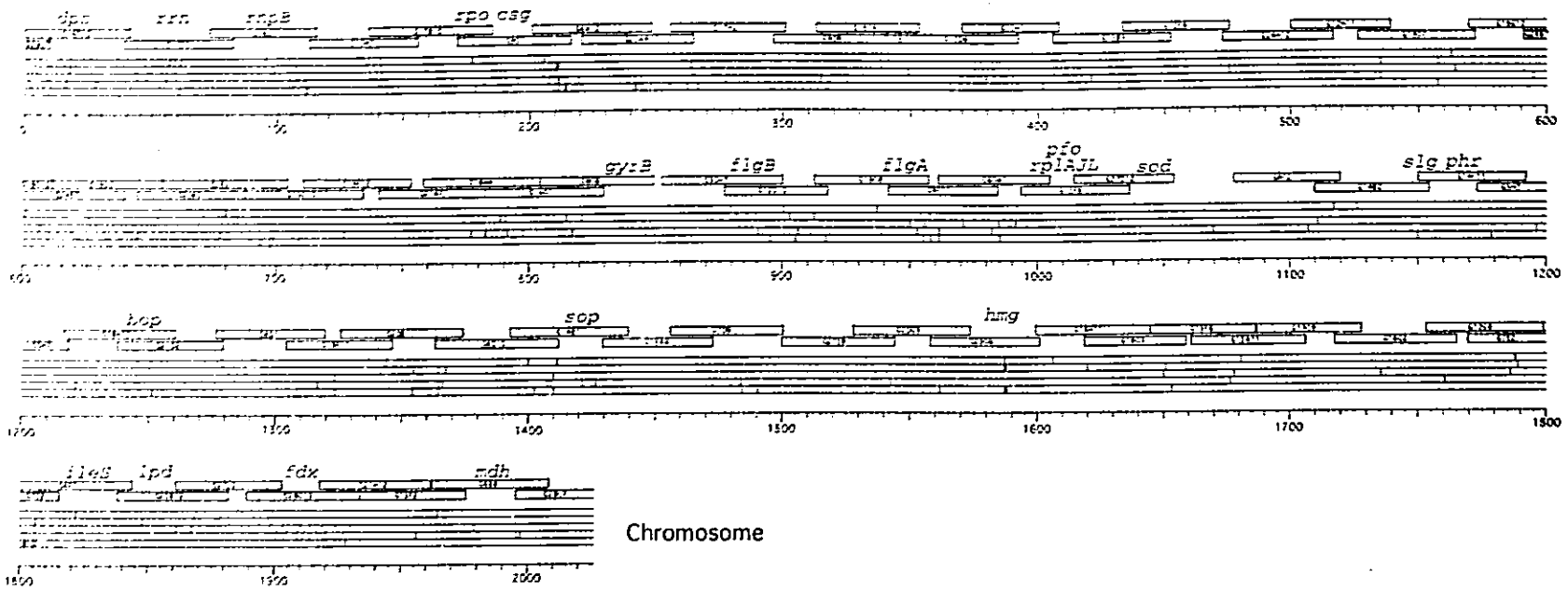
This work was funded by NSERC. We thank L.C. Schalkwyk for suggestions, and we gratefully acknowledge all of the laboratories who sent us gene probes, including those whose probes did not hybridize with GRB DNA and are thus not cited in Table 3.2. RLC is an Associate of the Canadian Institute for Advanced Research.

Update to Chapter 3

When this chapter was written, the relationship between strain GRB and other members of the genus *Halobacterium* had not been adequately worked out to assign it a proper specific epithet. Since that time, the detailed mapping efforts of N. Hackett and colleagues (Hackett et al., 1994) have shown that GRB is indeed a member of the species *salinarum*.

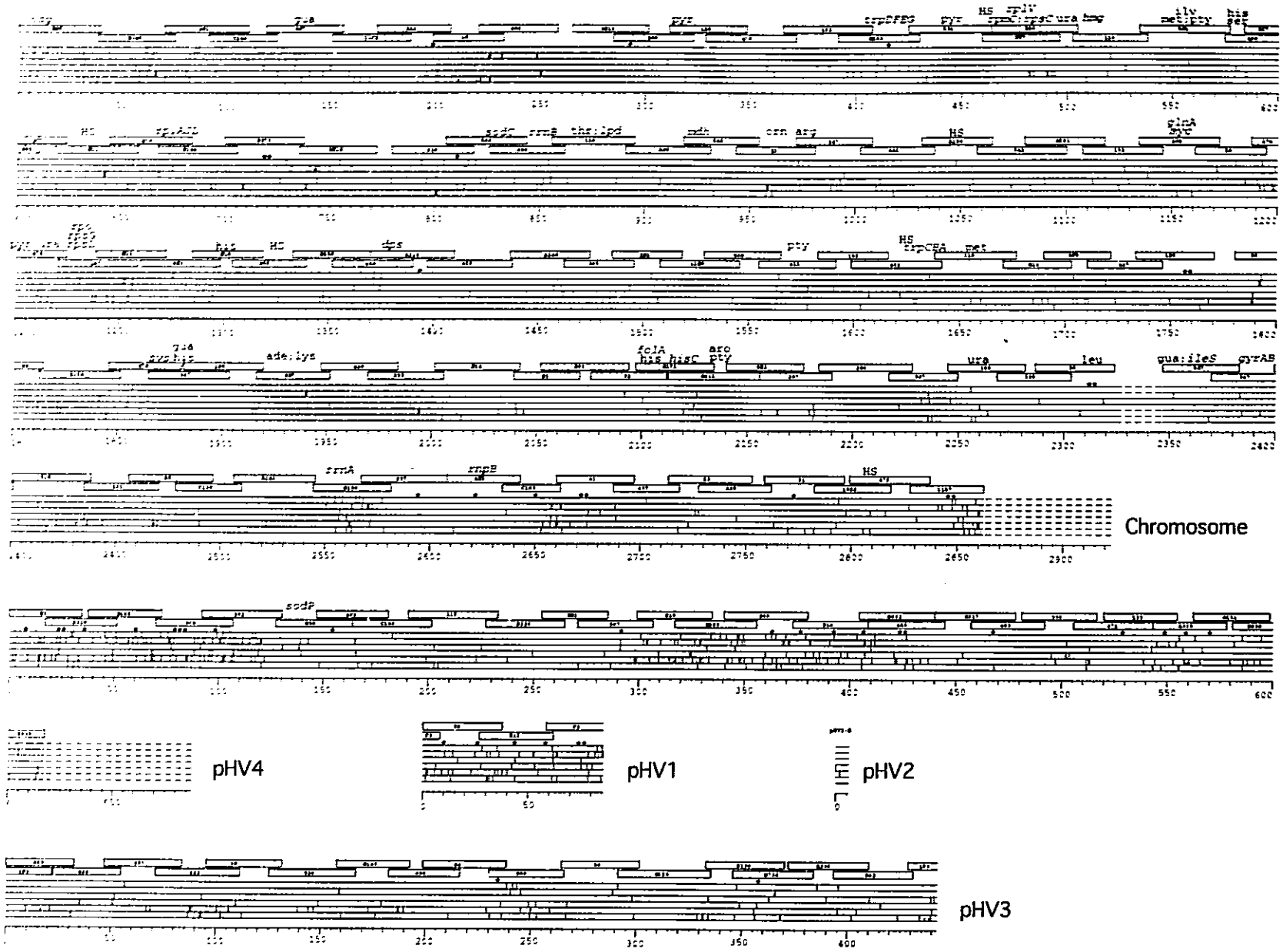
Continuing work on the GRB genomic map has placed an additional gene (*dps* cloned from *Hf. volcanii* [A. St. Jean, unpublished]) and closed one of the supposed gaps on the chromosome. The cosmids G16D10 and G29L7, which are the first and last chromosomal cosmids in both Fig. 3.1 and Fig. 3.4, were found to overlap, closing this gap. This changes the estimated size of the chromosome from 2038 kbp to 2026 kbp. Fig. 3.4 illustrates the *Halobacterium salinarum* GRB contig map reflecting this change and the additional gene. The contig map of the *Haloferax volcanii* DS2 genome is shown in Fig. 3.5. This map is included here for easy reference even though *Hf. volcanii* is not part of this chapter. For a comparison of these two genomes, see chapter 6.

Fig. 3.4. Updated physical and genetic maps of the genome of *Hb. salinarum* GRB. Cosmid clones representing 99% of the genome are shown as boxes above the restriction map. Enzymes used in the map are, from top to bottom; *Afl*II, *Ase*I, *Dra*I, *Hin*dIII and *Ssp*I. Scale bar is in kbp. Mapped genes are italicized above the cosmid clones. From Charlebois, in press with permission of the author.


rrn
rrpB
rpo csg
gyrB
flgB
flgA
pfo
rplAJL
sod
slg phr
hop
sop
hmg
fleQ
ipd
fdx
mdh
Chromosome
gvp
pGRB305
pGRB90
pGRB37
pGRB1

88-A

Fig. 3.5. Updated physical and genetic maps of the genome of *Hf. volcanii* DS2. Cosmid clones representing 96% of the genome are shown as boxes above the restriction map. Enzymes used in the map are, from top to bottom; *Bam*HI, *Bgl*II, *Dra*I, *Hin*dIII, *Pst*I, and *Ssp*I. Scale bar is in kbp. Genes mapped by hybridization are in italics. Genes mapped by complementation of auxotrophs are in plain text. Putative heat shock loci are indicated by 'HS' while members of the insertion sequence family ISH51 are shown by a star. From Charlebois, in press with permission of the author.


Chromosome
pHV4
pHV1
pHV2
pHV3

CHAPTER 4

Supercoiling and map stability in the bacterial chromosome

This chapter is published as:

Charlebois, R.L., and A. St. Jean. 1995. Supercoiling and map stability in the bacterial chromosome. J. Mol. Evol. **41**:15-23.

Abstract

A major goal of comparative genomics is an understanding of the forces which control gene order. This assumes that gene order is important, a supposition backed by the existence of genomic colinearity between many related species. In the bacterial chromosome, a polarity in the order of genes has been suggested, influenced by distance and orientation relative to the origin of DNA replication. We propose a model of the bacterial chromosome in which gene order is maintained by the adaptation of gene expression to local superhelical context. This force acts not directly at the genomic level but rather at the local gene level. A full understanding of gene-order conservation must therefore come from the bottom up.

My Contribution

The local context model started with a question: why do some prokaryotic genome comparisons discover significant conservation in the order of homologous loci while others find none? My contribution was to provide an answer to this question. If two genomes show conservation, it is because their particular arrangement of genes is adaptive

and changing this arrangement would be deleterious to the organisms. If two genomes lack conservation, then this indicates that at some point since the two lineages separated the adaptive force maintaining gene order was relaxed, allowing the genomes to diverge. R.L. Charlebois then proposed a mechanism—the interactions between DNA supercoiling and gene expression—through which gene order could be maintained. I also participated in the background research which led to this paper and helped write the manuscript.

Introduction

Comparative Genomics

Thanks to technological innovations within the past decade—in particular pulsed-field gel electrophoresis and strategies for sorting clone libraries—a collection of prokaryotic genome maps is building. Once limited to the more genetically amenable species, comparative genomics can now be applied to a broad assortment of organisms.

Comparative map and macrorestriction fragment analyses have begun to see use in strain identification and phylogeny and in the assessment of genomic variability within and between species. Genomic comparisons between strains or related species have revealed variable degrees of conservation in gene organization (Holloway et al., 1990; Pyle et al., 1990; Riley and Sanderson, 1990; Carlson et al., 1992; Ladefoged and Christiansen, 1992; Le Bourgeois et al., 1992; Taylor et al., 1992; Zuerner et al., 1993; Bukanov and Berg, 1994; Liu et al., 1994), whereas restriction maps are much more mutable (Ladefoged and Christiansen, 1992; Le Bourgeois et al., 1992; Liu et al., 1994). The range of map similarities observed from these comparisons is not surprising, given that the definition of a bacterial "species" is rather loose (Carlson et al., 1992).

Phylogenetically distant comparisons have generally revealed scrambled gene orders (Riley and Anilionis, 1978), although approaches for extracting meaningful similarities are being developed (Sankoff et al., 1990; 1992). Success in such deep comparisons depends on higher map resolution, which is currently affordable for only a few model organisms. Ultimately, technology will permit many more genomes to be intimately compared. Such alignment of maps will thoroughly reveal the extent of genomic rearrangement, but it may still do little to identify the forces which control gene order in the bacterial chromosome. What does it mean if genetic maps are similar, or dissimilar? Do these maps simply drift, or are they subject to selection?

The E. coli–S. typhimurium *Genome: A Paradigm of Stability*

Despite an estimated 120-160 million years of separate evolution (Ochman and Wilson, 1987) and an average sequence divergence of 15.6% (Sharp, 1991), the genetic maps of *Escherichia coli* and *Salmonella typhimurium* have remained strikingly similar (Riley and Sanderson, 1990). Though there is a large inversion and a number of deletions or accretions in one map relative to the other, they are otherwise congruent. Forces do exist to promote rearrangement in these genomes (Riley and Anilionis, 1978; Riley and Krawiec, 1987; Birkenbihl and Vielmetter, 1989; Krawiec and Riley, 1990; Naas et al., 1994)—plasmid-mediated events, transposition, and recombination between repeated sequences—yet these disruptive forces have somehow been countered by stronger forces which have maintained gene order in evolutionary time.

In an effort to understand functional and mechanistic barriers to rearrangement, several groups undertook to measure the inversion frequencies of various segments of the chromosome of *S. typhimurium* (Segall et al., 1988; Segall and Roth, 1989; Mahan et al.,

1990; Krug et al., 1994) and of *E. coli* (Rebollo et al., 1988; François et al., 1990). These analyses identified permissive and nonpermissive intervals for inversion of DNA. It is believed that the endpoints of nonpermissive intervals are unable to interact with one another, and thus that there are sometimes mechanical constraints to intrachromosomal recombination, due to nucleoid structure (Segall et al., 1988; Segall and Roth, 1989; Mahan et al., 1990; Krug et al., 1994), or perhaps to the bias in the orientation of chromosomal "organizing sequences" in the terminus region (Rebollo et al., 1988; François et al., 1990) or of the recombinationally important chi sites (Burland et al., 1993). Although some inversions occur at greatly reduced frequency (10^{-8}) relative to others, dependent on difficult exchanges between sister chromosomes (Segall et al., 1988; Segall and Roth, 1989), in theory these inversions should occur given the enormity of bacterial populations.

Do rearrangements affect fitness? Although the effects of inversion are not always obvious (Segall et al., 1988; Mahan et al., 1990), experiments have shown that inversion mutants are less competitive than the wild-type (Hill and Harnish, 1981), and that reinversion better restores fitness than does point mutation (Hill and Gray, 1988). Similarly, large transpositions mediated by recombination between *rrn* loci also have slight adverse effects on growth (Hill and Harnish, 1982).

Drift from one optimal map to another appears to be hampered by the prevalence of suboptimal maps in the intervening fitness landscape. The local optimal map thus appears to be maintained by one or more kinds of selective force. Few such forces are currently understood: they include the position and orientation of genes relative to *oriC* (discussed below), a desirability for the terminus of replication (*terC*) to be 180° from *oriC* in order to minimize the time required to replicate the chromosome bidirectionally, and the need

for a certain polarity of sequences near *terC* (Riley and Sanderson, 1990; Krawiec and Riley, 1990).

We propose another genome-stabilizing force, one which may be quite powerful, that being the effects of local superhelical tension on gene expression. Unlike the stabilizing forces described previously which act directly on genomic features, supercoiling opposes rearrangement from the bottom up, by imposing a contextual sensitivity to gene expression.

Gene Expression as a Function of Map Position

Position and Orientation Relative to oriC.

The chromosome of *E. coli* replicates bidirectionally from a point of initiation called *oriC*. Genes located nearer *oriC* will have a higher relative copy number than genes located more distally. Consequently, some genes may be expected to display a level of expression which reflects this gradient in copy number. This idea was substantiated in an experiment which moved a constitutively transcribed *hisD* gene around the chromosome (Schmid and Roth, 1987). Earlier, similar experiments involving integration of F' *lac* (Beckwith et al., 1966; Masters et al., 1985) also confirmed that chromosomal position influenced expression, though the results did not adhere as tightly to the expectations of the *oriC* gene-dosage model (Schmid and Roth, 1987). In this example, there was variation superimposed on the positional gradient, which has been tentatively attributed to the supercoiling or transcriptional context at the F' *lac* integration sites (Schmid and Roth, 1987).

In *E. coli*, genes are preferentially oriented to be transcribed in the same direction as DNA replication (Brewer, 1988; 1990; Burland et al., 1993). Inversions which flip large

blocks of genes such that their general orientation opposes DNA replication are not often seen or are detrimental to growth (Schmid and Roth, 1983; Louarn et al., 1985). This bias in strand usage is expected to reduce the number of head-on collisions between the DNA replication machinery and RNA polymerase. DNA polymerase can pass a transcriptional ensemble without disturbing transcription, at least in the same sense situation (Liu et al., 1993), but the wave of overwound DNA ahead of a transcriptional event (Liu and Wang, 1987; Wu et al., 1988; Figueroa and Bossi, 1988; Tsao et al., 1989; Rahmouni and Wells, 1992) may delay replication progress in the convergent situation (Brewer, 1988). A codirectional bias in the orientation of genes is even more pronounced in *Bacillus subtilis* (Zeigler and Dean, 1990; Sorokin et al., 1993).

Though these genomic-level forces are believed to promote stability in the bacterial chromosome, they seem incapable of opposing inversions symmetric about the origin or of restricting the movement of genes whose DNA copy number is not much of a factor in their expression. The limitations imposed by these forces undoubtedly contribute to genome structure, but they cannot fully account for the observed evolutionary stability.

Local Superhelical Context

Global superhelical tension in the nucleoid is controlled by a balance of topoisomerase activities: DNA gyrase introduces negative supercoils, and topoisomerase I relaxes them (Pruss et al., 1982; DiNardo et al., 1982; Richardson et al., 1984). The level of supercoil tension appears to be important, since it is regulated (Menzel and Gellert, 1983; Goldstein and Drlica, 1984; Tse-Dinh, 1985; Menzel and Gellert, 1987).

Several environmental factors can alter the regulated level of global supercoiling, and these reflect the need for different sets of genes to be expressed (Ni Bhriain et al.,

1989; Higgins et al., 1990). For example, anaerobic growth induces an elevation in global supercoiling in *E. coli* (Yamamoto and Droffner, 1985; Hsieh et al., 1991), as does an increase in osmolarity (Higgins et al., 1988). In contrast, starvation results in a relaxation of global supercoiling (Balke and Gralla, 1987; Jaworski et al., 1991). A study of the *E. coli* genome's transcriptional activity under different physiological conditions confirmed that osmotic shock, anaerobiosis and starvation all lead to numerous alterations in gene expression patterns, though unexpectedly the *topA10* nonsense mutation of topoisomerase I showed few differential effects (Chuang et al., 1993). Direct measurement of protein abundances from cells with altered superhelical tension has shown that for many genes, expression is remarkably sensitive to small global differences in titratable supercoiling (Drlica, 1987). Most promoters respond to the level of supercoiling, either increasing or decreasing expression (Sanzey, 1979; Jovanovich and Lebowitz, 1987; Menzel and Gellert, 1987).

Global supercoiling is regulated at the level of the domain, of which there are around 50 in the *E. coli* chromosome (Worcel and Burgi, 1972; Sinden and Pettijohn, 1981). It appears that DNA gyrase defines the boundaries of these domains by interacting with DNA at REP elements (Yang and Ames, 1988; 1990) or toposites (Condemine and Smith, 1990). The ability of domains to be supercoiled independently of one another is believed to protect the bulk of the nucleoid from becoming relaxed due to the nicks and gaps produced by DNA repair and replication (Drlica, 1987). It is not known if individual domains can be regulated to different levels of supercoiling (Drlica, 1984; 1987; Condemine and Smith, 1990), though it is possible that gyrase's differential affinity for individual REP elements contributes to some variability (Yang and Ames, 1990).

Local levels of supercoiling within the domain appear to be influenced to a large degree by local gene expression activity (Pruss and Drlica, 1989, Drlica et al., 1990). DNA is overwound ahead of a transcriptional bubble, and underwound behind it (Liu and Wang, 1987; Wu et al., 1988; Figueroa and Bossi, 1988; Tsao et al., 1989; Rahmouni and Wells, 1992), at levels sufficiently high to influence DNA conformation (Rahmouni and Wells, 1989; 1992). The effect is especially pronounced when the transcription-translation complex is anchored to the membrane, restricting free rotation of the macromolecular assemblies to a greater degree (Lodge et al., 1989; Lynch and Wang, 1993). Rho, a transcription terminator, also plays a role (Arnold and Tessman, 1988). Return of supercoiling to normal levels is not likely to occur immediately: adjustment of perturbed supercoil levels in the chromosome is known to be slow (Goldstein and Drlica,1984; Rahmouni and Wells, 1992). Topoisomerases are competent to eventually correct the imbalances caused by transcription (Figueroa and Bossi, 1988; Koo et al., 1990; Spirito et al., 1994), but in their limited numbers (Liu and Wang, 1987) they may be ill-suited to police rigorously any but the worst problems, where transcription is convergent, jamming progress, or where transcription is divergent, denaturing the intergenic DNA. (We suppose that most genes may be oriented in the same direction to prevent such embarrassments. Following such an alignment, blocks of genes are biased wholesale for one strand or the other. Whether the general orientation is with or against the direction of DNA replication then becomes important. This idea recognizes the problem of potential collisions between polymerases without requiring that more highly expressed genes be preferentially oriented, a correlation [Brewer, 1988; 1990] which has been challenged [Burland et al., 1993].)

Nucleoid DNA-binding proteins are known to affect superhelical tension as well (Rouvière-Yaniv et al., 1979; Broyles and Pettijohn, 1986; Hulton et al., 1990; Tupper et al., 1994). Indeed, half of the supercoils in the *E. coli* chromosome are constrained by proteins (Pettijohn and Pfenninger, 1980; Bliska and Cozzarelli, 1987). HU, IHF, and H-NS have been particularly well studied, and are known to participate in a number of pleiotropic interactions (Pettijohn, 1988; Hulton et al., 1990; Rouvière-Yaniv et al., 1990; Oppenheim et al., 1993; Painbéni et al., 1993). Their principal roles, of relevance here, may be to condition the superhelical context of sensitive genes (Oppenheim et al., 1993; Higgins et al., 1989; Owen-Hughes et al., 1992), to moderate the effects of extremes in supercoil tension (Pontiggia et al., 1993), and to interact with regulatory DNA-binding proteins to control that regulation. (For reviews, see Pettijohn, 1988; Pettijohn and Hodges-Garcia, 1990.) A pertinent example of such an interaction involves the management of a domain's supercoil level by HU's influence on the activity of DNA gyrase at REP elements (Yang and Ames, 1990).

There are numerous examples of the influence of superhelical tension on the expression of individual genes (Drlica, 1984; Jovanovich and Lebowitz, 1987; Pruss and Drlica, 1989; Drlica et al., 1990; Thompson et al., 1990; Figueroa et al., 1991; Chen et al., 1992; O'Byrne et al., 1992). The enhancement of promoter activity by negative supercoil tension is easy to envision, since the initiation of transcription includes a local denaturation at the promoter to initiate a viable transcriptional complex (Mishra and Chatterji, 1993). Too much supercoiling, however, can lead to unusual structures such as cruciforms which may block transcription directly (Bagga et al., 1990) or by sequestering supercoils away from the promoter (Horwitz, 1989). Each promoter may indeed have its own optimum level of twist, dependent on the spacing between promoter elements (Wang and Syvanen,

1992), and also on the promoter's sequence (Borowiec and Gralla, 1987) and G + C content (Figueroa et al., 1991; Chen et al., 1992). Topological coupling, mediated by transcriptional activity, further influences the expression of neighboring genes appreciably (Richardson et al., 1988; Lilley and Higgins, 1991; Chen et al., 1992; Tan et al., 1994).

Gene expression, however, is more complex than just an act of transcription by RNA polymerase. Often, proteins are interacting with the DNA sequences upstream of the cistron to regulate its expression. It is reasonable to suppose that many regulatory proteins will bind to their target sequences with different affinities depending on the degree of supercoiling twist (Wang and Syvanen, 1992) or writhe (e.g., Borowiec et al., 1987). For instance, catabolite repression has long been known to be sensitive to supercoil levels (Sanzey, 1979). There are several specific examples of the modulation of repressor binding by supercoil level, including the lactose repressor in *E. coli* (Whitson et al., 1987), and EarA, the repressor of *aniG* in *S. typhimurium* (Karem and Foster, 1993). Another notable example is the preferential binding of HU to the structures induced by DNA supercoiling (Pontiggia et al., 1993).

There are different kinds of regulatory DNA-binding proteins, and the consequence of changing the binding constant in different superhelical contexts is dependent on the type of regulation. Proteins which bind to their own gene's upstream sequence to effect a feedback inhibition would see their concentration change with a change in binding affinity. Proteins such as inducers and repressors, which alter conformation due to being modified or due to the binding of some ligand, may more-or-less easily induce or repress the cistrons under their control. Base levels of expression in the "off" state would vary with altered supercoiling, as would the derepressed or induced "on" levels. The influence of

context on altering gene expression may not be important for all genes, but the cascades which the sensitive expression may engender should be quite significant and pervasive.

Transcriptional Trespassing

Improper transcriptional termination of an upstream cistron could also affect a gene's expression. Failure to terminate may result in the transcription of a downstream gene, producing either an mRNA, or an antisense mRNA if the gene is oppositely oriented. An interesting case arises from the mechanism for regulating some phase-variable genes by promoter inversion or by cassette switching (Krawiec and Riley, 1990; Dybvig, 1993). Though this aspect of transcription is not directly related to supercoiling, it also influences the quality of a gene's positional context.

Stability of a Gene's Superhelical Context

Specific features inherent within the local sequence context—cruciform- or Z DNA-forming regions, bends, G + C content, recognition sites for DNA-binding proteins, and the activity and orientation of nearby genes—are what determine the local superhelical context. The local context, in turn, is coarsely influenced by the domain's regulated level of supercoiling, through the interaction of topoisomerases with specific sequences. Sequence features are relatively static, and hence the context experienced by each transcriptional unit is stable as long as gene expression does not have to deal with environmental parameters outside of a predescribed range. It is of adaptive value to the cell to be able to predict its responses to the environment—to know its own life history—and it is crucial that gene expression be regulated and controlled to work within that prediction. We expect that a given locus on the chromosome experiences a range of

superhelical tensions, some loci suffering larger fluctuations than others. Genes will be found in contexts which are conducive to their function (Pruss and Drlica, 1989; Drlica et al., 1990). In organisms like *E. coli* and *S. typhimurium*, which change their gene expression patterns while in the gut partly by raising global supercoil levels, the risk of moving genes around is doubly great since the genes must be adapted to (at least) two sets of contexts.

Implications

Contextual Adaptation of Gene Expression

Since the regulation of gene expression has adaptive value, and since the level of regulation may well differ under different local superhelical contexts, it follows that selective pressures exist to keep the gene in the context that works best. We discussed how this force applies to the conservation of gene order between *E. coli* and *S. typhimurium*, but the model should apply equally well to other prokaryotes, since supercoiling is ubiquitous (e.g., see Yang and DasSarma, 1990; Grau et al., 1994).

Our model predicts that genes are placed in positions which maximize their adaptive value throughout the normal range of environments encountered by the cell. If the life history of the species is invariant, each gene will further adapt by mutation to improve its performance within its own positional context. In a predictable environment, it becomes increasingly counteradaptive to move genes around, because the superhelical context of these genes would likely be altered. Thus, if gene expression and its regulation can anticipate each of a life history's environmental challenges, there is genomic stability. This is true even in a complex environment such as faced by *E. coli* (in and out of the animal

gut), since *E. coli*'s elaborate regulatory network makes its perceived environment stable and predictable.

If a lineage were to exploit a new niche, to change the characteristics of its life history, one possibility is that a new stable map might emerge. Gene expression patterns would fall outside the range predicted from the abandoned life history such that local gene contexts would become maladapted. The force to maintain gene order would then relax. Mutations and haphazard rearrangements would both occur to adjust individual gene expression levels, satisfying the cell's new needs. After this initial relaxation of constraint on rearrangement, leading to a multitude of possible maps, one or more of the best new gene arrangements would stabilize as the adaptation phase matured.

Alternatively, while adapting to a new life history, mutations might occur to increase the general adaptability of gene expression, minimizing the perceived effect of the environmental change by adjusting regulatory strategies. This kind of adaptation would result not in the replacement of one life history for another, but rather in their integration. This is the generalist's strategy. Selective pressures would attempt to mutate or to reposition genes such that expression levels became more favorable in both situations. Genes may tend to settle at loci shielded from transcriptional and superhelical inconsistencies.

One way of insulating genes from the activities of their neighbors would be to minimize the influence of supercoiling itself on gene expression. Supercoiling is of course not a dispensable characteristic, since it has a role in packaging the nucleoid. However, local variation in supercoil tension might be moderated considerably by buffering aberrant regions more effectively with topoisomerases or with DNA-binding proteins. A reduction

in the importance of the supercoil context would result in a continued relaxation of the constraints on genome rearrangement and an increased rate of genomic drift.

Elimination of the effect of supercoiling on gene expression is not an automatic consequence of adaptation to multiple environments, however. Cells like *E. coli*, which periodically inhabit very different habitats, have developed a means to regulate large sets of genes by adjusting global supercoil tension. This dependency on the sensitivity of gene expression to supercoiling precludes its loss.

What does it mean if genetic maps are similar, or dissimilar? If maps of bacterial chromosomes are similar, then the life history shared by each of the strains being compared has differed little from that of their ancestor. If the maps are not the same, then at some point in the history of the strains, environmental challenges have differed significantly. A currently unstable map order implies either that adaptive selection is in progress, or that the forces maintaining gene order are, in that lineage, weak. When comparing very divergent organisms, one must also take into account that, like DNA sequences for conserved genes, there will be drift in the maps. We predict that environmental challenges increase the rate of drift dramatically, by relaxing constraints on genome rearrangement. Under selection, a map optimized for the new environmental conditions might be found through genomic drift, and a succession of rearrangements to perfect the new map would ensue.

Most likely, several alternative maps could adapt the organism to its new niche. The population's drift towards various maps could be considered an act of speciation, since divergent map order effectively blocks recombinational exchange (Krawiec, 1985). Phylogenetically, an act of adaptation to a new niche would thus be seen as an event of evolutionary radiation.

Disruptive Forces

Though the chromosome of *E. coli* and *S. typhimurium* has resisted shuffling of its gene order, a number of small genomic alterations have been incorporated in addition to the single conspicuous inversion. Relative to *E. coli*, the *S. typhimurium* chromosome has 14 missing regions, and 15 additional loops (Riley and Krawiec, 1987; Riley and Sanderson, 1990), representing instances of transposition, accretion and deletion. We expect that these alterations would have had consequences on local superhelical tension, and thus that these events disturbed optimal settings. It is interesting that a number of minor and local changes in gene order have occurred immediately adjacent to the insertion/deletion loops (Riley and Anilionis, 1978; Riley and Krawiec, 1987), and we speculate that they represent the result of a minor cascade of rearrangements spurred by the initial perturbation. That these rearrangements are local (within about a minute) is supported by the observation that nucleoid structure most favors recombinational interaction in the 40–80-kbp proximal range (Krug et al., 1994).

Gene order seems generally static, but the *E. coli* and *S. typhimurium* genomes are by no means immutable. There are the deletions and insertions of DNA regions, discussed above, which contribute to the phenotypic differences observed (Riley and Anilionis, 1978; Riley and Krawiec, 1987; Riley and Sanderson, 1990; Groisman et al., 1992). There has also been the occasional transposition. In addition, the complement of insertion sequences varies in strains within and between the two species (Biserùić and Ochman, 1993a,b), reflecting a history of movement. Finally, and most importantly, there is strong evidence that the chromosome is a mosaic of sequences transmitted horizontally (Milkman and Stoltzfus, 1988; Milkman and Bridges, 1990). This allelic replacement phenomenon is perhaps one of the strongest forces in organismal adaptive evolution.

How does supercoiling oppose the various disruptive forces? Rearrangements can be stable or unstable, and they may or may not directly alter gene order. Deletions and insertions alter spacing, but not order. Non-duplicative transposition alters spacing and order, whereas duplicative transposition alters spacing and, in only a special sense, order as well. Each of these events is stable (difficult to revert), and can alter local superhelical context. As argued in this paper, altered context is predominantly disadvantageous and therefore counterselected. Addition of useful new genetic functions by insertion may overcome the detrimental effect of the interruption, though the choice of insertion site and the ease by which the context can be readjusted are likely selective factors as well. Deletions and transpositions, though, seem to do nothing positive: they may remove or disrupt genes and they alter contexts. We propose that cryptic genes may be kept as much for their effects on local context as for their rare coding value.

Another stable genomic modification is replacement, where homologous sequences are exchanged. The replacing DNA may be structurally different, with deletions or insertions (Stoltzfus et al., 1988), but the contextual details of the replacing DNA are already tried and trusted, having come from cells common enough to be spreading their genes around. Therefore, adaptive context replacement accompanies allelic exchange. This brand of genomic evolution, which is so common (Milkman and Stoltzfus, 1988; Milkman and Bridges, 1990), is nondisruptive.

There are two kinds of unstable rearrangement. Tandem duplications, unless the duplication is favored by selection, will easily revert to single copy. These duplications do little to alter gene order anyhow. Of the forces which can change gene order, inversion is potentially the most powerful. Nested and overlapping inversion events, if unrestricted, would soon lead to a shuffled gene order. Inversions mediated by recombination, however,

are reversible. Since the superhelical contexts of the endpoints of the inversion isomers will differ, the isomer historically adapted will prevail, and will become increasingly rooted.

Comparative Genomics

Comparison of evolutionarily distant genomes usually reveals scrambled gene orders. High-resolution comparisons may nevertheless reveal blocks of genes whose order is conserved. Pairwise similarities may be a consequence of chance, but when a sufficient number of genomes are examined to rule out chance, a block may be conserved because of some general adaptive value. The extreme case of gene order conservation is the bacterial operon, where genes for a biochemical pathway have found each other and through clipping and trimming of intergenic spaces have become cotranscriptionally regulated. We expect transcriptionally distinct cistrons to be clustered as well, occasionally, due to a mutually beneficial conditioning of each other's context. The best example of a well-conserved gene cluster is near the origin region (Fujita et al., 1989). Supercoiling, protein binding and transcriptional activation are known to be important for the function of *oriC* (Zyskind, 1990), and thus the particular gene arrangement around *oriC* may have been selected to maintain a particular context.

Genomes must be compared at the highest resolution, if any conclusion other than one about the stability of life histories is to be made. Gene order is very important within the genome, but between genomes, its overall conservation is a consequence of individual historical demands on gene expression. When, through multiple high-resolution comparisons, blocks of evolutionarily conserved gene order are discovered, it will be necessary to identify what the driving force is which maintained that block. For operons, it

is cotranscriptional regulation. For *oriC*, it may be the genetic context around *oriC* itself. Some genes may also remain linked due to an interdependency driven by slippage repair (Ornston et al., 1990), such that rearrangement becomes mutagenic. Other blocks of conserved gene order may well have their own functional or historical idiosyncrasies keeping them together, and we expect that in many, it will be a generally favorable superhelical context.

Acknowledgments

For support, we thank the Natural Sciences & Engineering Research Council and the Canadian Institute for Advanced Research.

Update to Chapter 4

The degree and pattern of DNA supercoiling relates to such widely studied phenomena as the regulation of gene expression and DNA packaging. As such, it is not surprising that the literature on the subject has increased since the time this chapter was submitted for publication. Table 4.1 lists further examples of promoters with activities affected by their level of supercoiling. These papers add support for one of the necessary conditions for our model to work, that gene expression be susceptible to changes in local context, in this case due to changes in the supercoiling of DNA.

This chapter only touches upon the global (genome or domain-wide) adjustments to DNA supercoiling seen in bacteria during changes in growth conditions, covered in the section of chapter 1 dealing with the nucleoid. This is because the local context model — as I will call the model described in this chapter — implies nothing regarding the global regulation of gene expression by DNA supercoiling and thus whether bacteria possess this ability or not does not affect the model's validity. We mention that global changes in DNA

Table 4.1. More genetic loci with promoters sensitive to changes in DNA supercoiling.

	Locus	Function	Organism	Reference
1	*spv*	plasmid-linked virulence	*Salmonella typhimurium*	O'Byrne and Dorman, 1994
2	*fimA*	Type 1 fimbrial subunit	*Escherichia coli*	Dove and Dorman, 1994
3	*tyrT*	$tRNA^{tyr}$	*Escherichia coli*	Free and Dorman, 1994
4	*ptsH*	phosphoenolpyruvate:glycose phosphotransferase system	*Escherichia coli*	Ryu and Garges, 1994
5	*nrd*	ribonucleotide reductase	*Escherichia coli*	Sun and Fuchs, 1994
6	*tdc*	threonine:serine transport and metabolism	*Escherichia coli*	Wu and Datta, 1995

supercoiling in response to changes in the environment can have the effect of increasing the selection for a conserved gene order because genes must be adapted to more than one context. A lack of these additional contexts does not eliminate the need to maintain a gene's context, it just makes such maintenance less stringent, at least in theory. The concern raised by Cook et al. (1989) that DNA supercoiling in bacteria may not significantly change due to changes in the environment does not perturb the local context model. Related to this question is whether the domains that make up the bacterial nucleoid are regulated to different supercoiling levels. Two independent studies conclude that they are not (Miller and Simons, 1993; Pavitt and Higgins, 1993) but again, whether or not domains are regulated to different levels should not affect the local context model. The forces affecting a gene's context, transcriptional trespassing and changes in local supercoiling due to transcriptional activity of neighboring genes, still apply regardless of the average supercoil level of the domain in question.

These latter two studies do, however, present results that challenge the local context model. Both studies measure the expression of supercoil-dependent fusion genes at different positions around the chromosome of either *E. coli* (Miller and Simons, 1993) or *S. typhimurium* (Pavitt and Higgins, 1993). The local context model predicts that the expression of these fusion genes should differ at the different locations but in both studies, the only significant difference in expression is claimed to be due to gene dosage effects caused by position relative to *oriC*. Closer examination of their results does not bear this out.

Miller and Simons (1993) used two promoters (the *lac* operon and *gyr*A promoters) affected oppositely by changes in DNA supercoiling, each fused to a reporter gene. By measuring the expression of both gene fusions inserted at the same locus on the *E. coli*

chromosome, they could detect differences in expression caused by different levels of supercoiling. Miller and Simons did detect differences in the gene expression ratio as great as 45% but they dismiss this as indicating an insignificant difference in supercoiling at the different chromosomal loci studied. For their purposes maybe so, since they were concerned with detecting stable differences in DNA supercoiling between domains. The local context model is concerned with any difference in local supercoiling which can affect the expression of genes, no matter how small or transient. The differences in gene expression detected by Miller and Simons may be all that is necessary to maintain a selective advantage for one particular gene arrangement since it is known that even very small changes in the fitness of a bacterium can lead to rapid extinction (Hill and Harnish, 1981; 1982).

A similar scenario occurs with the study by Pavitt and Higgins (1993). A different promoter-reporter gene combination was used compared to the previous study but otherwise it is very similar. Again, differences in the observed expression of the fusion gene product was attributed to position relative to *oriC*. However, anomalous expression results (either high or low) were detected that could not be explained this way. These anomalous data were said to be due to local effects on transcription and were eliminated from their analysis. For Pavitt and Higgins then, the experimental noise of their study is the very effect which is central to the local context model. Thus this study tends to support our model rather than refute it.

Apart from a requirement that gene expression be sensitive to changes in DNA supercoiling, another assumption of the local context model is that the level of supercoiling actually changes at or near the site of the promoter in vivo. It does not matter how sensitive a promoter is to supercoiling if it is shielded in some way from the effects of

neighboring transcription bubbles or other processes which can perturb the local context This is of concern because, as we mention in this chapter, certain nucleoid proteins can bind to DNA to constrain supercoils. Indeed, approximately half of the supercoils of the *E. coli* chromosome are constrained in some way (Pettijohn and Pfenninger, 1980; Bliska and Cozzarelli, 1987). It has been shown that nucleoid proteins bind at the surface of the nucleoid which is where transcription is taking place (Ryter and Chang, 1975; Dürrenberger et al., 1988; Dürrenberger et al., 1991). This suggests that the DNA in the regions of genes being actively expressed may be constrained so that local changes in supercoiling are negated.

H-NS is a nucleoid protein that affects the expression of a large number of unlinked genes in *E. coli* and where it was investigated, these genes are also affected by changes in supercoiling (Higgins et al., 1990). H-NS binds preferentially to DNA forming a sequence-induced curve in its tertiary structure (Yamada et al., 1990) and such DNA has been found in the upstream region of promoters affected by H-NS (Yoshida et al., 1993). Two models of H-NS action have been proposed; 1) that the protein interacts directly with specific sequences in target promoters to influence gene expression (Göransson et al., 1990) or 2) that it binds to DNA to change the local supercoil level. This change in supercoiling would then be the direct cause in altered gene expression (Higgins et al., 1988). I propose that elements of both models are correct and that gene regulation by H-NS still allows local context to affect the expression of genes under its control.

Curved DNA upstream of a promoter can increase expression of that promoter in both *E. coli* and *B. subtilis* (McAllister and Achberger, 1988; Hsu et al. 1991). It is thought that the curved DNA wraps around RNA polymerase to stabilize its binding to the promoter thus making initiation more efficient (see Pérez-Martin et al., 1994 for a review).

H-NS will also bind to this curved DNA (Yoshida et al., 1993). This may block or otherwise interfere with the binding of RNA polymerase to the promoter and it would certainly make the curved DNA unavailable for assisting in RNA polymerase stabilization. This explains the down-regulation effect of H-NS on numerous loci (Yoshida et al., 1993; O'Byrne and Dorman, 1994) and supports the model that H-NS binding directly affects gene expression.

The enhancement of expression afforded by curved DNA interacting with RNA polymerase has been found to be greatly reduced or abolished when the DNA is relaxed as opposed to supercoiled (Bracco et al., 1989; Gartenberg and Crothers, 1991; Nickerson and Achberger, 1995). A study in *B. subtilis* also found that the spacing of the curved DNA is important (McAllister and Achberger, 1989). Changing the spacing between the curved DNA and promoter by anything but multiples of one helical repeat impaired promoter function. Since changes in supercoil level change the spacing between segments of DNA, this strongly suggests that the level of supercoiling will affect the strength of the interaction between the curved DNA and RNA polymerase. Thus, the susceptibility to changes in supercoiling that H-NS-influenced promoters show is explained. With H-NS bound to the curved DNA upstream of the promoter, RNA polymerase is blocked or binds and initiates at the promoter at a reduced rate. When H-NS is removed, RNA polymerase can use the curved DNA to increase binding and initiation to a degree influenced by the local supercoil level.

The use of curved DNA by RNA polymerase is not limited to those genes affected by H-NS. Recognition sequences for the catabolite activator protein (CAP) can be replaced with naturally curved DNA to activate genes normally controlled by CAP (Bracco et al., 1989; Gartenberg and Crothers, 1991) suggesting that it is the DNA

bending ability of CAP that is important for initiation. The same is true for the transcription activator NR_I (Brahms et al., 1995; Révet et al., 1995). The wide use of curved DNA by bacterial promoters and its apparent dependence on supercoiled DNA for activity provides a specific example of how changes in the local context of a gene can change its level of expression.

A second example exists in the form of the spacing between the -10 and -35 regions of bacterial promoters. Promoter activity is strongly affected by the twist angle between these two elements which is influenced by the number of nucleotides separating them (spacer length) and the level of local supercoiling (Borowiec and Gralla, 1987; Aoyama and Takanami, 1988). Wang and Syvanen (1992) have argued that this sensitivity to supercoiling means that promoters with short spacers (less than 17 bp) will increase their activity as the level of supercoiling is reduced while promoters with long spacers (above 17 bp) will increase in activity as supercoiling is increased. Promoters with intermediate length spacers are predicted to show highest activity at intermediate supercoil levels. A study investigating the effects of supercoiling on the expression of proteins in *E. coli* has been used to investigate this hypothesis (Steck et al., 1993). In this study, differences in the twist angle between the -10 and -35 regions of promoters were found to correlate with differences in expression levels in a *topA* mutant (increased DNA supercoiling) and a *gyrB* mutant (decreased DNA supercoiling) in agreement with Wang and Syvanen's predictions. Since the deviations in supercoiling levels caused by the mutations (±20% from mean) were within the range observed for wild type cells grown under different conditions, it is likely that these changes in gene expression reflect responses to supercoil levels that actually occur in nature. Steck et al. (1993) also shows that quite modest changes in supercoiling can have an effect on gene expression.

While a great deal of work has been done that indicates that gene expression can be influenced by changes in supercoiling (Drlica, 1987; Dorman, 1995), and that upstream and downstream genes can also affect a gene's expression (Rahmouni and Wells, 1992; Tan et al., 1994), much less has dealt specifically with the effects of altering the position and orientation of genes relative to their neighbours. Nevertheless, this is an important point to consider since so much of the local context model rests upon the assumption that moving genes does indeed alter their expression.

One way to test this assumption is by altering the relative positions of specific genes on chromosomal DNA in an otherwise wild-type host. Chromosomal loci should be used to avoid any possible differences in effects between chromosomal and plasmid DNA. Topoisomerase mutants and topoisomerase-affecting antibiotics should be avoided since these necessarily impact on the normal regulatory mechanisms that control supercoiling within cells, often in very artificial ways, and the object of this investigation would be to quantitate changes in gene expression that could occur in a natural environment. Reporter genes with expression levels that can be quantitatively measured would be fused to supercoil-sensitive promoters and linked in different orientations and orders together with a marker gene. The resulting 'expression cassettes' would be integrated into specific sites dispersed around the host cell's chromosome. Differences in expression of the reporter genes from different cassettes integrated at the same locus would indicate that altering gene order does indeed affect gene expression. Additionally, strains with different cassettes integrated into the same locus, or the same cassette integrated into different loci, could be brought into direct competition by growing more than one strain together in a chemostat. If the expression of genes flanking the cassettes have been altered, the changes may be enough to impart a competitive advantage to one of the strains which would then

come to dominate the chemostat over time. It would be important to use more than one chromosomal locus as the integration site in this case since there seems to be no reasonable justification to assume that the expression of every gene in a genome must be influenced by neighbouring genes. Indeed, the tightly coupled interdependencies of regulatory circuits and the large numbers of genes that may be involved in a particular metabolic pathway means that this is not a necessary condition for the local context model to operate in any case. A single gene sensitive to local changes in supercoiling may have a profound effect on the cell it resides in depending on the particular role it plays.

It is important to perform such tests of the local context model since it is not the only theoretical framework that is consistent with the observed data. A strictly gradual accumulation of rearrangements over time does not seem to be consistent for at least some lineages although in others insufficient data are available to rule out this possibility. The essence of this chapter deals with rates of genetic recombination, and most recombination—be it homologous or illegitimate—is at least indirectly influenced by genes involved in DNA repair (Mahajan, 1988; Radding, 1988; Ishiura et al., 1989; 1990). Variations in mechanisms, rates and specificities across and within lineages may contribute significantly to the observed patterns of conservation and divergence in genomic comparisons. Whatever the specific mechanisms, the forces that promote and inhibit genomic change almost certainly operate through complex interactions with one another that will be both troublesome and rewarding to investigate.

CHAPTER 5

Genomic stability in the archaea *Haloferax volcanii* and *Haloferax mediterranei*

This chapter is published as:

López-García, P., A. St. Jean, R. Amils, and R.L. Charlebois. 1995. Genomic stability in the archaea *Haloferax volcanii* and *Haloferax mediterranei*. J. Bacteriol. 177:1405-1408.

Abstract

Through hybridization of available probes, we have added nine genes to the macrorestriction map of the *Haloferax mediterranei* chromosome, and five genes to the contig map of *Haloferax volcanii*. Additionally, we hybridized 17 of the mapped cosmid clones from *H. volcanii* to the *H. mediterranei* genome. The resulting 35-point chromosomal comparison revealed only two inversions and a few translocations. Forces known to promote rearrangement, common in the haloarchaea, have been ineffective in changing global gene order throughout the nearly 10^7 years of these species' divergent evolution.

My Contribution

This work was done in collaboration with R. Amils' lab in Spain. I prepared Southern blots of the minimal cosmid set of the *Haloferax volcanii* genome for use in Spain. In turn, I received Southerns of pulsed-field gels of the *Haloferax mediterranei*

genome which I hybridized with probes to place the nine genes mentioned in the abstract and listed in Table 5.1 on the macrorestriction map. I used these data combined with the data generated in Spain to produce the alignment of the two chromosomes shown in Fig. 5.1.

Introduction

One of the most notable characteristics of extremely halophilic archaea is their genetic instability (Charlebois and Doolittle, 1989; DasSarma, 1989; Derkacheva et al., 1993; Pfeifer, 1988). The haloarchaea are rich in insertion sequences which can disrupt genes at frequencies as high as 10^{-2} in the case of *Halobacterium salinarium* (Weidinger et al., 1979; Pfeifer et al., 1981b). Most of the activity of these insertion sequences is confined to plasmid DNA or to FII DNA (which has a lower moles percent G+C content) (Pfeifer et al., 1982; Ebert and Goebel, 1985;Pfeifer and Betlach; 1985), but chromosomal genes are not spared from disruption. The *bop* gene, for instance, is inactivated by at least eight different types of insertion sequences (Pfeifer, 1988; DasSarma, 1989), resulting in a combined risk of about 10^{-4} per generation (Pfeifer et al., 1981b). *H. salinarium* is known to possess hundreds of insertion sequences, in dozens of families (Sapienza and Doolittle, 1982). The resulting transpositional cost to the cell is compounded by the potential recombinational chaos mediated by interaction between members of each insertion sequence family. It has been suggested that a genomically pure clone of *H. salinarium* cannot be grown because of continual rearrangements which occur in plasmid DNA (Sapienza et al., 1982; Pfeifer, 1988; Pfeifer et al., 1989).

The genus *Haloferax* is not as severely infested with insertion sequences as is *H. salinarium*. Nevertheless, *Haloferax volcanii* possesses at least 49 copies of the ISH51

family distributed throughout the genome (Cohen et al., 1992), and there is good evidence for the existence of several other types of repeated sequences as well (Sapienza and Doolittle, 1982; Schalkwyk et al., 1993). Though *H. volcanii* is not as prone to genetic disruption as is *H. salinarium*, neither is it immune. Less is known about the number or distribution of insertion sequences in *Haloferax mediterranei*. Interestingly, the populous ISH51/27 family shared between *H. volcanii* and *H. salinarium* (Pfeifer and Blaseio, 1990) is absent from *H. mediterranei* (Schalkwyk et al., 1993). Regardless, repeated sequences which can potentially facilitate genomic rearrangement are present (Antón et al., 1994).

With physical and genetic maps available, we can now begin to address genomic stability in the haloarchaea. The first researchers to assess the degree of rearrangement in the haloarchaeal chromosome by using a comprehensive comparative mapping approach (Hackett et al., 1994) found that much of the *H. salinarium* chromosome is conserved in structure among several strains, despite different complements of repetitive sequences in their genomes. Differences in the maps were observed to be confined to a few discrete, hypervariable blocks. *H. salinarium* NRC-1 and *H. salinarium* S9 maintain hundreds of insertion sequences within their genomes, and yet gene order is preserved relative to that of *H. salinarium* GRB, which contains virtually no repetitive sequences. The work was a clear demonstration of the existence of map inertia in the haloarchaea, at least over the short period of time—evidenced by a high conservation of restriction sites—since the strains diverged from one another.

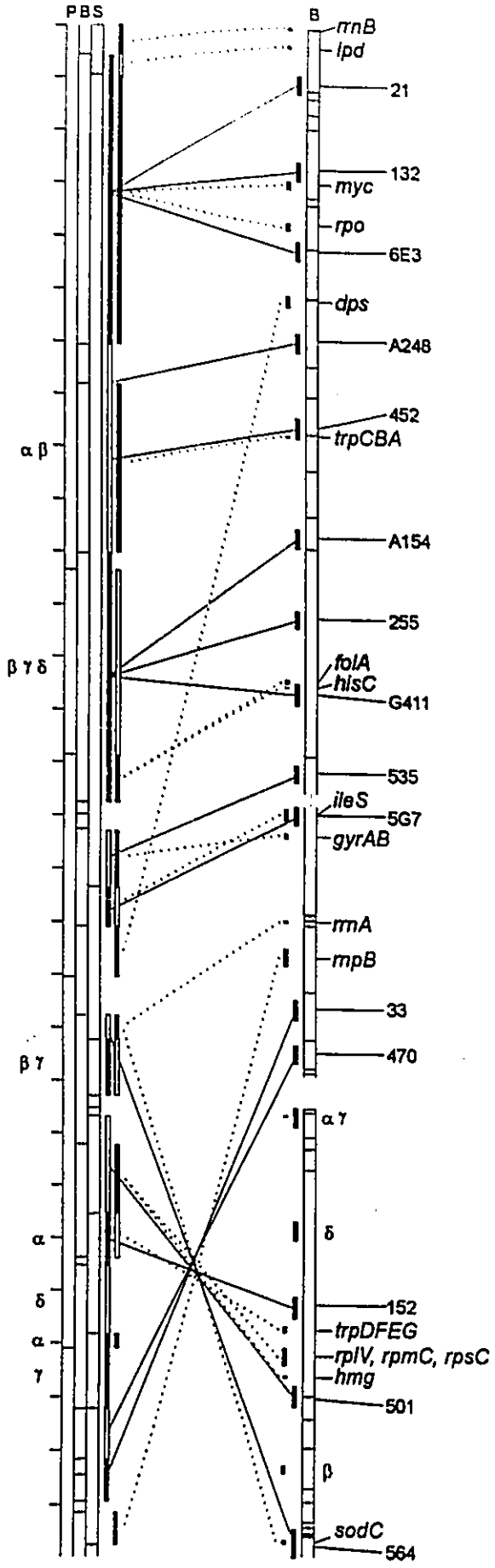
Two other haloarchaeal genomic comparisons, spanning larger time scales, are currently feasible. Comparison of the *H. salinarium* GRB (St. Jean et al., 1994) and *H. volcanii* DS2 (Charlebois et al., 1991) contig maps will help us to examine genomic

structural divergence at the genus level. Before undertaking this project, however, it was useful to assess genomic stability at the species level. This paper focuses on a comparison of the contig map of the *H. volcanii* DS2 genome (Charlebois et al., 1991) and the chromosomal macrorestriction map of *H. mediterranei* ATCC 33500 (López-García et al., 1992; Antón et al., 1994), which we have here supplemented with additional loci to facilitate alignment. Numerical taxonomy (Torreblanca et al., 1986) and DNA hybridization studies (Gutiérrez et al., 1989) clearly categorize *H. volcanii* and *H. mediterranei* as distinct species. A 98.4% similarity in 16S rRNA sequence (Kamekura and Seno, 1992) implies a divergence time of about 80 million years (Ochman and Wilson, 1987). We show here that despite sufficient opportunity for genomic rearrangement, the chromosomal maps of *H. volcanii* and *H. mediterranei* are chiefly congruent.

Comparison of the maps

Restriction maps are mutable, and they reflect sequence divergence. The *Bam*HI maps of the *H. volcanii* and *H. mediterranei* chromosomes are consequently dissimilar (Fig. 5.1), but this is not necessarily a result of genomic rearrangement. In order to examine map stability, we had to compare gene order. Since we had only seven shared loci from our previous mapping efforts (Charlebois et al., 1991; Antón et al., 1994), it was necessary to augment the number of these loci in order to improve the resolution. We first elected to complete the mapping by hybridization of available cloned genes as much as possible. With dot blots or Southern blots (St. Jean et al., 1994) of *H. volcanii* cosmid clones, we hybridized (López-García et al., 1993; St. Jean et al., 1994) probes representing *dps*, *gyrA*, *ileS*, a *myc*-like gene, and the S10 ribosomal operon. With pulsed-field gel blots (López-García et al., 1993) of *H. mediterranei* genomic DNA, we

Fig. 5.1. Comparison of the chromosomal maps of *H. volcanii* and *H. mediterranei*. Both chromosomes are circular and are of the same length, 2.9 Mbp. On the left, the macrorestriction map of the *H. mediterranei* chromosome is shown, with sites for *Pac*I (P), *Bam*HI (B) and *Swa*I (S). On the right, the *Bam*HI (B) map of the *H. volcanii* chromosome is displayed. Dotted interruptions in this latter map indicate regions unmapped for *Bam*HI. Genes and *H. volcanii* cosmid clones with unique positions in the *H. mediterranei* map are linked to the *H. volcanii* map by dotted and solid lines, respectively. Genes and cosmid clones mapping to several locations on the *H. mediterranei* chromosome—*csg* (α), *rplAJL* (β), cosmid G86 (γ), and cosmid 464 (δ)—are shown as symbols without connections (see text). Genes were mapped to *Bam*HI, *Pac*I and *Swa*I *H. mediterranei* fragments, whereas cosmids were mapped only to *Bam*HI and *Swa*I fragments. Where loci could map anywhere within a restriction fragment, we traced lines to the center of the fragment. Tick marks on the scale bar on the left are placed at 100-kbp increments.


PBS
B
rrnB
lpd
21
132
myc
rpo
6E3
dps
A248
452
trpCBA
α β
A154
255
β γ δ
folA
hlsC
G411
535
ileS
5G7
gyrAB
rrnA
mpB
33
470
β γ
α γ
α
δ
δ
152
trpDFEG
α
rplV, rpmC, rpsC
γ
hmg
501
β
sodC
564

hybridized (St. Jean et al., 1994) probes for *dps*, *hisC*, *hmg*, *ileS*, *lpd*, *rplAJL*, *rpo*, *trpCBA* and *trpDFEG*. We thus added nine genetic loci to the *H. mediterranei* macrorestriction map, and five to the contig map of *H. volcanii* (Table 5.1 and Fig. 5.1).

The ordered cosmid clone collection representing the *H. volcanii* genome (Charlebois et al., 1991) supplied ideal physical markers to supplement the limited number of genetic loci available. We chose 17 *H. volcanii* cosmid clones, spaced evenly around the chromosome, to hybridize (López-García et al., 1993) with *H. mediterranei* macrorestriction fragments. These cosmids were not known to carry repeated sequences (Cohen et al., 1992), so that we might avoid extraneous cross hybridization. At high stringency, most of the cosmid probes gave a unique signal. The multiple signals produced by cosmids G86 and 464 (Fig. 5.2) arose either from transpositional fragmentation of the sequences within these cosmids or from events of duplicative transposition. Interestingly, the single-copy *csg* locus from *H. volcanii* maps to three places on the *H. mediterranei* chromosome (Antón et al., 1994). However, cosmid G86, which includes *csg*, hybridizes convincingly only to one of *csg*'s *H. mediterranei* residences. We interpret this to mean that the lower-specific-activity whole cosmid probe requires more extensive lengths of sequence to hybridize visibly. Therefore, we can conclude that there are three regions of the *H. mediterranei* chromosome that have extensive homology with G86, and there are two other regions that have specific homology with *csg*. Duplicative transposition of *csg* is apparent; we cannot rule out other duplicative or nonduplicative transpositions involving short sequences such as those of single genes based on the cosmid-to-chromosome hybridizations, though we can be sure that the bulk of each cosmid's homology has been

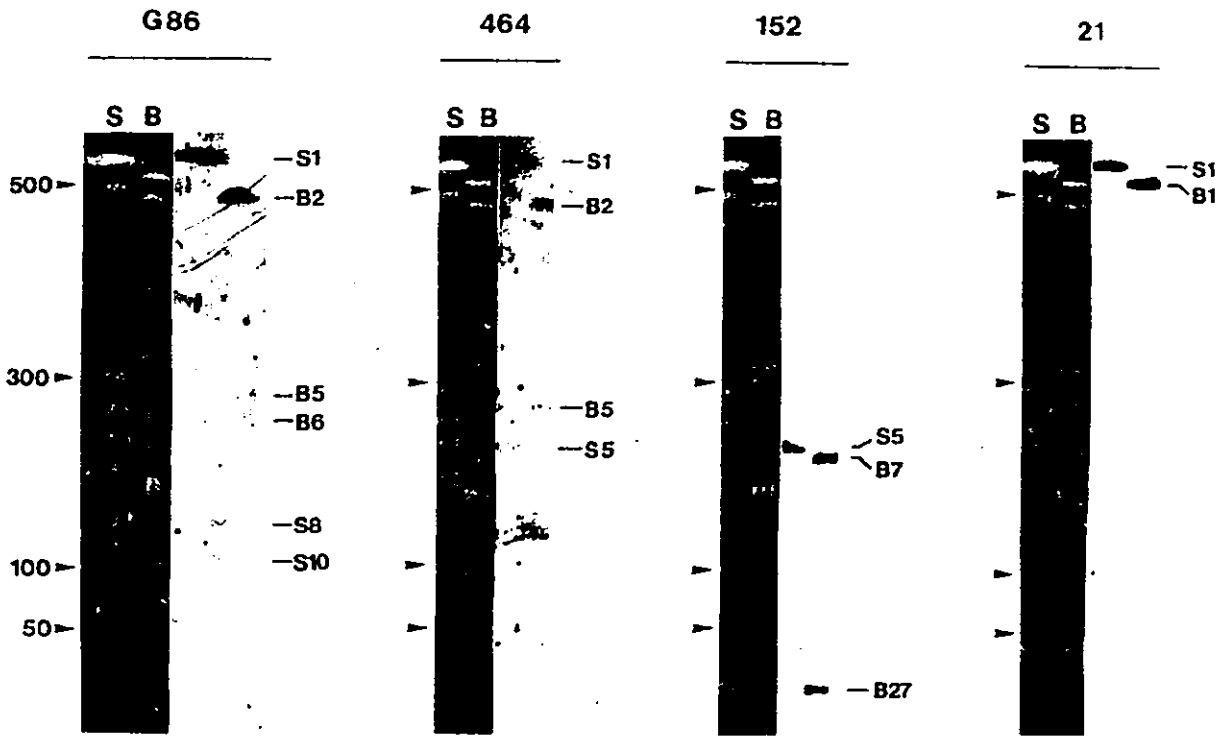
Table 5.1. Locations of genetic markers newly mapped in this study.

Gene or Operon	Description	*H. volcanii* cosmid(s)[a]	*H. mediterranei* *Bam*HI/*Swa*I/*Pac*I bands[b]	Reference
dps	DPS family heat shock	456 and A210	B9b/S3/P3	St. Jean and Charlebois, unpublished
gyrA	DNA gyrase, subunit A	547		Holmes and Dyall-Smith, 1991
hisC	Histidinol phosphate aminotransferase		B2/S1/P3	Conover and Doolittle, 1990
hmg	3-hydroxy-3-methylglutaryl coenzyme A reductase		B7/S7/P2	Lam and Doolittle, 1989
ileS	Isoleucyl tRNA synthetase	5G7	B8/S3/P3	Daniels, 1992
lpd	Dihydrolipoamide dehydrogenase	126	B1/S11/P1	Vettakkorumakankav and Stevenson, 1992
myc	Protein homologous to myc product	460		Ben-Mahrez et al., 1988
rplAJL	Ribosomal protein A operon		B2/S1/P4, B3/S1/P1, and B6/S10/P2	Itoh, 1988
rplV *rpmC* *rpsC*	Ribosomal protein S10 operon	D57 and 266		Spiridonova et al., 1989
rpo	RNA polymerase operon		B1/S1/P1	Leffers et al., 1989
trpCBA	Tryptophan biosynthesis		B3/S1/P1	Lam et al., 1990
trpDFEG	Tryptophan biosynthesis		B7/S5/P2	Lam et al., 1992

[a] See Charlebois et al., 1991.

[b] See Antón et al., 1994.

Fig. 5.2. Southern hybridization of *H. volcanii* cosmid clones with genomic DNA from *H. mediterranei*. Clones G86, 464, 152 and 21 were among the 17 cosmids used as probes in hybridizations with pulsed-field gel-fractionated *Swa*I (S) and *Bam*HI (B) genomic digests of *H. mediterranei*. G86 and 464 map to multiple chromosomal locations, whereas all other cosmids, including 152 and 21, map to single locations. The contour-clamped homogeneous electric field gels shown were run for 40 h at 10 V/cm, with a switching time of 10 s. As size standards, we used lambda concatemers and *Saccharomyces cerevisiae* YN295 chromosomes. Sizes shown on the left are in kilobase pairs.


G86
464
152
21
S B
S B
S B
S B
500
300
100
50
S1
B2
B5
B6
S8
S10
S1
B2
B5
S5
S5
B7
B27
S1
B1

accurately mapped. Evidence for an event of nonduplicative transposition was observed in the case of one of our gene probes, that for *dps*.

An alignment of the enriched chromosomal maps, showing all shared loci derived from this work and from previous work, reveals an unexpected congruence (Fig. 5.1). When we found multicopy loci, one copy always matched the corresponding location on the other chromosome. Thus, apart from a large inversion presumably at the pair of inverted (Charlebois et al., 1991) *rrn* loci, another small inversion involving *ileS* and *gyrAB*, and a few transpositions, the chromosomes are colinear. Previously, during mapping the *H. volcanii* genome (Charlebois et al., 1989), a hybrid cosmid clone which suggested recombination between the *rrn* loci was found (R.L. Charlebois, unpublished). This large inversion may therefore not be a stable difference between the chromosomes.

Forces affecting rearrangement

Insertion sequences are known to cause frequent and extensive rearrangement of plasmid DNA in *H. salinarium*, and they can consequently alter plasmid-encoded phenotypes (Weidinger et al., 1979; Pfeifer et al., 1981b; Pfeifer, 1988). Most transpositional activity is phenotypically invisible (Sapienza et al., 1982), however, which underscores the importance of insertion sequences in the evolution of haloarchaeal genome structure—there must be more underlying activity than we can observe through occasional phenotypic alterations. There is strong physical evidence for such disruptive potential in *Haloferax* spp. as well (Sapienza and Doolittle, 1982; Cohen et al., 1992; Schalkwyk et al., 1993). We have discovered, though, that this destabilizing pressure has been largely ineffective in changing chromosomal organization in *Haloferax* spp. as far as the resolution of our comparison could show. A few minor differences are all that

distinguish the current maps. The genetic core of the *Haloferax* genome, its chromosome, is remarkably stable. Plasmids, which can vary considerably (Rosenshine and Mevarech, 1989; Charlebois et al., 1991; López-García et al., 1992; López-García et al., 1993), remain the predominant seat of haloarchaeal genomic diversity.

Powerful stabilizing forces must exist in the haloarchaeal chromosome to maintain a particular gene order. We know little about the significance of gene order in a replicon. A catalog of genetic functions seems insufficient to describe the genome of an organism, since a highly conserved genetic map implies that the arrangement of genes is important as well. In *Escherichia coli* and *Salmonella typhimurium*, for which the most data exist (Krawiec and Riley, 1990; Riley and Sanderson, 1990; Charlebois and St. Jean, 1995), chromosome structure seems to be influenced by the distance and orientation of genes relative to the origin of DNA replication, by the desirability of terminating bidirectional replication 180° from the origin, by the need to maintain a certain polarity of sequences near the terminus of replication, and by a contextual sensitivity of gene expression. We know nothing about gene orientation, nucleoid structure, or the mechanics of chromosomal replication in the haloarchaea, but it is likely that similar forces influence gene location. In the haloarchaea, plasmid DNA seems to be less constrained by the factors which control the structure of the chromosome. Perhaps herein lie additional clues which will help us to discover what shapes a genome.

Acknowledgements

This work was supported in Madrid by the Spanish Interministerial Commission for Science and Technology (CICYT), and in Ottawa by the Natural Sciences and Engineering Research Council (NSERC).

CHAPTER 6

Comparative genomic analysis of the *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB contig maps reveals extensive rearrangement

This chapter is published as:

St. Jean, A., and R.L. Charlebois. 1996. Comparative genomic analysis of the *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB contig maps reveals extensive rearrangement. J. Bacteriol. **178**:3860-3868.

Abstract

Anonymous probes from the genome of *Halobacterium salinarium* GRB and twelve gene probes were hybridized to the cosmid clones representing the chromosome and plasmids of *Hb. salinarium* GRB and *Haloferax volcanii* DS2. The order and pairwise distances between 35 loci uniquely cross-hybridizing to both chromosomes were analysed in a search for conservation. No conservation between the genomes could be detected at the 15-kbp resolution of this study. We found distinct sets of low-copy-number repeated sequences in the chromosome and plasmids of *Hb. salinarium* GRB indicating some degree of partitioning between these replicons. We propose alternative courses for the evolution of the haloarchaeal genome: (i) that the majority of genomic differences that exist between genera came about at the inception of this group, or (ii) that the differences

have accumulated over the lifetime of the lineage. The strengths and limitations of investigating these models through comparative genomics are discussed.

My Contribution

Apart from the credits given in the acknowledgements and advice and suggestions given by R.L. Charlebois and my supervising committee, this chapter is all mine. I performed all the experimental procedures and analyses of the resulting data, I wrote the program COMPAGEN to aid in the analysis of the data and I am the principal author of the manuscript .

Introduction

In recent years much effort has been devoted to genome-level analysis among prokaryotes (Cole and Saint Girons, 1994; Fonstein and Haselkorn, 1995). Most of this effort involved the use of pulsed-field gel electrophoresis (PFGE), either in the construction of physical maps or in the direct comparison of restriction fragment patterns. The latter method allows for a large number of genomes to be quickly and relatively easily compared but suffers from poor resolution, though it has been useful in the typing of strains (Grothues and Tümmler, 1991; Huber and Selenska-Pobell, 1994; Lück et al., 1995). A large number of studies has been conducted using PFGE to construct physical maps (for a review see Cole and Saint Girons, 1994). Most often, genetic markers are then localized to specific regions of the map through hybridization allowing the investigation of gross rearrangements at the genomic level. Genetic loci occurring on the same restriction fragment, however, cannot be ordered on the map, masking any differences in their arrangement. This limitation restricts such comparisons to closely related genomes.

The majority of genomic comparisons performed to date use PFGE-derived maps to compare genomes at the strain or species level. The results of these comparisons have prompted their division into two groups (Fonstein and Halelkorn, 1995), organisms with highly conserved genetic maps and those with divergent maps. Examples of members of the former group include *Escherichia coli* and *Salmonella typhimurium* (Riley and Krawiec, 1987), *Borrelia* spp. (Ojaimi et al., 1994; Casjens et al., 1995), *Clostridium perfringens* (Canard et al., 1992), *Lactococcus lactis* (Le bourgeois et al., 1995), *Mycoplasma* spp. (Ladefoged and Christiansen, 1992; Peterson et al., 1995), and *Streptomyces lividans* (Leblond et al., 1993). Members of the second group include *Bacillus* spp. (Carlson et al., 1992; Carlson and Kolstø, 1993), *Rhodobacter* spp. (Fonstein et al., 1995), and *Leptospira interrogans* (Zuerner et al., 1993). The criteria for deciding in which group a comparison will be included has not been rigorously established, however. Typically, no objective measure is used to determine the degree of similarity between genomes which makes relating the results of one comparison to those of another problematic. Indeed, genomes showing a moderate number of differences could be considered either conserved or divergent, depending on the context of the study.

Of the more than 100 chromosomal maps available, only 10 are from the domain Archaea. To date, three archaeal genomic comparisons have been performed, one between two methanogens (Stettler et al., 1995) and two between different members of the haloarchaea (Hackett et al., 1994; López-García et al., 1995). The most detailed maps among the Archaea are derived from the extreme halophiles. Maps of ordered cosmid libraries are available for *Haloferax volcanii* DS2 (Charlebois et al., 1991) and *Halobacterium salinarium* GRB (St. Jean et al., 1994), while highly detailed macrorestriction maps constructed by using two-dimensional gel electrophoresis are

available for two additional strains of *Hb. salinarium* (Hackett et al., 1994). The higher resolution of these maps compared to what is achieved with standard PFGE maps allows for more detailed and potentially informative comparisons to be made.

Comparison among the three *Hb. salinarium* strains showed almost complete conservation in the physical maps, with many identical restriction sites being found in all three. Only two variable regions were identified, the first spanning 240 kbp involving numerous changes to the restriction map, a large insertion-deletion and an inversion and the second an insertion-deletion spanning approximately 10 kbp (Hackett et al., 1994). While one would normally expect such conservation when looking at strains of the same species, the genetic instability of two of the strains in the comparison did not make this a forgone conclusion. Many strains of *Hb. salinarium* harbor active insertion sequences of different types in up to hundreds of copies (Sapienza and Doolittle, 1982). These elements are known to cause frequent insertional inactivation of chromosomal and plasmid genes (Charlebois and Doolittle, 1989; Dassarma, 1989; Derkacheva et al., 1993). For some time, the genetic instability of *Hb. salinarium* was assumed to apply to the physical map as well (Sapienza et al., 1982). We now know that the map can be preserved despite the potential for rearrangement. Among members of the domain Bacteria, chromosomal rearrangements between repeated DNA sequences, both *rrn* operons and insertion sequences, are not uncommon (Krawiec and Riley, 1990; Liu and Sanderson, 1995a; Zuerner et al., 1993).

A second comparison involving two halophilic Archaea of different species, *Haloferax volcanii* and *Haloferax mediterranei*, also found highly conserved maps (López-García et al., 1995). In this case, two inversions (one involving the two *rrn* operons found in this genus) and one transposition involving a single locus were found,

while the *Bam*HI restriction maps of the two chromosomes were completely different, as expected from interspecific sequence divergence. Because only a macrorestriction map exists for *Hf. mediterranei*, some of the 35 probes used could not be ordered on its map. Within the resolution of the comparison, however, no other differences could be detected. This is despite the fact that *Hf. volcanii* is known to possess active insertion sequences though not in the same numbers as in some strains of *Hb. salinarium* (Charlebois and Doolittle, 1989; Schalkwyk et al., 1993).

Because of the degree of conservation found in these two comparisons, we wished to discover if similar conservation applied to more distantly related halophilic Archaea. If the maps were conserved, an alignment would allow us to localize the homologs of cloned genes whose sequences might have diverged too much to be accessible through hybridization or PCR. Whether the maps were conserved or scrambled, the comparison would provide useful data for the study of forces which maintain or disrupt gene order (Charlebois and St. Jean, 1995). Of particular interest was the tempo and mode of genome-level change, set in a phylogenetic context. Such quantitation required implementing analyses which could provide a more objective measure of the degree of similarity between the genomes being compared than have been used in the past. The availability of detailed maps and cosmid libraries for the chromosome and plasmids of *Hf. volcanii* DS2 and *Hb. salinarium* GRB made these two organisms the logical choice for this next haloarchaeal genomic comparison.

Materials and Methods

Archaeal strains and cosmid libraries

DNA was obtained from the previously prepared ordered cosmid libraries of *Haloferax volcanii* DS2 (Charlebois et al., 1991) and *Halobacterium salinarium* GRB (St. Jean et al., 1994). These libraries included the chromosomes and the three largest plasmids found in each archaeon.

DNA dot blot and Southern blot preparation and hybridization

Dot blots were prepared by using the minimal cosmid libraries of *Hf. volcanii* DS2 and *Hb. salinarium* GRB. Approximately 50 ng of each cosmid was mixed with an NaOH solution to a final concentration of 0.4 M and spotted onto GeneScreen nylon membranes (DuPont). For Southern blots used to verify ambiguous dot blot signals, cosmid DNA was digested using *Mlu*I, *Bam*HI or a combination of both enzymes. Southern blots also used GeneScreen membranes and were prepared with approximately 0.5 μg of cosmid DNA per lane. DNA was transferred to the membranes by using a Tyler VT-20 vacuum transfer unit according to Tyler's protocol.

Hybridizations were performed as described in St. Jean et al. (1994) with some minor changes. The prehybridization and hybridization temperatures were always 40°C, while the 1-h wash with 2X SSC (1X SSC is 0.15 M NaCl plus 0.015 M sodium citrate)—1% sodium dodecyl sulfate was done at 70°C. Most probes were fragments prepared from cosmid DNA digested with *Mlu*I, *Bam*HI or both and isolated from agarose gels by using GeneClean (Bio101). Twelve genes cloned from various organisms (listed in Charlebois et

al., 1991; St. Jean et al., 1994; López-García et al., 1995) were also used as probes. In both cases probes were prepared by the random-priming method.

For higher-resolution mapping of the six pairs of probes that could not be ordered with the dot blots, Southern blots of the relevant cosmids plus flanking cosmids were prepared. Cosmids were digested with various combinations of one, two or three of the following enzymes: *Bam*HI, *Bgl*II, *Dra*I *Eco*RI, *Hind*III, *Mlu*I, and *Ssp*I for *Hf. volcanii* cosmids and *Afl*II, *Bam*HI, *Bgl*II, *Eco*RI, *Hind*III, *Mlu*I and *Xho*I for *Hb. salinarium* cosmids. Each Southern blot was then hybridized with the appropriate pair of unordered probes. Partial restriction maps of the cosmids were prepared by using data from the digestions and the hybridizations so that the probes could be ordered on the genome.

Computer and Statistical Analysis

The computer program DERANGE II (provided by M. Blanchette and D. Sankoff, Université de Montréal) was used to determine the number of changes (inversions, transpositions and inverted transpositions) necessary to transform one chromosome into the other given the order on each chromosome of a set of homologous loci. Data were run through DERANGE II with a variety of parameters: values of 4, 5 and 6 were used for "look ahead," while weights for transpositions and inverted transpositions were 1, 2, 2.5, 4 and 10. Values of 1 for inversion weight and 0 for length coefficients were used in all cases. The results were compared to values obtained by using 100 random permutations, and the significance level of the difference between the experimental and random results was calculated as 1 plus the number of randomized permutations having a "total cost" or "total number of moves" less than or equal to that for the experimental data, divided by 1 plus the number of randomizations. Total cost and total number of moves are values

produced by DERANGE II that measure the degree of divergence between two DNA segments.

The performance of DERANGE II was tested with permutations containing known numbers of inversions. Successive random inversions were introduced into permutations of 35 loci until a total of 60 inversions had been done. Ten such sets of permutations were constructed. DERANGE II was then used to solve each permutation, and the results were averaged between sets. For this test, look ahead was set to 6 and transposition and inverted transposition weights were set to 2.5 or 10.

Conservation in the distances between loci on the two chromosomes was investigated by calculating every pairwise distance between the loci included in the test. These values were plotted, and a regression analysis was performed to look for any correlation between the two chromosomes.

Results

Hybridization of probes to dot and Southern blots

The ordered cosmid libraries of *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB were used to prepare DNA dot blots representing the chromosome and the three largest plasmids of each genome: pHV4 (690 kbp), pHV3 (440 kbp) and pHV1 (86 kbp) for *Hf. volcanii* and pGRB305, pGRB90 and pGRB37 for *Hb. salinarium*. *Hb. salinarium* cosmids were digested with *Mlu*I, *Bam*HI or both enzymes, and selected anonymous fragments were used to probe the dot blots. Positive controls consisted of hybridization to the cosmid from which the probe was taken, and an equimolar amount of lambda DNA provided the negative control.

DNA for probes was taken from *Hb. salinarium* because it possesses fewer repeated sequences than does *Hf. volcanii* (St. Jean et al., 1994), making interpretation of the results simpler. Also, there was concern that many probes from the 4.1-Mbp *Hf. volcanii* genome would not hybridize to the 2.5-Mbp *Hb. salinarium* genome. A total of 143 anonymous probes ranging in size from 0.4 to 16 kbp were hybridized in this way, 120 from the chromosome of *Hb. salinarium* and 23 from its plasmids. Probes were as evenly distributed around the genome as possible (usually two probes per cosmid). In addition to these, 12 previously cloned genes were hybridized to both genomes. The dot blots allowed the hybridization signals to be localized to roughly a third of a cosmid, either the middle non-overlapping portion, or within the regions of overlap with neighboring cosmids. This provided average resolutions of 15 kbp for *Hb. salinarium* and 14 kbp for *Hf. volcanii*, depending on the sizes of the individual cosmids and the degrees of overlap with their neighbors.

Of the 143 anonymous probes that had been hybridized to the dot blots, 74 were found to give ambiguous hybridization signals (Fig. 6.1). To resolve these ambiguities, sets of Southern blots were prepared which included every cosmid showing an equivocal signal for a particular probe on the dot blots, plus positive and negative controls. We confirmed or refuted signals from 46 probes in this way.

In six instances, a pair of probes from different parts of the *Hb. salinarium* chromosome hybridized to a common locus. To dissect these signals, Southern blots of cosmids from the four unresolved loci on the *Hf. volcanii* chromosome and the two on the *Hb. salinarium* chromosome were prepared. Single, double and triple digests of the cosmids allowed partial restriction mapping of these cosmids, ordering three pairs of signals.

From the total of 155 hybridizations performed, 127 gave reliable results after the additional screening described above (Table 6.1). Of these, 70 probes cross-hybridized between the two genomes, including twelve gene probes. The sensitivity of the hybridization procedure used was tested by hybridizing a probe for ISH51 (the best characterized insertion sequence family in *Hf. volcanii* [Cohen et al., 1992]) to the dot blots of both genomes (Fig. 6.2). No signals were observed for *Hb. salinarium* (strain GRB lacks the closely related ISH27 found in some other strains [Pfeifer and Blaseio, 1990]), and all but one previously identified copy of the element (Cohen et al., 1992) was found in *Hf. volcanii*. Since ISH51 sequences can differ by at least 15% (Hofman et al., 1986), this demonstrated that the procedure being employed could find homologous though divergent loci.

Repeated sequences

Twenty-nine probes gave multiple signals on one or both genomes (never numbering above five per genome) (Fig. 6.3). In a previous study using whole cosmids as hybridization probes (St. Jean et al., 1994), we identified five duplicated sequences in the *Hb. salinarium* GRB genome, four of them within or between plasmids. In that survey, hybridizations involving all cosmids representing plasmid DNA and cosmids representing 40% of the chromosome resulted in a conclusion that this strain's genome is quite repeat-poor, in contrast to those of most other characterized strains. Here, our probes sampled 19% of the genome: 20% of the chromosome and 14% of the plasmid sequences. Ten of the 113 chromosomally derived probes hybridized to between two and five loci in the chromosome (none hybridized to plasmid DNA), whereas 6 of the 14 plasmid-derived probes hybridized to two or three plasmid loci and another plasmid-derived probe

Fig. 6.1. Example of DNA dot-blot hybridization of the *Hb. salinarium* and *Hf. volcanii* genomic cosmid libraries giving ambiguous signals and the Southern blot used to resolve the ambiguities. (a) Dot blots were probed with a gel-isolated fragment of *Hb. salinarium* cosmid G19D7 giving a strong homologous hybridization in the overlap with an adjacent cosmid (numbered 1). Positive signals from two additional pairs of adjacent cosmids are shown with arrows; four weaker signals are numbered 2 to 5. (b) Southern blot of cosmids giving equivocal signals in the dot blots. The homologous hybridization is in lane 1, while the four questionable signals are in lanes 2 to 5. (Lane numbers correspond to the signal numbers in panel a.) Lane 6 is a negative control producing no signal on the dot blots. Three of the four ambiguous cosmids (lanes 3-5) produced signals on the Southern blot. Since the dot blot signal numbered 2 could not be reproduced on the Southern blot, we excluded it from the genomic comparison. The band to the left of lane 1 in panel b is a nonspecific hybridization signal to a lambda marker band.

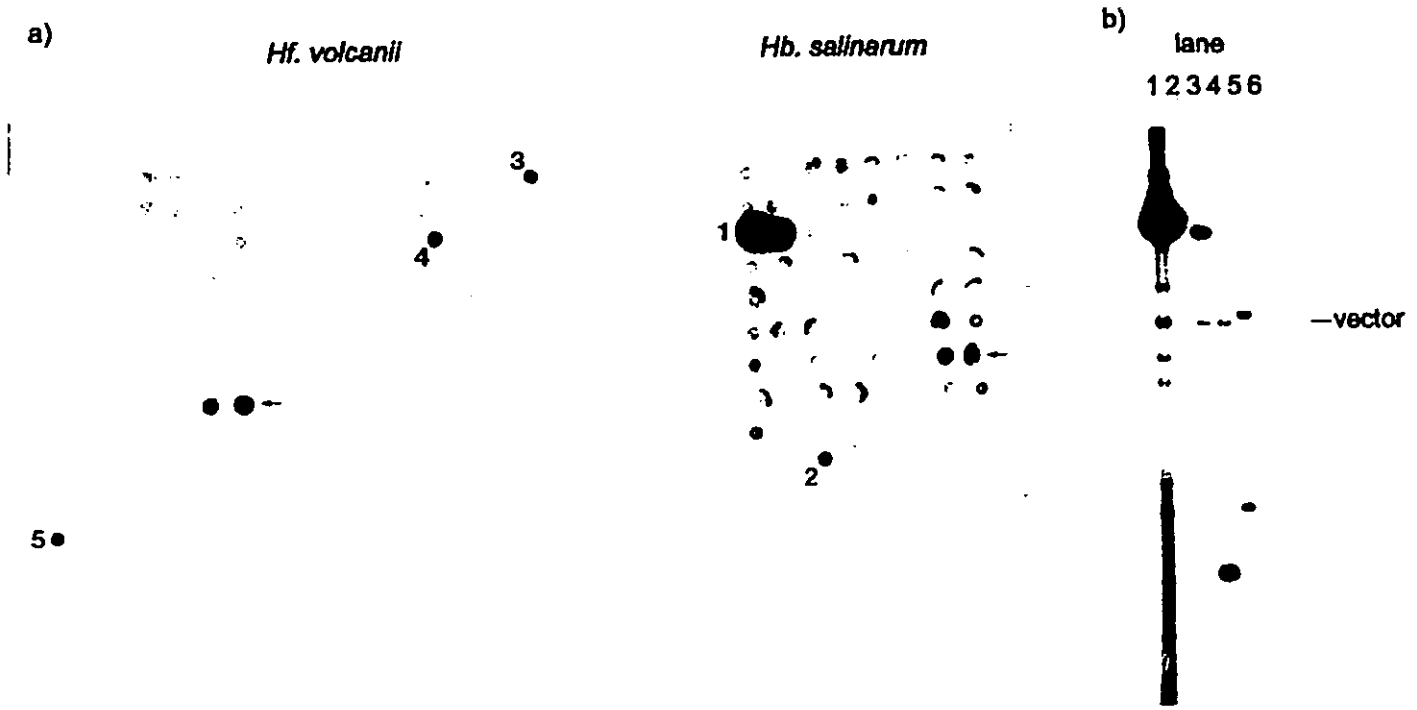

a)
Hf. volcanii
Hb. salinarum
b)
lane
1 2 3 4 5 6
3
4
1
2
5
—vector

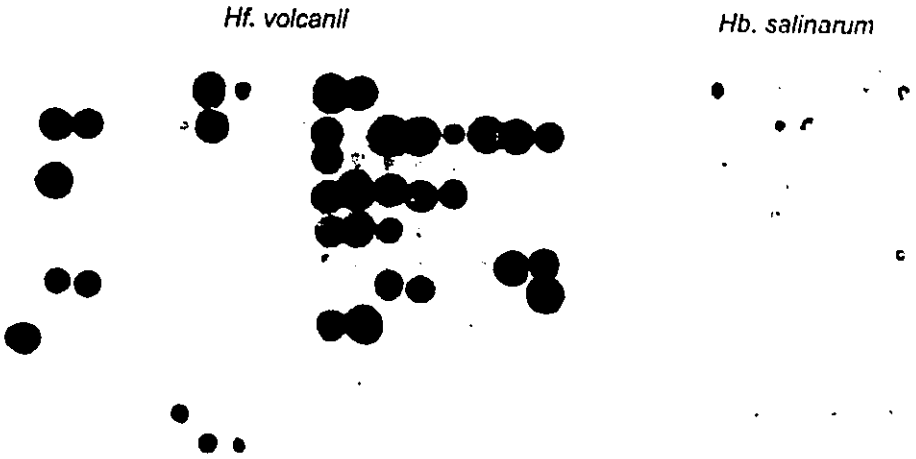
Table 6.1. Summary of hybridization results.

Source of Probe		Number of Probes	Unambiguous Hybridizations	Unambiguously Cross-Hybridizing Probes
Chromosome	anonymous probe[a]	120	101	53
	gene probe[b]	12	12	12
Plasmid	anonymous probe[a]	23	14	5
Total		155	127	70

[a] Restriction fragments from cosmids representing the *Hb. salinarium* genome.

[b] Previously cloned from various haloarchaea (Charlebois et al., 1991; López-Garcia et al., 1995; St. Jean et al., 1994).

Fig. 6.2. Hybridization of the insertion sequence ISH51 to dot blots of cosmid libraries of the *Hf. volcanii* and *Hb. salinarium* genomes. This insertion sequence does not occur in *Hb. salinarium* GRB.


Hf. volcanii
Hb. salinarum

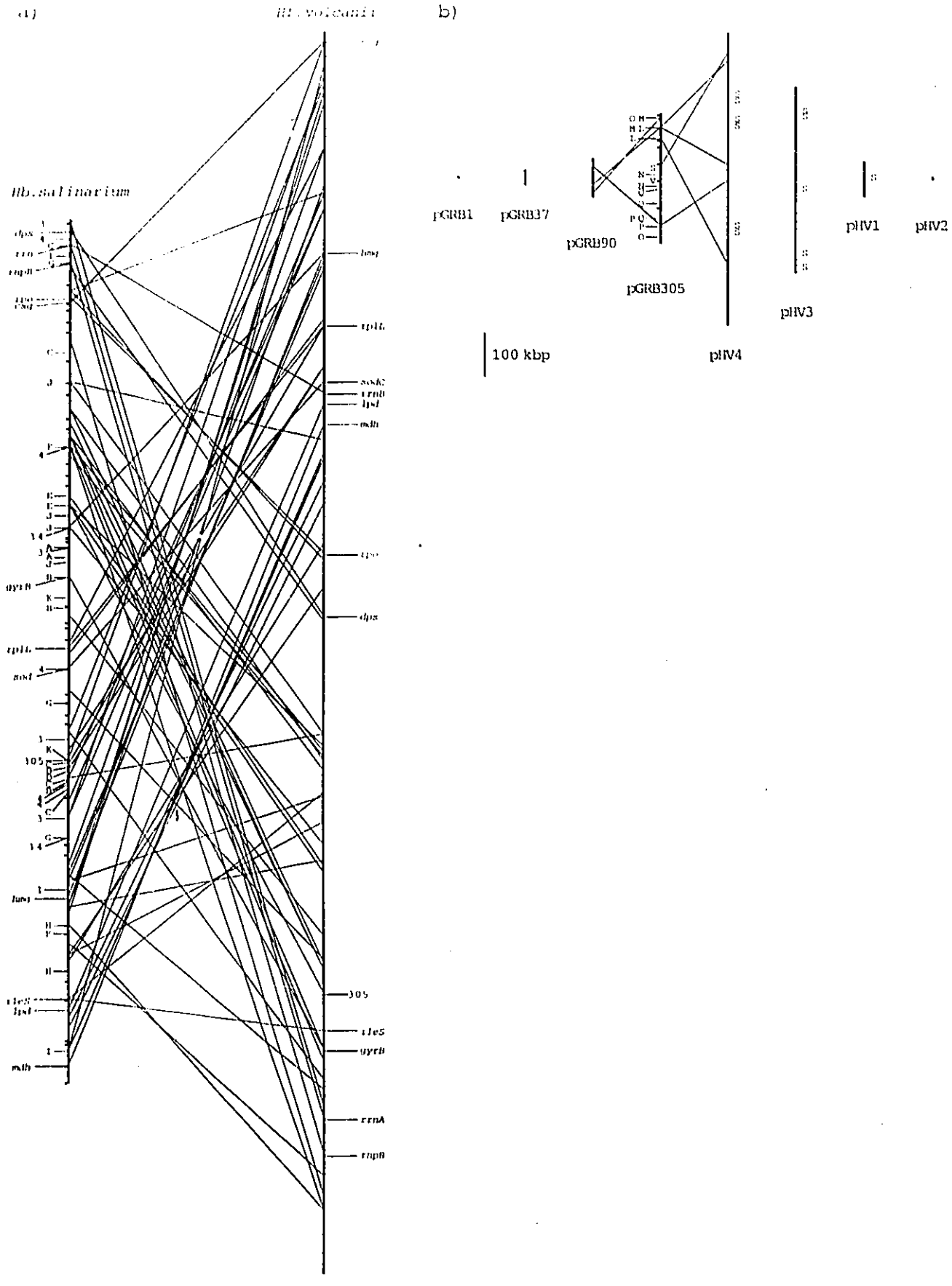
hybridized to a single chromosomal locus. We previously found that probes prepared from restriction fragments could be more sensitive than those prepared from whole cosmids (López-García et al., 1995); hence, the increased detection of repeated sequences is not surprising. A greater specific activity as well as a dissection of compound repeats may well be the explanation.

Extrapolating from the present study, we can estimate the repeat content within the *Hb. salinarium* GRB genome to be approximately 50 low-copy-number repeats in the 2.03-Mbp chromosome and another 50 in the 0.43 Mbp of plasmids. Besides intragenomic repeats, 21 chromosomal probes and 1 plasmid probe hybridized to between two and four loci in the *Hf. volcanii* DS2 genome (Fig. 6.3). Ten of these 22 probes gave multiple signals in both genomes. We do not know the nature of the repeated sequences; they may represent gene families as have been documented for the haloarchaea (Gerl and Sumper, 1988; Horne et al., 1988; Sanz et al., 1988; Joshi and Dennis, 1993; Antón et al., 1994), insertion sequences (especially in the plasmids [Charlebois and Doolittle, 1989] though of uncharacterized types [Ebert et al., 1986]), or other repeat sequence structures (Mojica et al., 1995).

Cross-hybridizations

Seventy of the 127 probes giving clear hybridization results linked homologous loci between the two genomes (Fig. 6.3). It is likely that the 57 probes uniquely hybridizing to *Hb. salinarium* simply diverged from *Hf. volcanii* in sequence beyond the threshold of detection, although major differences in genetic inventory between the two species cannot be ruled out. Of these 70 probes, 41 involved unique loci, whereas the other 29 included multiple signals (see above). Thirty-five of the 41 were interchromosomal, and 6

Fig. 6.3. Comparison of the *Hb. salinarium* and *Hf. volcanii* genomes. All replicons are drawn to the same scale and are circular molecules represented as vertical lines for clarity. The top of each replicon corresponds to map position 0 (Charlebois et al., 1991; St. Jean et al., 1994). Diagonal lines connect loci that cross-hybridize between replicons, while dots beside the vertical lines represent loci that do not cross-hybridize. All anonymous probes used were from *Hb. salinarium*. (a) The numbers 1, 3, and 4 are loci hybridizing to *Hf. volcanii* plasmids pHV1, pHV3, and pHV4, respectively. Two loci hybridizing to the *Hb. salinarium* plasmid pGRB305 are indicated by 305. The letters A through K represent low-copy-number repeated sequences on the *Hb. salinarium* chromosome. Twelve gene probes are indicated by name. (b) The letters S and V represent loci hybridizing to the chromosome of *Hb. salinarium* and *Hf. volcanii*, respectively. The letters L through Q indicate low-copy-number repeated sequences on pGRB305.


a)
Hb. salinarium
b)
pGRB1
pGRB37
pGRB90
pGRB305
pHV1
pHV2
pHV3
pHV4
100 kbp

connected the *Hb. salinarium* chromosome to *Hf. volcanii* plasmids (pHV4, three links; pHV3, two links; and pHV1, one link). Although the *Hf. volcanii* plasmids make up 29.5% of its genome, only 17.1% of uniquely cross-hybridizing (18.5% of all cross-hybridizing) *Hb. salinarium* chromosomal probes found a *Hf. volcanii* plasmid homolog. Most chromosome-to-plasmid, and all plasmid-to-plasmid, connections involved repeated sequences.

Analysis of comparison data

An overview of the chromosomal comparison led us to believe that extensive rearrangements had occurred but that some conservation remained in certain regions. The 35 probes producing one signal on each chromosome were used in analyses designed to quantify this impression.

The first analysis used the program DERANGE II (Kececioglu and Sankoff, 1995; Blanchette et al., 1996), which determines the minimum number of moves (using inversions, transpositions and inverted transpositions) needed to transform one set of ordered loci into another. DERANGE II measures similarity in terms of total cost (where each move adds a set value to the total, with transpositions and inverted transpositions costing more than inversions) and number of moves (the total number of all moves needed for the transformation). Although DERANGE II can deal with circular DNA molecules, it inputs the order of loci as a linear set. In case this has an effect on the outcome of the analysis, each of the 35 circular permutations of loci was run through DERANGE II using the set of parameters listed in Materials and Methods. These results were compared to 100 randomized permutations of 35 loci (Fig. 6.4). Figure 6.4 shows results for total cost,

which are similar to those given by number of moves (data not shown), indicating that there was no statistically significant ($p > 0.14$) difference between the randomized and experimental data.

In order to test how many changes were needed between the two chromosomes before DERANGE II would find no conservation, we had DERANGE II solve permutations of 35 loci with known numbers of inversions (Fig. 6.5). A linear relationship between total cost and the number of inversions is seen initially, eventually reaching a plateau. After roughly 40 steps, further inversions cease to have an impact on the ability of DERANGE II to solve the permutation. The total cost to solve random permutations of 35 loci closely follows this plateau, as does the total cost to solve the experimental data. Substituting number of moves for total cost gives identical results.

Another analysis of the experimental data involved determining whether conservation in the distances between pairs of loci within each chromosome exists. The position of a locus was estimated to be the center of the portion of the cosmid to which the probe hybridized. In a regression analysis (Fig. 6.6), no significant difference between the random and experimental data was found ($r^2 = 0.0148$).

Discussion

There is no conservation in the order of loci on the chromosome or in their pairwise distances in the genomes of *Halobacterium salinarium* GRB and *Haloferax volcanii* DS2. Although a lower density of homologous connections was observed between *Hb. salinarium* chromosomal DNA and *Hf. volcanii* megaplasmid DNA than between the two chromosomes was observed, more than one in six chromosomal probes found a plasmid homolog. Genomic rearrangements which encompass all major replicons, shuffling loci

Fig. 6.4. Comparison between the total cost of rearranging 35 loci from the chromosomes of *Hf. volcanii* and *Hb. salinarium* and random data. Numbers on the abscissa indicate the "look ahead" and "weight for transpositions and inverted transpositions" parameters of DERANGE II used in each run. Diamonds and vertical bars indicate average significance levels for the 35 circular permutations of the experimental data with associated standard deviations. Horizontal bars indicate the maximum and minimum values for each set of parameters used.

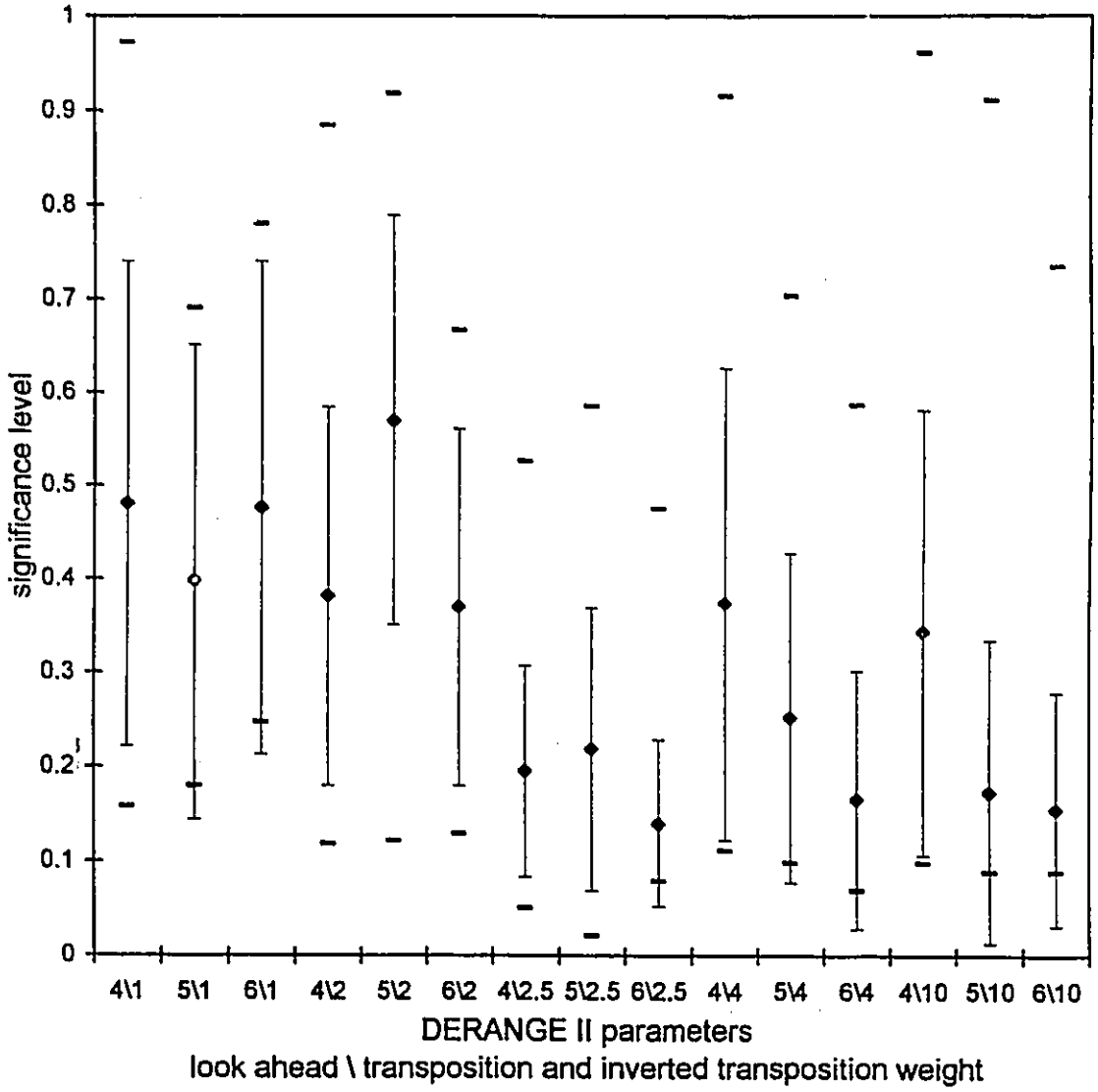
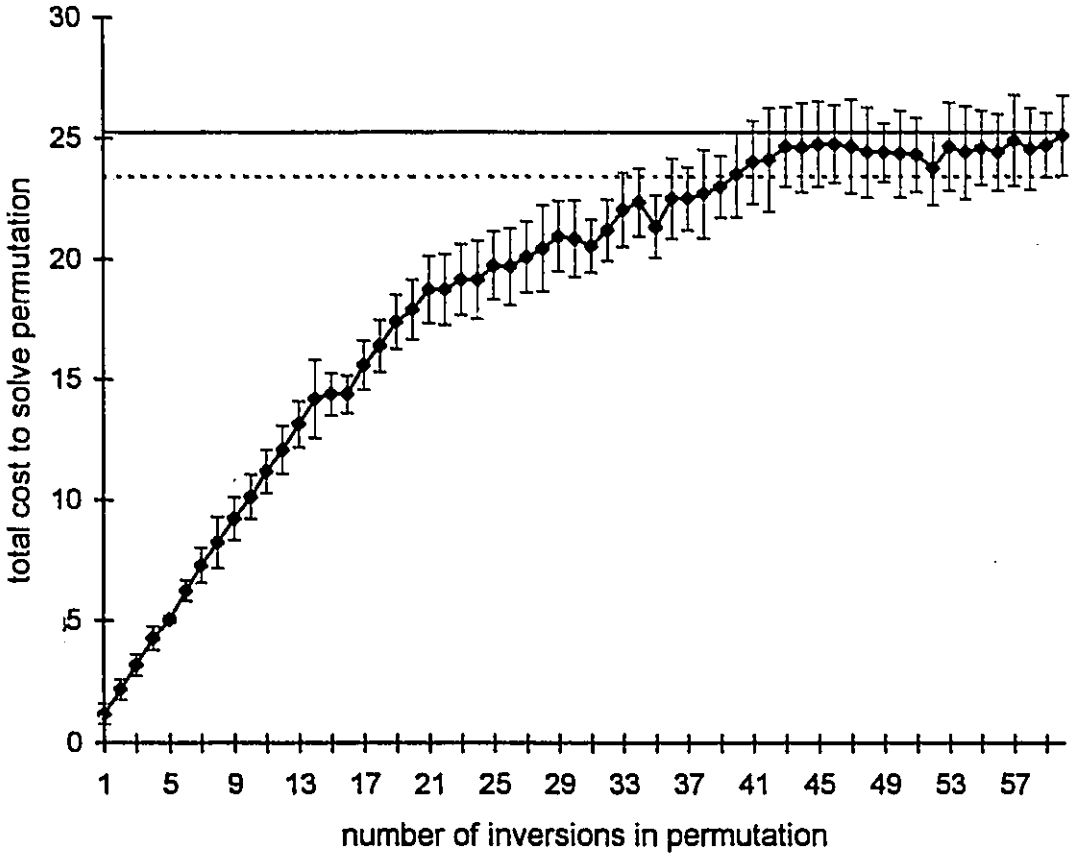

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
significance level
4\1
5\1
6\1
4\2
5\2
6\2
4\2.5
5\2.5
6\2.5
4\4
5\4
6\4
4\10
5\10
6\10
DERANGE II parameters
look ahead \ transposition and inverted transposition weight

Fig. 6.5. Performance of DERANGE II on permutations of 35 loci containing known numbers of randomly generated inversions based on total cost. Permutations contain between 1 and 60 inversions. The DERANGE II parameters used were a "look ahead" of 6 and "transposition and inverted transposition weights" of 2.5. The solid horizontal line indicates the average total cost to solve 100 random permutations of 35 loci. The dashed horizontal line indicates the average total cost to solve the 35 circular permutations of the experimental data. Vertical bars indicate standard deviations.

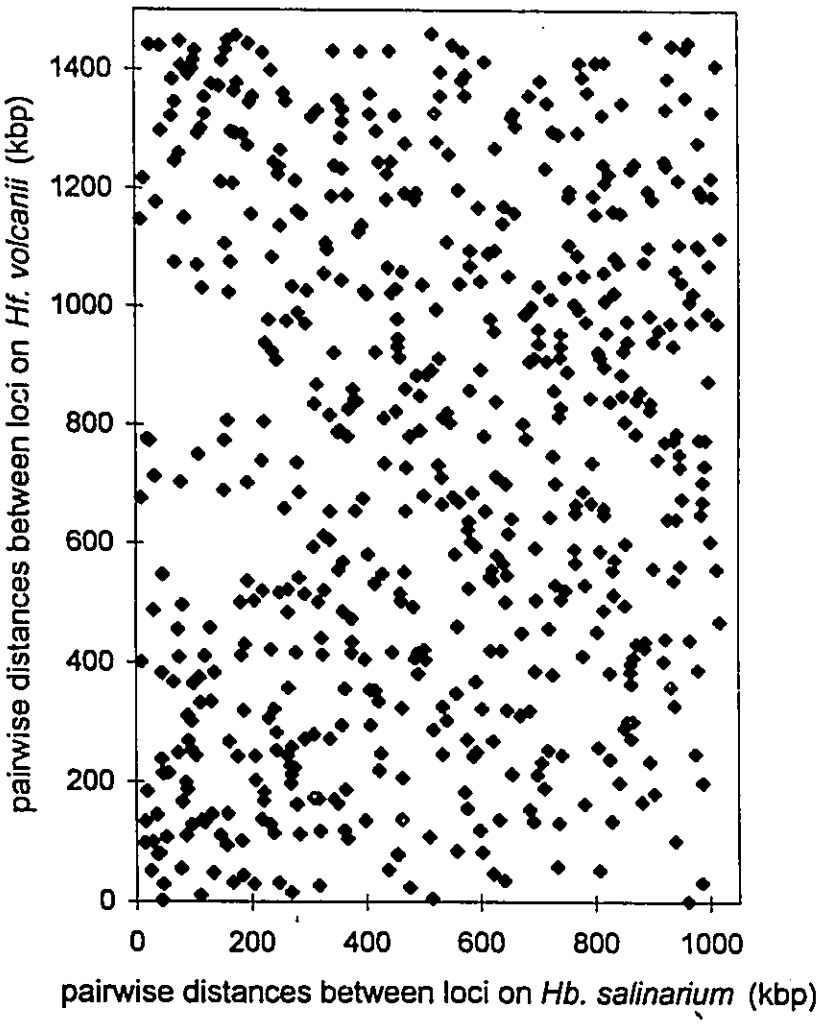

30
25
20
15
10
5
0
total cost to solve permutation
1
5
9
13
17
21
25
29
33
37
41
45
49
53
57
number of inversions in permutation

between them, have occurred. The nature of plasmid-encoded loci is still unknown, however, except that few identified genes map to them (Cohen et al., 1992). Probes hybridizing between plasmids and between chromosome and plasmids tended to confine themselves to FII and FI regions of plasmid DNA (Charlebois et al., 1991; St. Jean et al., 1994) respectively.

Half of all cross-hybridizing probes found repeated sequences, numbering between two and five copies. These can include gene families, insertion sequences and noncoding repeats. Insertion sequences concentrate in FII DNA (Pfeifer and Betlach, 1985; Cohen et al., 1992), and haloarchaeal plasmids are often FII (Charlebois and Doolittle, 1989; St. Jean et al., 1994), but *Hb. salinarium* GRB is a genetically stable isolate (Soppa and Oesterhelt, 1989) without any of the known '*Hb. halobium*'-type insertion sequences (Ebert et al., 1986). Still, it is likely that the FII repeats are insertion sequences, albeit of new types, since haloarchaeal repeats are often insertion sequences (Hofman et al., 1986) and insertion sequences cluster in FII (Pfeifer and Blaseio, 1990; Cohen et al., 1992). We have observed the rare sectored colony of strain GRB, indicative of their presence and activity. We postulate that *Hb. salinarium* ancestrally maintains an adapted set of moderate insertion sequences but that the *Halobacterium halobium*-type strains recently inherited novel virulent types now wreaking havoc in their genomes.

Repeated sequences (especially insertion sequences) are often more conserved between haloarchaeal genomes than are unique sequences (Sapienza and Doolittle, 1982; Pfeifer and Blaseio, 1990). Our hybridization-based survey naturally demanded detectable, hence conserved, signals and may have overemphasized mobile genetic elements in the comparison and alignment. If strain GRB has been insulated from recent horizontal

Fig. 6.6. Scatter plot of the distances between all pairs of 35 loci found on the *Hb. salinarium* chromosome and homologous pairs on the *Hf. volcanii* chromosome. An r^2 value of 0.0148 was obtained by using 595 datum points.


pairwise distances between loci on *Hf. volcanii* (kbp)
1400
1200
1000
800
600
400
200
0
0
200
400
600
800
1000
pairwise distances between loci on *Hb. salinarium* (kbp)

acquisition of novel insertion sequences, as evidenced by its lack of the most potent types, the multicopy signals are more likely members of older gene families.

Fifty-seven of 127 *Hb. salinarium* probes did not hybridize to the larger *Hf. volcanii* genome. The lack of success of this 45% of the anonymous probes can result from major differences in genetic inventory between the species or from sequence drift beyond the threshold of detection by hybridization. This has consequences relevant to the interpretation of comparative genomic analyses such as the one presented here. Suppose, for instance, that a pair of duplicated genes map to parallel loci in two genomes. A probe including one of these genes would find both the paralogous and orthologous copies, revealing the duplicity. The hybridization may not give an indication of which copy is the ortholog, resulting in an inability to use that probe's results in a chromosomal alignment. If, however, sequence drift occurs such that only the intergenomic paralogous pair is detected by hybridization (unlikely but possible if the copies are truly redundant and not under divergent selective pressures), a false genomic rearrangement is observed. Especially misleading and more common would be the cases where alternate orthologs are deleted, leaving only a paralogous pair. Forty-five percent of our probes did not cross hybridize, and 50% of those that did found multicopy loci.

Probes not cross-hybridizing were evenly distributed about the *Hb. salinarium* chromosome, and probes which did cross-hybridize found signals evenly distributed about the 44% larger *Hf. volcanii* chromosome. Either there have been multiple small insertions or deletions in one genome relative to the other, or larger blocks have been scrambled by recombination. The latter scenario requires a more ancient origin of the chromosomal size difference. Though the *Haloferax mediterranei* chromosome matches that of *Hf. volcanii* in size (López-García et al., 1992), implying constancy since their divergence, a

comparison within strains of *Hb. salinarium* (Hackett et al., 1994) revealed recent small insertion-deletion events. Thus, neither mechanism can be excluded at this time.

There are now three haloarchaeal genomic comparisons available, at three phylogenetic depths: intraspecific, interspecific, and intergeneric. The first two demonstrated genomic stability, which is remarkable given the potential for disruption in haloarchaeal genomes. The present study found no amount of conservation in the order of loci. It is likely and expected that some assortments of genes have been maintained, operons especially. Macrorestriction maps were sufficient to align the genomes of *Hf. volcanii* and *Hf. mediterranei* (López-García et al., 1995), and they were ideal in the alignment of *Hb. salinarium* genomes (Hackett et al., 1994). Here, we leap from an estimated interspecies divergence of 80 million years (López-García et al., 1995) to very roughly 600 million years, on the basis of a 16S rRNA divergence of 12% (Mankin et al., 1985) and a clock rate of 1% per 50 million years (Ochman and Wilson, 1987). True rearrangement as well as complications involving gene families and sequence divergence (described above) effectively prevent chromosomal alignment, even with the use of high resolution contig maps and a program like DERANGE II. It is important to remember that the *Halobacterium* and *Haloferax* genomes are derived from a common ancestor. Although extensive rearrangement, the full elucidation of which will require sequence-level comparison, was observed between them here, this rearrangement is a process driven by a balance of forces. What does it mean that these genomes have shuffled the order of their genes at a scale finer than our mean resolution of 15 kbp? Or indeed have they done so, given that we could have been misled by the dynamics of gene-family sequence divergence?

We entertain two models of genomic restructuring: gradual and punctuated. The gradualist would see the process of rearrangement as a function of time. A more saltatory mode of genome evolution would see the genome abandon its map for another as a consequence of wholesale selection for an altered pattern of gene expression (Charlebois and St. Jean, 1995). Our present data, which are limited to three comparisons, are consistent with both models. It is useful to address this issue of tempo, since rearrangement can effectively block recombination and thus contributes to speciation. Biodiversity, even among microbes, is a necessary component of survival and provides the foundation from which adaptations and innovations can arise.

If the haloarchaeal genome was designed at the base of their evolutionary radiation into (currently) eight genera through massive adaptive reorganization of an ancestral genome, there should essentially be eight maps. On the other hand, if rapid genomic reengineering did not coincide with the founding of the haloarchaeal lineage but rather has evolved to its present state through occasional rearrangements, there will be many maps which together may help us to construct detailed phylogenies and to recount the history of genomic events. The punctuated model would likely provide few clues in this direction, since the events would be condensed into a short span of time, and would effectively block a view beyond the origin to the earlier parent.

The lesson learned in this experimental study has been that, at least in the haloarchaea, genomic change is complex. Therefore, further studies should focus on fine detail in one or a few specific regions of the genome, with the assumption that similar things are happening elsewhere in the genome. Sequence-level comparisons can be used to measure the rate of genomic change relative to sequence divergence in many members of a

lineage, although a reference sequence of an entire genome would be useful in order to solve the problem of paralogy.

Data need not be so extensive or so expensive as those obtained by genomic sequencing in order to answer many pertinent questions. The present study does, however, illustrate the need for tests to measure objectively the degree of similarity when genomic comparisons are performed. To date, most comparisons have involved low-resolution PFGE maps of closely related organisms, usually of the same genus or species. The small number of common loci used in these comparisons often makes the degree of similarity present easy to determine simply by looking at the data. Examples of comparisons where many loci were used, such as *E. coli* and *S. typhimurium* (Riley and Krawiec, 1987) and *Rhodobacter capsulatus* (Fonstein et al., 1995), showed clear-cut results. Such easily interpreted results are unlikely to make up the majority of future comparisons.

The need for many cross-hybridizing signals in a comparison, at the highest possible resolution, is also clear. This becomes more important as the phylogenetic distance between the compared genomes increases, as shown by the present study, sometimes even necessitating large-scale sequencing. The existence of whole genome sequences of both prokaryotes and eukaryotes will encourage comparisons between distantly related genomes. These comparisons will have the most to gain from tests such as those performed in this study as well as future generations of increasingly sophisticated analytical tools. The move towards more objective measures and methodologies will be a necessary step if comparative genomics studies hope to answer any but the most vague questions about genomic level change and evolution.

Acknowledgements

We gratefully thank D. Sankoff and M. Blanchette for providing the program DERANGE II. Assistance given by E. Weiher with the statistical analyses is also greatly appreciated.

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. R.L.C. is an Associate of the Canadian Institute for Advanced Research.

Update to Chapter 6

Many of the conclusions and inferences drawn in this chapter rely on the results of the program DERANGE II. Because of this, a thorough understanding of this program's strengths and limitations is necessary to allow for the appropriate interpretation of these results. Some of these issues are explored here while additional information on the workings of DERANGE II can be found in the introduction to chapter 7 and in Blanchette et al. (1996).

Analysis with DERANGE II indicated that this program could detect no conservation in the order of homologous loci between *Hf. volcanii* and *Hb. salinarum*. The question remained, however, as to the nature of the results obtained if two more closely related chromosomes were compared. This question was addressed to a degree by Fig. 6.5 where simulated DNA segments containing known numbers of rearrangements were analysed. Additionally, an identical analysis as described in this chapter was performed on the data from the *Hf. volcanii-Hf. mediterranei* comparison (López-Garcia et al., 1995 [chapter 5]). In this case, sixteen chromosomal loci were used as it was necessary to collapse many of the original 35 loci that hybridized to common regions of

the *Hf. mediterranei* chromosome. Comparing this data set with randomly generated data, DERANGE II identified the two chromosomes as consistently showing conservation. A significance level of 0.01 was obtained with all parameters used. When the transposition and inverted transposition weights were set to 2, DERANGE II went so far as to use the same two inversions and single transposition in its solution that were identified in chapter 5 (results not shown). These two tests indicate that DERANGE II can indeed distinguish conserved DNA segments from random.

CHAPTER 7

Compare-A-Genome (COMPAGEN): a database program for the management and analysis of comparative genomic data

Introduction

'The more you get, the more you want', is a sentiment that has been voiced many times. It is particularly true of genomic studies. Computational methods and tools for the management of the rapidly accumulating amounts of map and sequence data are no longer merely desirable, but absolutely necessary. As well, new kinds of metrics and analyses designed to take advantage of this growing amount of information need to be implemented in a way that allows them to be used in a reasonable amount of time. 'Reasonable time' usually means recourse to a computer program when dealing with genomics projects. Accessibility, ease of use, and integration with other tools are also factors in the 'reasonable time' equation. Unless a tool can be obtained, learned and used in concert with existing tools quickly, it is in danger of being ignored, especially in a rapidly evolving field such as genomics.

Such considerations inspired the development of COMPAGEN. COMPAGEN was written in order to take advantage of another program, DERANGE II (Blanchette et al., 1996). DERANGE II provides a metric of the differences between pairs of DNA segments by tracking the number of changes needed to rearrange homologous loci on the two segments until the two are colinear. The length of the DNA segment is not a factor in the calculations—only the number of homologous loci—so whole chromosomes may be analysed in this manner.

DERANGE II operates by using three type of rearrangements—inversions, transpositions, and inverted transpositions—to reduce the number of breakpoints in one DNA segment relative to another. A breakpoint is said to occur between two loci whenever those two loci are adjacent in one of the segments but are not in the other. Each inversion introduced into the solution can reduce the number of breakpoints by a maximum of two, while transpositions and inverted transpositions can eliminate as many as three breakpoints. DERANGE II does not perform an exhaustive search of all possible series of rearrangements as this would be computationally impractical. Instead, only rearrangements that reduce the number of breakpoints at each step are considered. This technique is combined with a limited look ahead which is set by the user. DERANGE II uses the look ahead to test different combinations of rearrangements—the number of rearrangements in each trial equaling the look ahead value—with the combination eliminating the largest number of breakpoints being implemented. One of the more important features of DERANGE II is the ability to specify weight values to the three types of rearrangements implemented. The weight value represents the cost to DERANGE II to use a type of rearrangement meaning that as the weight value increases, DERANGE II is more likely to find alternative solutions that do not use, or use fewer, of the weighted rearrangement. The most obvious application of this feature is to specify differences between the likelihood of inversions as opposed to transpositions (both types) to occur in a genome. The two values produced by DERANGE II to measure the dissimilarity between two DNA segments are 'number of moves' and 'total cost'. Number of moves is simply the sum of all the rearrangements introduced by DERANGE II in the course of solving a permutation. Total cost is the sum of the costs for each rearrangement introduced by DERANGE II and as such, is influenced by the weight value assigned to

each rearrangement type. The program tries to reduce total cost in preference to number of moves so depending on the weight values used, a solution with more rearrangements (a higher value for number of moves) might be accepted if the total cost is lower than an alternative solution. DERANGE II does not presume to reconstruct the evolutionary history of the DNA segments it examines. Rather, it provides a theoretical mathematical minimum distance separating two DNA segments. DERANGE II is a command line program which can make analysis of multiple data sets rather cumbersome. It also outputs its results in a complicated text file which is not conducive to use by other software packages such as statistical or presentation programs. This is where COMPAGEN steps in.

COMPAGEN provides a graphical user interface to DERANGE II where one can specify any of the parameters used by DERANGE II. Both the input file used by DERANGE II and the output file generated each time DERANGE II is run can be viewed from within COMPAGEN. DERANGE II's output file can also be parsed into a set of database table files. This provides much easier access to the results by other programs and allows summaries of the results to be generated quickly and easily. Manipulations of the data, such as deleting data sets run with a specific set of parameters or even deleting only part of a data set are greatly facilitated.

COMPAGEN: What do I need?

COMPAGEN was written using the Borland® Paradox® database program version 7, a 32-bit program that runs natively under Microsoft Windows® 95 and Windows NT™. As such, the end user needs a copy of Paradox to use COMPAGEN. DERANGE II must also be included for COMPAGEN to be useful. DERANGE II is provided as a separate

executable program allowing newer versions to be used as they become available. However, for COMPAGEN to work properly, the formats of both the input parameters passed to DERANGE II and the output file produced by it must be respected. Any change in either of these will likely cause errors to occur when running COMPAGEN.

COMPAGEN: How do I use it?

As stated in the introduction, COMPAGEN's purpose is to make DERANGE II easier to use and to store DERANGE II's output in an easily maintained and retrievable form. COMPAGEN does this by providing a Microsoft® Windows graphical user interface for setting up and running DERANGE II, by parsing DERANGE II's output into a series of database table files and by generating tabular and graphical summaries of this output which can be printed as a hard copy.

Fig. 7.1 shows a screen shot of COMPAGEN's DERANGE II setup screen. From here, all parameters used by DERANGE II can be specified, input and output files can be named and viewed using a text editor, DERANGE II can be run and its output parsed into table files, and user preferences set. Input files of random permutations for comparison with experimental data can also be generated from this screen. The importance of generating random permutations will be discussed in the next section. After running DERANGE II and parsing the output text file into table files, one can select the saved data set for analysis. More than one data set can be analysed at a time.

Analysing a data set will generate summary information for that data set. Figs. 7.2 and 7.3 show screen shots of the summary information COMPAGEN provides. Information can be viewed for only one data set at a time as shown in Fig. 7.2 but the graphical display allows multiple data sets to be compared directly (Fig. 7.3). With this

summary information, one can compare the number of changes introduced by DERANGE II, the number of each type of change (inversion, transposition, and inverted transposition), the distributions of these types of changes (whether inversions tend to occur before transpositions for example), and the number of loci each change encompasses.

One can access the data stored in COMPAGEN's table files in two ways. COMPAGEN allows the export of information belonging to selected data sets in a variety of formats. Exported formats include fixed length text, delimited text, Excel, Quattro Pro and Lotus 1-2-3. Paradox is also compliant with the Open DataBase Connectivity (ODBC) standard. Thus any program that is also compliant with this standard can access COMPAGEN's table files directly without the need to export the data. This adds another potential advantage in that ODBC-compliant database querying tools can be used to retrieve COMPAGEN's stored data allowing for much more control over the retrieval process.

The need for easy access to COMPAGEN's stored data is important because COMPAGEN itself has very limited analytical and graphing capabilities. COMPAGEN is useful for gaining a first impression of the output produced by DERANGE II but for more rigorous analysis, another software package would almost certainly be needed. For example, in chapter 6, data stored in COMPAGEN table files was accessed directly using Microsoft® Excel™. Excel's graphing capability was then used to produce Fig. 6.4.

Fig. 7.4 shows COMPAGEN's data management screen. It is here that one can quickly find out how many data sets are stored, view any of the stored data sets, rename data sets, and delete selected data sets or parts of data sets.

Fig. 7.1. The DERANGE II setup screen of COMPAGEN. From here parameters used by DERANGE II can be set, input and output files of DERANGE II can be viewed, random data can be generated, data can be selected for analysis or exported in a variety of formats, and analyzed data can be viewed.

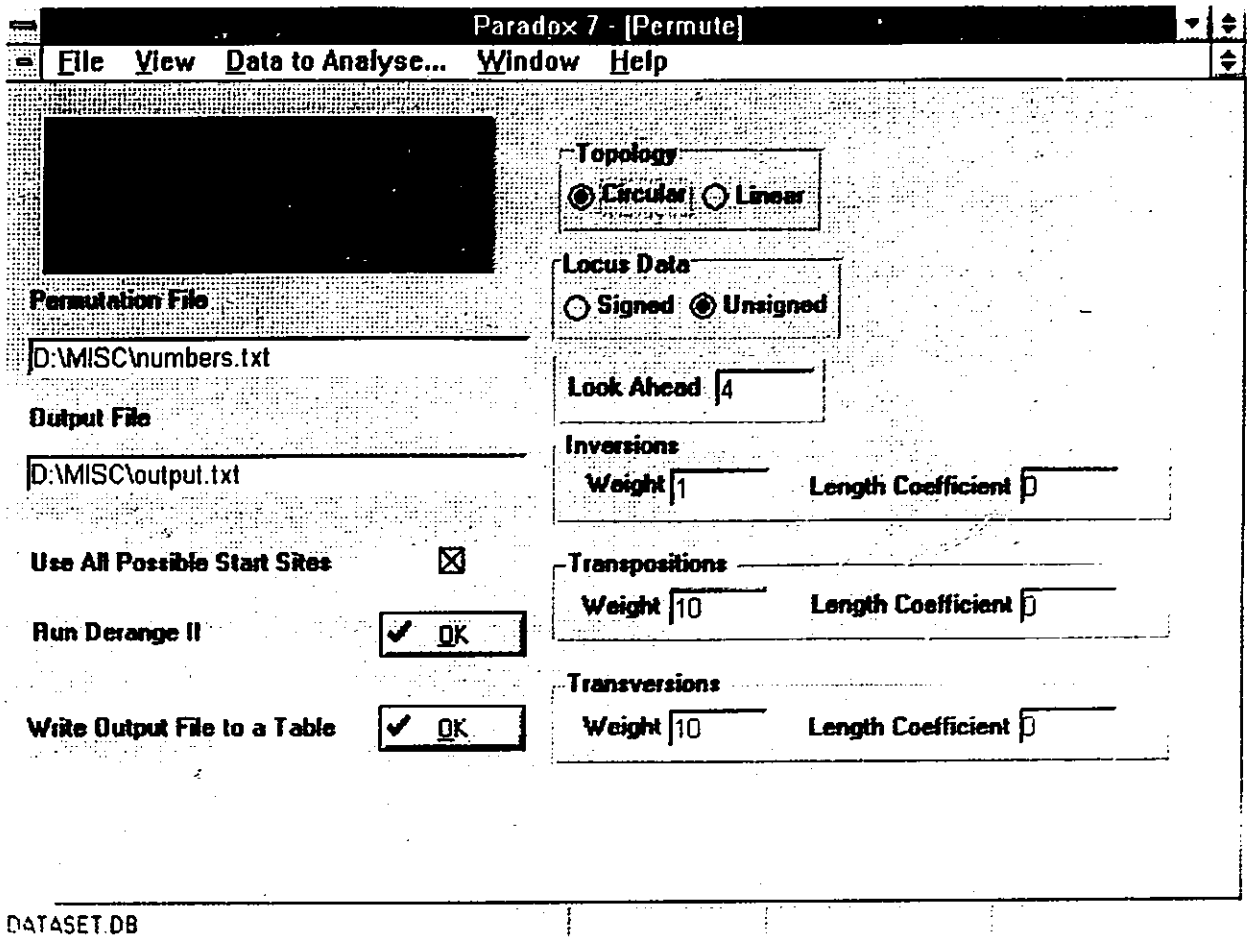

Paradox 7 - [Permute]
File View Data to Analyse... Window Help
Permutation File
D:\MISC\numbers.txt
Output File
D:\MISC\output.txt
Use All Possible Start Sites
Run Derange II
OK
Write Output File to a Table
OK
Topology
Circular
Linear
Locus Data
Signed
Unsigned
Look Ahead 4
Inversions
Weight 1
Length Coefficient 0
Transpositions
Weight 10
Length Coefficient 0
Transversions
Weight 10
Length Coefficient 0
DATASET.DB

Fig. 7.2. Data analysis performed by COMPAGEN displayed in tabular form. The DERANGE II parameters used to generate the data are also shown on this screen. Results of each analyzed data set can be viewed sequentially.

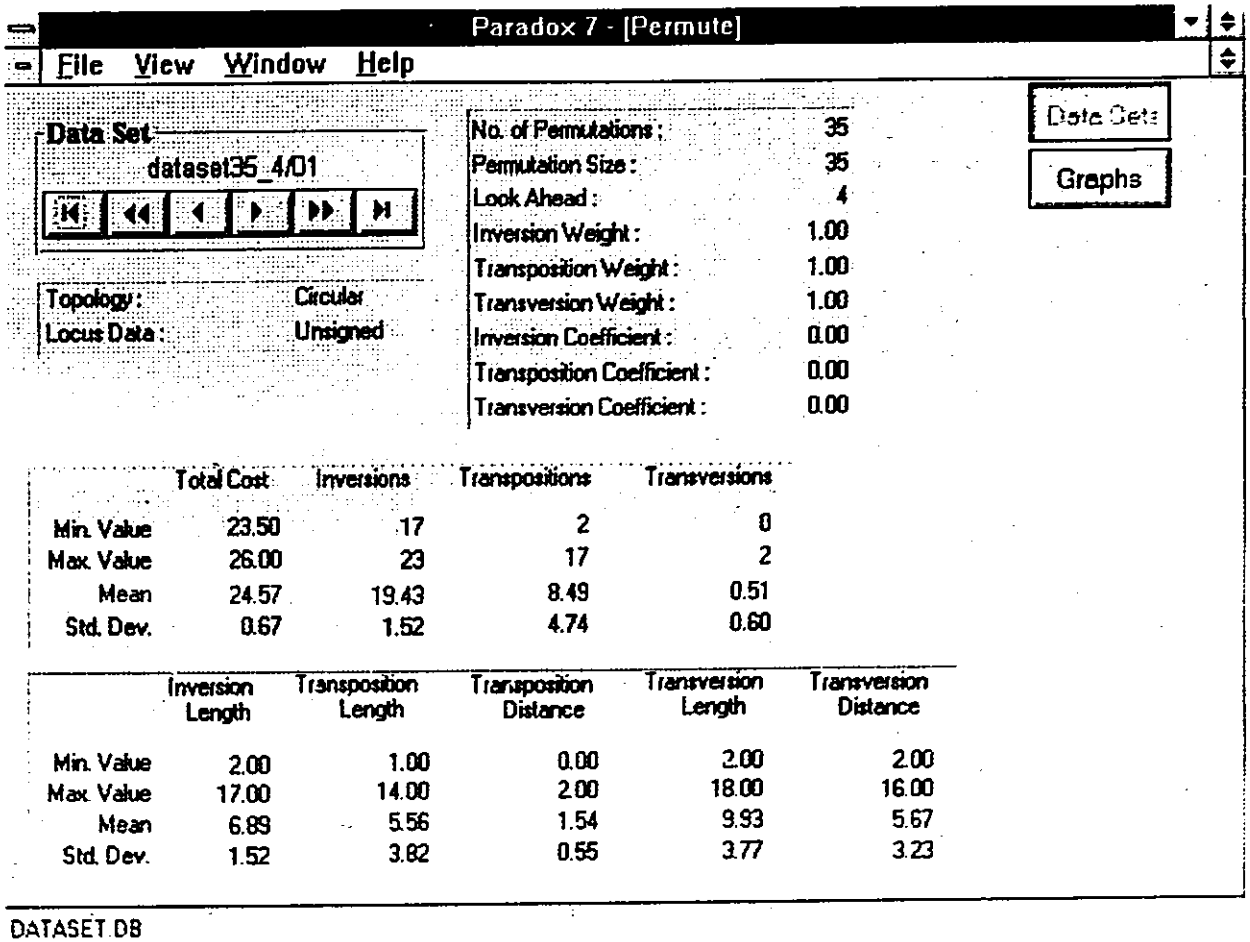
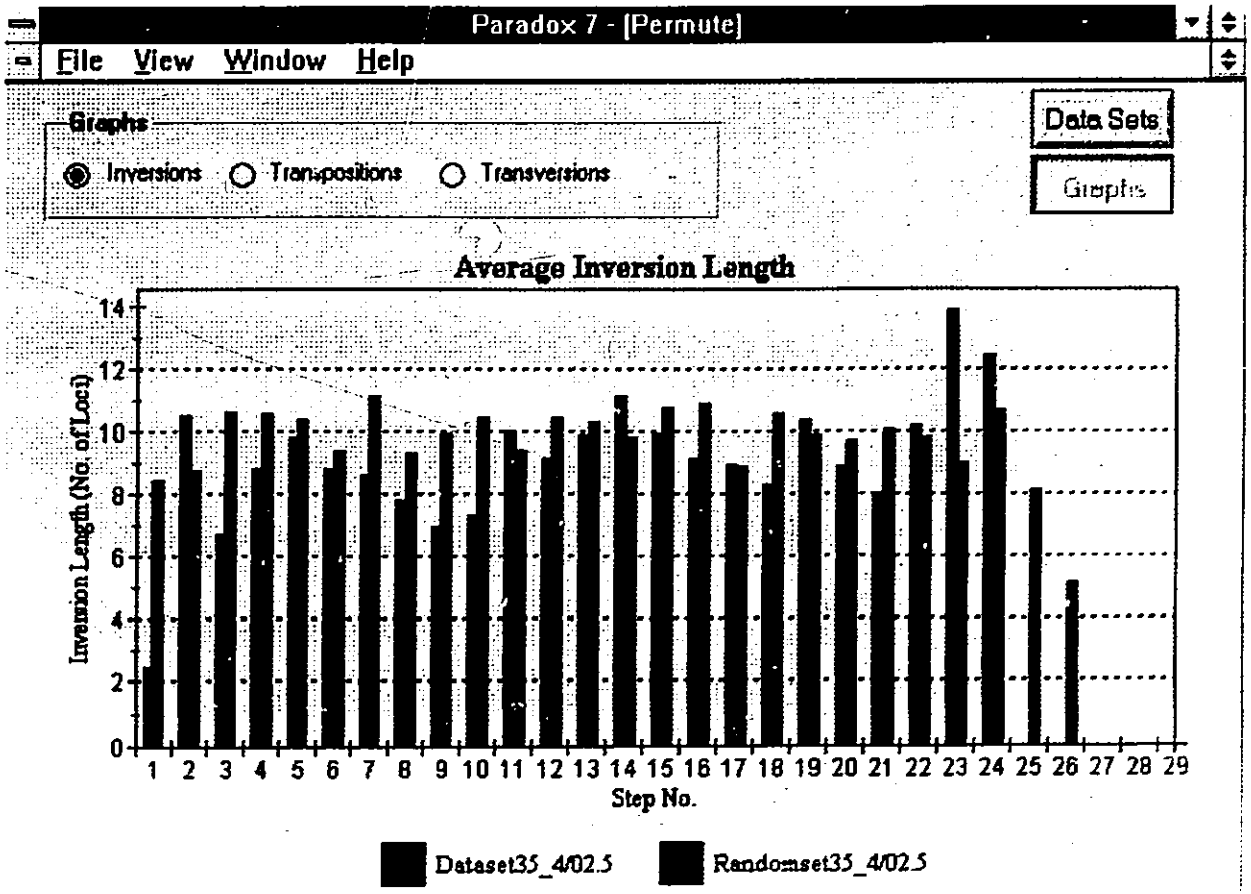
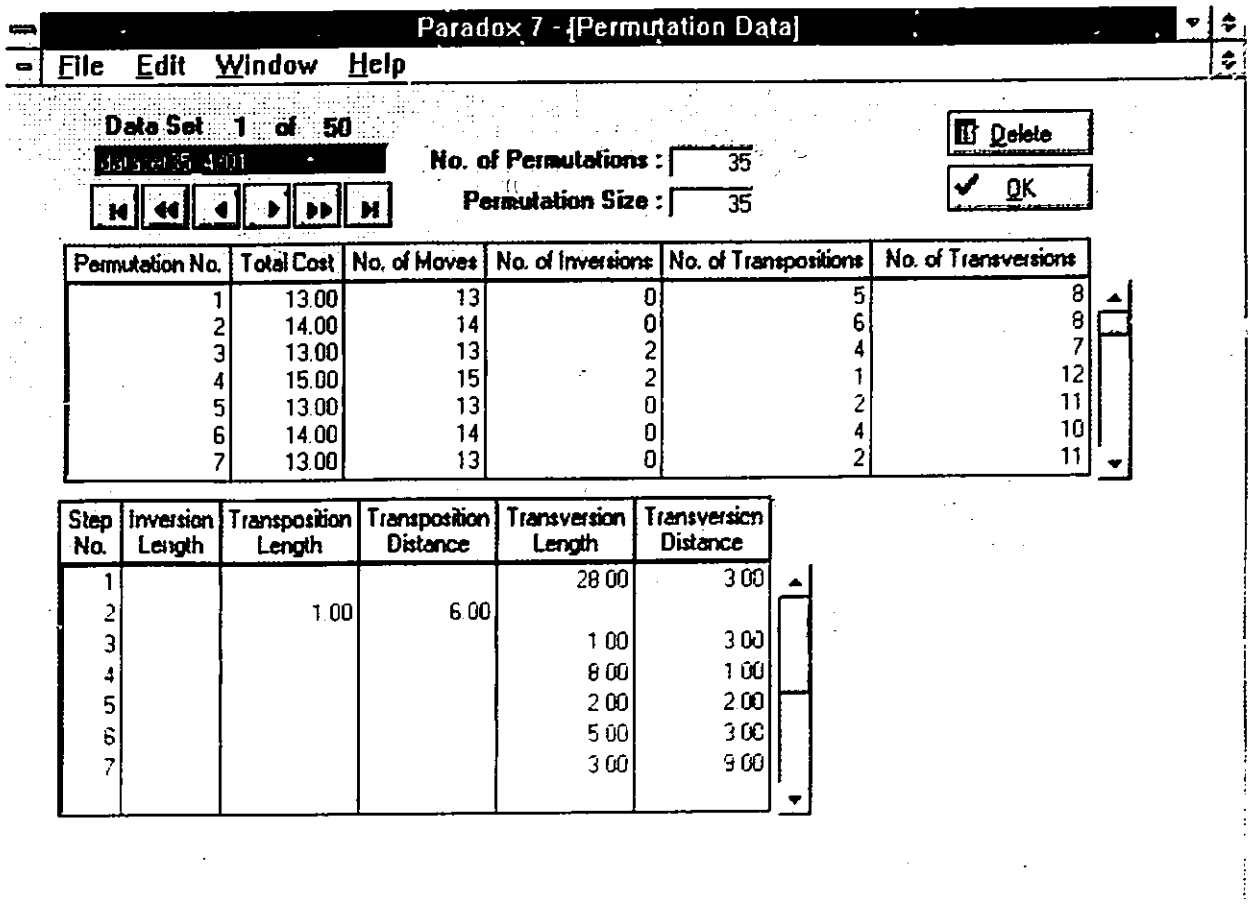

Paradox 7 - [Permute]
File View Window Help
Data Set
dataset35_4/01
Topology: Circular
Locus Data: Unsigned
No. of Permutations: 35
Permutation Size: 35
Look Ahead: 4
Inversion Weight: 1.00
Transposition Weight: 1.00
Transversion Weight: 1.00
Inversion Coefficient: 0.00
Transposition Coefficient: 0.00
Transversion Coefficient: 0.00
Data Sets
Graphs
Total Cost Inversions Transpositions Transversions
Min. Value 23.50 17 2 0
Max. Value 26.00 23 17 2
Mean 24.57 19.43 8.49 0.51
Std. Dev. 0.67 1.52 4.74 0.60
Inversion Length Transposition Length Transposition Distance Transversion Length Transversion Distance
Min. Value 2.00 1.00 0.00 2.00 2.00
Max. Value 17.00 14.00 2.00 18.00 16.00
Mean 6.89 5.56 1.54 9.93 5.67
Std. Dev. 1.52 3.82 0.55 3.77 3.23
DATASET.DB

Fig. 7.3. Graphical display of the data shown in Fig. 7.2. Multiple data sets can be viewed simultaneously for each of the three types of changes introduced by DERANGE II; inversions, transpositions, and inverted transpositions.


Paradox 7 - [Permute]
File View Window Help
Graphs
Inversions
Transpositions
Transversions
Data Sets
Graphs
Average Inversion Length
Inversion Length (No. of Loci)
0 2 4 6 8 10 12 14
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29
Step No.
Dataset35_4/02.5
Randomset35_4/02.5

DATASET DB

Fig. 7.4. The data management screen of COMPAGEN. All the data sets stored by COMPAGEN can be viewed from this screen. Data sets can also be renamed and deleted.


Paradox 7 - [Permutation Data]
File Edit Window Help
Data Set 1 of 50
No. of Permutations : 35
Permutation Size : 35
Delete
OK
Permutation No. | Total Cost | No. of Moves | No. of Inversions | No. of Transpositions | No. of Transversions
1 | 13.00 | 13 | 0 | 5 | 8
2 | 14.00 | 14 | 0 | 6 | 8
3 | 13.00 | 13 | 2 | 4 | 7
4 | 15.00 | 15 | 2 | 1 | 12
5 | 13.00 | 13 | 0 | 2 | 11
6 | 14.00 | 14 | 0 | 4 | 10
7 | 13.00 | 13 | 0 | 2 | 11
Step No. | Inversion Length | Transposition Length | Transposition Distance | Transversion Length | Transversion Distance
1 | | | | 28 00 | 3 00
2 | | 1 00 | 6 00 | |
3 | | | | 1 00 | 3 00
4 | | | | 8 00 | 1 00
5 | | | | 2 00 | 2 00
6 | | | | 5 00 | 3 00
7 | | | | 3 00 | 9 00
1 of 50 [DATASET.DB]

Random numbers and COMPAGEN

DERANGE II performs binary comparisons of DNA segments. In chapter 6 these DNA segments were the chromosomes of two haloarchaea. To see if any conservation existed in the order of homologous loci between the two chromosomes, it was necessary to find out how many changes DERANGE II would introduce into random data with the same number of loci using the same parameters. To accomplish this, 100 random permutations of numbers simulating a DNA segment were generated and run through DERANGE II. The results were stored in COMPAGEN and compared to the results given by the data from the two chromosomes. No conservation was found because the experimental data did not look significantly different from the random data. Thus it can be seen that an assessment of the randomness of the generated numbers is vitally important.

COMPAGEN uses the pseudo-random number generator defined in the file random.cpp which is distributed as part of the Standard Template Library (STL) by Hewlett-Packard. This generator has been modified in that the constant seed value defined in the original has been replaced by a seed derived from the elapsed time in milliseconds since January 1, 1970. This provides a unique seed each time a random data set is generated.

It can be assumed that the numbers generated are not truly random as this is impossible to achieve using an algorithm, even one that incorporates unique seed values (Bratley et al., 1983). However, strictly random numbers are often not necessary in practical terms and it is important only that the numbers be random enough. For example, if a generator produces numbers that correlate with previously generated numbers after 3 million iterations, this would not impact on an application that requires the generation of

only 10,000 random numbers. With this in mind, the generator used by COMPAGEN was tested with 1000 data sets of 50 numbers each, far more data (50,000 numbers) than was used in the analysis described in chapter 6 (3,500 numbers). If the numbers are randomly distributed, then each number (from 1 to 50) is expected to appear in each position of the series an average of 20 times given a series of 50 numbers and 1000 replicates. The actual number of times each number appeared at each position was calculated and a χ^2 analysis performed using 20 as the expected frequency of appearance in all cases. This analysis resulted in a value of 0.279, three orders of magnitude below the critical value of χ^2 at a confidence interval of 0.001, indicating that the numbers are randomly distributed.

Random.cpp is implemented in a dynamic link library called compagen.dll written in C++. Only one function is exported from this DLL, randomize(), allowing it to be replaced so that another random number generator can be used if desired. For anyone wishing to do this, the required documentation is provided in Appendix B.

Chapter 8

General Discussion

The Current State of Haloarchaeal Genomics

The haloarchaea are remarkably well represented when it comes to genomic map and comparative map data considering the time frame in which the work was done: a mere eight years. High resolution contig maps have been constructed of two genomes—*Hf. volcanii* DS2 (Charlebois et al., 1991; Cohen et al., 1992) and *Hb. salinarum* GRB (St. Jean et al., 1994 [chapter 3])—with PFGE maps available for the chromosomes of two additional strains of *Hb. salinarum* (Hackett et al., 1994) and five strains of *Hf. mediterranei* (López-García et al., 1992; López-García et al., 1993). Comparative pulsed-field gel mapping has also been conducted on the largest plasmid, pHM300, of five strains of *Hf. mediterranei* (Antón et al., 1995). The chromosomal comparisons so far conducted span three different time frames. In the case of the three strains of *Hb. salinarum* phylogenetic distance has been judged by numerical taxonomy (Ebert et al., 1984; Ebert et al., 1986) and the conservation evident in the chromosomal restriction maps (Hackett et al., 1994). Percent similarity in 16S rDNA can be applied to the other two comparisons: 98.4% between *Hf. volcanii* and *Hf. mediterranei* (Kamekura and Seno, 1992), and 88% between *Hf. volcanii* and *Hb. salinarum* (Mankin et al., 1985).

These comparisons indicate that the insertion sequences inhabiting *Hb. salinarum* NRC-1 and S9 and *Hf. volcanii* DS2 have been largely ineffective in altering the chromosomal maps of these genomes. Plasmid DNA seems to be under less stringent constraints, displaying sometimes high degrees of polymorphism between isolates of the

same species (Antón et al., 1995) or the same strain (Pfeifer et al., 1989). The chromosomal comparisons between members of the same genus showed varying degrees of conservation between their maps concomitant with their presumed phylogenetic relationships while the intergeneric comparison showed no detectable conservation, also in keeping with their phylogenetic placement. Nothing untoward seems to come from this pattern of conservation and divergence which follows the phylogeny of the organisms involved without exception. However, when one considers the branching pattern of the haloarchaea as a whole and adds the local context model and its consequences, an interesting scenario suggests itself.

The local context model predicts that an organism's genome may suffer an increased rate of rearrangements in response to a change in conditions that cause its patterns of gene expression to become suboptimal. These rearrangements could lead to new genomic map orders being fixed in the population and since rearrangements are an effective way to prevent homologous recombination, this may represent an adaptive radiation (Charlebois and St. Jean, 1995 [chapter 4]). Phylogenies of the haloarchaea based on molecular data often show short branch lengths between the various genera suggesting that they diverged from one another in a relatively short time span. This is seen in the tree of McGenity and Grant (1995) with its short branch lengths and low bootstrap values and Kamekura and Dyall-Smith (1995) who opted to collapse many of the branches between genera because of the uncertainty in their ordering. Could it be that the genomic comparisons are actually showing the results of an adaptive radiation at the base of the haloarchaea with different genera finding their own chromosomal maps which have been preserved ever since? The transition from methanogen to extreme halophile certainly involved a drastic alteration in gene expression; the haloarchaea are aerobic while all methanogens are strict anaerobes

due to their unique energy metabolism (Balch et al., 1979). An influx of exogenous DNA may also have been involved as there is some evidence for a cyanobacterial origin for certain genes (Hase et al., 1977; Walker et al., 1984; Pfeifer, et al., 1993). If this scenario is true, one would predict that each phylogenetically distinct group radiating from the base of the haloarchaeal tree should have a distinct chromosomal map order. Given the degree of conservation found between *Hf. volcanii* and *Hf. mediterranei* (López-Garcia et al., 1995 [chapter 5]), it may be sufficient to use genetic maps based on pulsed-field gel physical maps to test this hypothesis. Ideally, at least two distantly related isolates from each haloarchaeal genus should be mapped to conduct a proper test.

On the other hand, the various genera within the haloarchaea may simply have diverged far enough in the past for rearrangements occurring at a low rate to accumulate until all signs of map conservation have been eliminated. To determine whether a punctuated or gradualist pattern of evolution prevails within the haloarchaea, genomic maps separated by a variety of time periods would need to be examined. By plotting the degree of difference between genomes (using DERANGE II or a similar tool) against divergence time (measured using 16S rDNA sequence for example), one could find whether the genomic maps have been changing at a more or less constant rate or if punctuated evolution is the norm.

The Questions People Ask

Why Genomics

In recent years, a great deal of work has been devoted to the area generally known as genomics. This is especially true of studies directed at DNA sequence itself—fingerprinting, mapping and sequencing—as opposed to those dealing with nucleoid

structure for example. Why might this be the case? One reason is certainly because the ability to do these sorts of studies is relatively new and science and scientists are generally enamored of anything novel. Any single explanation is surely too simplistic, however, to account for the enthusiasm with which genomics is pursued today and other reasons are evident. One important consideration can be termed the 'tyranny of technology' The purpose of technology is to allow people to do things they previously could not or to facilitate things they could. Like most people, researchers are often under internal and external pressures to do more with less so it is little wonder that PFGE-based genome mapping and automated sequencing have been pursued with such vigor. Investigations using these technologies are certainly contributing to our knowledge in constructive ways and will continue to do so in the future yet there are consequences to consider. Such wholehearted commitment to a small number of techniques does tend to channel inquiry in certain directions while leaving others fallow. This effect is exaggerated when the technology advances in one area far more than in others as it has with nucleic acid sequencing. Our ability to sequence DNA has far outstripped our ability to deal with all the data. Evidence for this is the amount of ongoing work devoted to simply organizing the data in a form that allows easy access and retrieval (Médigue et al., 1993; Médigue et al., 1995; Moszer et al., 1995; Nierlich, 1996). A great deal of effort is also being put into categorizing open reading frames and putative genes into families (Riley, 1993), assigning functions to ORFs by sequence similarity to known homologs usually at the amino acid level (Gonnet et al., 1992; Casari et al., 1995), and the compilation of so-called gene inventories (Fleischmann et al., 1995; Fraser et al., 1995; Bult et al., 1996). Considering the types of investigations being conducted, what kinds of answers can one expect to get?

It has become standard procedure in many labs to search the sequence databases (genbank, embl, swissprot, and pir) for possible homologs once a new gene has been sequenced as it is often the fastest and easiest way to identify and assign a function to newly sequenced DNA. This methodology has been quite successful as evidenced by the numbers of genes that have been identified in the *Haemophilus* and *Mycoplasma* genomes (Fleischmann et al., 1995; Fraser et al., 1995). It is easy to forget that years of biochemical and genetic investigation, usually in *E. coli*, lies behind this success rate. All database searches ultimately rely on the painstaking, often time-consuming process of biochemical characterization of a gene's product. Since database searches are much easier and faster to perform, genes identified this way are increasing at a much faster rate than are genes identified biochemically. The consequence of this is that only genes that are highly conserved or closely related to characterized genes can be identified with much confidence. As newly sequenced genes diverge from previously characterized genes, the amount of information that can be gleaned from database matches quickly drops. While genes can often be identified as belonging to a particular family (a class of dehydrogenases for example), specific functions are often lacking. This sort of broad assignment is of value for certain types of studies but not others. After all, just how useful is the revelation that one's gene of interest is related to an antigenic protein in *Treponema* if no other information is available?

In an analogy to the tyranny of technology, this can be termed the 'tyranny of prior knowledge'. Genes are identified only if they are related to genes that have already been identified. Thus the discovery of novel genes is greatly hindered. Novel genes include genes specific to a lineage that may impart important properties to that lineage such as bacteriorhodopsin in *Hb. salinarum*. A related problem is the misidentification of genes. A

gene may be identified through database searches as a maltose transporter. However, it is hard to say whether the few amino acid substitutions present in the newly sequenced gene do not alter the specificity of its substrate so that it is not involved in the transport of maltose at all. If the database match that initially led to the identification was itself identified using the database, then the assignment is made even more tenuous. This is a problem because of the tendency of database assignments to take on the authority of fact regardless of the quality of the initial match.

A lot of stock has been put into the value of gene inventories. However, there are large holes in the prokaryotic genome sequences already completed in the form of unidentified ORFs (Fleischmann et al., 1995; Fraser et al., 1995). The speed with which DNA is sequenced today is encouraging researchers to look for ways to identify these ORFs at a comparable rate. Classical biochemical techniques are far too slow to keep up, so various comparative methods are being developed (Gonnet et al., 1992; Casari et al., 1995; Labedan and Riley, 1995a, 1995b). Of course, this leads back to the tyranny of prior knowledge problem discussed above. For the most divergent genes, comparative identification involves motifs discernible at the amino acid level which are used to assign certain functions or properties to genes. The Blocks (Henikoff and Henikoff, 1991) and Prosite (Bairoch, 1992) databases and the Ancient Conserved Regions (ACR's) described in Green et al. (1993) and Koonin et al. (1995) are examples of this. This sort of work is valuable in that it can be applied to elucidating the evolution of genes and gene families, perhaps discerning between common descent and sequence convergence and thus contributing to our understanding of how a genome reached its current state. As more sequence is obtained, the dream of every biochemist comes closer to hand: namely the prediction of the specific function, substrate, and dynamics of a protein solely from

sequence data. To realize this goal, however, the genes that underlie such predictions will have to be positively identified and thoroughly characterized in a way that comparative sequence analysis cannot achieve. The practical upshot of all this is that while gene inventories will indeed tell us some things about prokaryotic genome functioning and evolution, much more follow up work will be necessary before their potential can be realized.

An interesting effect of all the sequence data that is being produced is that it is becoming increasingly difficult to publish sequence data without accompanying it with an analysis of some sort. This analysis might include expression studies, a multiple alignment with related genes, or mutagenesis of the coding region or the promoter to elucidate function. This trend will probably be the single biggest factor inducing researchers to take a closer look at the characterization of the genes they sequence in the future. This will only help genomics by providing more detailed information on large numbers of genes mitigating the problem outlined in the previous paragraph.

In addition to applying genome sequencing to the study of individual genes or operons, investigations into the structure and functioning of the genome itself has already gained much and is poised to gain even more. The large scale expression studies mentioned in chapter 1 (Trieselmann and Charlebois, 1992; Chuang et al., 1993; Ferrer et al., 1996) are an example where the distribution of genes on the genome expressed under different environmental conditions can be investigated, perhaps providing insights into nucleoid structure. Work such as that by Williamson et al. (1993) and Eyre-Walker (1995, 1996) which use sequence data to investigate the distribution of genes on the chromosome are other examples.

Another important aspect of genome functioning that is rapidly gaining ground is a result of the comparative genomic work that is being conducted. In the past, studies on genomic recombination and rearrangements have been limited to a small number of well characterized genomes such as *E. coli* and *S. typhimurium*. Now, thanks to rapid genome fingerprinting and mapping techniques, alterations to genomes can be investigated in a wide variety of organisms. This sort of work has already proved its worth for characterizing the pattern of rearrangements between genomes (Hackett et al., 1994; Fonstein et al., 1995), identifying the locations of virulence factors on a genome (Canard et al., 1992), elucidating the phylogenetic relationships between organisms (Carlson and Kolstø, 1993; López-García, 1993), and inspiring testable hypotheses about the evolution of certain groups of prokaryotes (this thesis). The lack of resolution inherent in many of the comparisons done so far will be lessened as more prokaryotic genomes are sequenced. As more such data become available, the pressure to use it will also increase until low-resolution maps will no longer be in favour, much like raw sequence data is considered insufficient now. DNA fingerprinting and sequencing have already been combined to elucidate that the source of the polymorphisms seen in nine strains of *Enterococcus faecalis* are more often due to rearrangements rather than point mutations (Hall, 1994). Comparative genomics using map information will provide answers not on the evolution of genes, but on the evolution of genomes; how they change, how they don't change, the rates of change and perhaps the circumstances that can promote change.

Future Directions

What does the future hold for prokaryotic genomics? More genome sequences certainly. Table 1.4 lists 18 bacterial and archaeal genomes that are in the process of being

sequenced in addition to the three that have already been completed and are in the public domain. Most of these projects involve collaborations between groups of laboratories. This sort of work constitutes an enormous investment by both the laboratories involved and the funding agencies that support them. To date, most of the projects listed in Table 1.4 have not published any results of their sequencing and this, combined with the monetary pressures of genome sequencing, probably means that the drive to initiate new projects is nearing its end if it is not already there. This is not to say that additional prokaryotic genomes will not be sequenced. Now that organizations like TIGR have been established, they will do their best to perpetuate their own existence and that means more sequencing. Given the interest in the field, sequencing technology will continue to improve for some time, allowing the price per base to come down, as well as speeding up the process. Whether technical advances will be enough to allow individual labs to conduct their own sequencing projects is difficult to say but if it does, enough genomes will probably have already been sequenced to make the exercise unattractive unless it is part of a larger investigation involving extensive analysis. In other words, whole genome sequences won't be enough, just as gene sequences are no longer enough today.

How will the genome sequencers of today respond to this? The more evolutionarily minded will likely turn towards comparisons. Today, top-down maps are the tools of choice when conducting genomic comparisons. Tomorrow, as model genomes from diverse prokaryotic lineages become available, sequence level comparisons will become more prevalent. This is absolutely necessary for comparing distantly related genomes as witnessed by St. Jean and Charlebois (1996 [chapter 6]) and the demand for more than just sequence data will quicken the adoption to these more detailed methods. To keep costs down, rather than comparing whole genomes, selected regions may be compared in

detail using one or a few completed genome sequences as references. Population genetics will probably receive another boost as numbers of closely related strains are sequenced in whole or in part to investigate the dynamics of genetic diversity in various lineages. This will also be a step forward in the elucidation of the relative roles of homologous and illegitimate recombination in the evolution of prokaryotic genomes. Comparison studies using DNA fingerprinting will survive for some time because of the particular demands of the clinical setting; high volume and short turnaround combined with ease of use. Comparisons using PFGE maps with a few genes placed on them will not be adequate for much longer, however, and methods providing more detail will have to be adopted. Also, much of the comparative work that is conducted today is purely descriptive with very little to say about possible causes. This is due in part to the low resolution of many comparisons which will change as more comparisons based on sequence data are conducted. Hopefully, works such as Charlebois and St. Jean (1995 [chapter 4]) and St. Jean and Charlebois (1996 [chapter 6]) will help to stimulate more consideration for the theoretical aspects of comparative genomics.

At the same time as all this sequencing is going on, the computationally minded will continue to produce tools for the maintenance and analysis of sequence data. As dedicated sequencing facilities such as TIGR continue to churn out prokaryotic genomes at a rate of one or two a year, more and more people will turn to the analysis side of the equation simply because there will be too much data to ignore. So far, much of the effort in sequence analysis has gone into the identification of ORFs, the assignment of functions for these ORFs, and the categorizing of genes into functional groups and evolutionarily related gene families. All this illustrates just how new whole genome sequences are. Once again, the need for the direct characterization of ORFs is apparent to identify novel

functions and gene families. The comparison of rearranged segments of DNA is one area of genomics that has attracted researchers for some time (Watterson et al., 1982). The development of computer programs to deal with comparative genomics continues (Bafna and Pevzner, 1995; Blanchette et al., 1996) and the use of DERANGE II in St. Jean and Charlebois (1996 [chapter 6]) is an example of the practical application of one of these programs. High resolution comparative studies that make use of sequence data will likely rely heavily on such programs in the future. In addition to this, other types of genomic sequence analysis is needed. Investigations such as those by Eyre-Walker (1995, 1996), and Deschavanne and Filipski (1995) that deal with the causes behind genomic level structure are few and more are sorely needed.

Finally, it must not be forgotten that the genome of a prokaryote and its nucleoid are two ways of looking at the same thing. It is often hard to relate one to the other; a collection of genes arranged singly or in operons one after the other on a linear or circular molecule of DNA or a compact, constantly changing 3-dimensional structure complete with associated proteins. The reconciliation of these two points of view is vitally important, however, if we are to fully understand why genomes look and change the way they do. Efforts in this direction have and are being made through electron microscopy, investigations into the effects of topoisomerases and histone-like proteins on DNA, and the patterns and mechanisms of genomic rearrangements. An assumption made by many of these studies is that various replicons (chromosomes, plasmids, cosmids) all possess a similar structure within the cell. In other words, plasmids and cosmids are just smaller versions of the chromosome. Ishiura et al. (1990) called this assumption into question by finding that plasmids and cosmids responded differently with respect to rearrangements to specific mutations in an *E. coli* host. Many investigations into protein interactions with

DNA use plasmids and infer similar effects for chromosomal DNA. Perhaps other methods should be used instead which measure effects on the chromosome directly. Because of the phylogenetic position of the Archaea relative to the Bacteria and Eucarya and the lack of knowledge concerning archaeal nucleoids, the opportunities for discovery in this area are great. Hopefully the *Methanococcus jannaschii* and *Sulfolobus solfataricus* genome sequences and the investigations into archaeal histone-like proteins will encourage work in this area.

In any case, genomics will be around for quite a while. It promises to provide insights into the tempo and mode of evolution of the genome, how it is regulated both globally and locally, and the interplay between the forces acting upon it to change or remain conserved. The future will reveal more detailed information on more prokaryotic organisms than we have ever seen before and this will provide the raw material for a deeper understanding of what being a prokaryote is all about.

Appendix A

Protocols and Recipes

Low Salt Medium for *Hf. volcanii* and *Hf. mediterranei* (Daniels et al., 1984)

For 1 L of medium.

Salt Solution (final volume 950 mL in distilled water)	
NaCl	125 g
KCl	10 g
$MgSO_4 \cdot 7H_2O$	10 g
$MgCl_2 \cdot 6H_2O$	45 g
$CaCl_2 \cdot 2H_2O$	1.34 g
agar (plates only)	15 g

Food Solution (final volume 50 mL in distilled water)	
Bacto tryptone	5 g
Bacto yeast extract	3 g

Autoclave solutions separately for 20 minutes then combine.

High Salt Medium for *Hb. salinarum* (Cline and Doolittle, 1987)

For 1 L of medium.

Salt Solution (final volume 950 mL in distilled water)	
NaCl	250 g
Na_3 citrate	3 g
KCl	2 g
$MgSO_4 \cdot 7H_2O$	20 g
$CaCl_2 \cdot 2H_2O$	0.2 g
agar (for plates only)	15 g

Food Solution (final volume 50 mL in distilled water)	
Bacto tryptone	5 g
Bacto yeast extract	3 g

Autoclave separately for 20 minutes and combine.

YT Medium for *E. coli* (Messing, 1983)

For 1 L of medium.

tryptone	8 g
yeast extract	5 g
NaCl	5 g
agar (for plates only)	15 g

Autoclave 20 minutes. Add 30 µg/mL kanamycin sulfate for selection of cosmid DNA.

Alkaline Extraction of plasmid and cosmid DNA from *E. coli* (Sambrook et al., 1989)

Materials

TEG	20 mM Tris•HCl pH 7.6
	50 mM EDTA
	1% (wt/vol) glucose
0.2 M NaOH/1% (wt/vol) SDS	
7.5 M $NH_4OAcetate$	
1:1 phenol/chloroform in TE pH 7.6	
isopropanol	
95% ethanol	
80% ethanol	84 mL ethanol
	20 mL water
TE pH 7.6	10 mM Tris•HCl pH 7.6
	1 mM EDTA

Procedure

1. Grow *E. coli* in 12 mL cultures at 37°C with shaking to late log phase.
2. Centrifuge cells 10 minutes at 4 300 rpm.
3. Pour off supernatant and resuspend cells in 200 µL TEG. Use a pipette tip to break up pellet.
4. Add 400 µL NaOH/SDS and mix. Add 300 µL $NH_4OAcetate$ and mix.
5. Transfer liquid to a 2 mL tube and centrifuge 7 minutes at 12 000 rpm in a benchtop centrifuge.
6. Remove pellet with a sterile toothpick and discard. Add 400 µL phenol/chloroform to supernatant and mix.
7. Centrifuge 3 minutes at 12 000 rpm and transfer aqueous phase to a fresh 2 mL tube.
8. Add 0.6 volumes of isopropanol and mix. Centrifuge 7 minutes at 12 000 rpm and remove supernatant.

9. Add 400 µL TE and incubate at 65°C-75°C for 10 minutes.
10. Add 0.5 volumes $NH_4OAcetate$ and 2 volumes ethanol and mix. Centrifuge for 7 minutes at 12 000 rpm.
11. Remove supernatant and add 600 µL 80% ethanol and mix. Centrifuge 5 minutes at 12 000 rpm.
12. Remove supernatant and resuspend pellet in 50-100 µL TE. Store at 4°C or -20°C.

Total DNA Extraction from Haloarchaea (Lam and Doolittle, 1992)

Materials

salt solution (100 mL)	NaCl	20 g
	KCl	0.37 g
	$MgSO_4 \cdot 7H2O$	3.7 g
	10 µM $MnCl_2$	17 µL
	2 M Tris•HCl pH 7.2	2.5 µL
TE pH 8.0	50 mM Tris•HCl pH 8.0	
	1 mM EDTA	
STE	50 mM Tris•HCl pH 7.6	
	50 mM EDTA	
	1% (wt/vol) N-lauroylsarcosine (sarkosyl)	
1:1 phenol/chloroform in TE pH 7.6		
95% ethanol		
TE pH 7.6	50 mM Tris•HCl pH 7.6	
	1 mM EDTA	

Procedure

1. Grow cells with shaking to late log phase. Volumes as small as 4 mL are adequate for this procedure and cultures should be divided into aliquots small enough to allow the use of a benchtop microcentrifuge.
2. Centrifuge cells at 10 000 rpm for 7 minutes.
3. Remove supernatant and resuspend cells in 1/10 growth volume of salt solution.
4. Add ½ growth volume TE pH 8.0 to lyse cells and vortex.
5. Add 1 volume phenol/chloroform, vortex and spin at 12 000 rpm for 10 minutes.
6. Remove aqueous phase to a fresh tube and repeat phenol/chloroform extraction until interface is clean.
7. Add 2 volumes ethanol and mix. Centrifuge 7 minutes at 12 000 rpm.
8. Remove supernatant and air dry pellet. Resuspend in 50-100 µL TE or sterile water. Store DNA at 4°C or -20°C.

Chocolate Chip, Sour Cream Cake

For 10 to 12 servings.

Materials

soft butter or margarine	6 tblsp.
sugar	1 cup
eggs	2
unsifted all purpose flour	11/3 cups
baking powder	1 ½ tsp.
baking soda	1 tsp.
cinnamon	1 tsp.
sour cream	1 cup
semi-sweet chocolate chips	1 pkg. (6 oz.)
sugar	1 tblsp.

Procedure

1. In a large bowl, beat butter and 1 cup sugar until blended.
2. Beat in eggs one at a time.
3. In a second bowl, add together flour, baking powder, baking soda, and cinnamon and mix. Add to creamed mixture.
4. Fold in sour cream.
5. Pour batter into a greased and floured 9 X 13 inch baking pan and scatter chocolate chips evenly over the batter.
6. Sprinkle the 1 tblsp. of sugar over top and bake at 350°F for 35 minutes or until cake just begins to pull away from the sides of pan. Allow to cool before serving.

Vacuum Transfer of DNA for Southern Blots using a Tyler Research Instruments VT-20

Materials

depurination solution	0.25 M HCl
denaturation solution	1.5 M NaCl
	0.5 M NaOH
transfer solution	0.4 M NaOH
2X SSC	0.3 M NaCl
	35 mM Na_3 citrate

Procedure

1. Load vacuum transfer unit as follows: frit wetted with distilled water, nylon membrane wetted with distilled water, mask, and agarose gel to be transferred.
2. Apply vacuum and check for leaks in gel or mask. Leaks can be sealed using molten agarose.
3. Pour enough depurination solution over gel to just cover surface. Let stand for 3 minutes for regular agarose gels and up to 6 minutes for pulsed-field gels.
4. Tilt transfer unit and remove depurination solution with a pipette.
5. Pour enough denaturation solution over gel to just cover surface. Let stand for 3 minutes.
6. Remove denaturation solution as above. Add transfer solution until transfer unit is half full. Let stand for 30 minutes.
7. Remove transfer solution. Remove gel and then turn off vacuum.
8. Rinse membrane in 2X SSC and air dry. If gel was stained with ethidium bromide, transfer can be checked by viewing membrane on an ultra violet light box.
9. If membrane is to be used repeatedly, irradiate membrane in an ultra violet light box for 5 minutes.

Random-Primer Labelling of DNA with ^{32}P and Hybridization to Southern and Dot Blots Using a Tyler Research Instruments HI 16000 Hybridization Incubator

Materials

hybridization solution (for 500 mL)

NaCl	29.2 g
1 M Tris pH 7.6	25 mL
SDS	25 g
formamide	250 mL

up to 500 mL with distilled water

herring sperm DNA	10 µg/µL
random hexamers	2.5 µg/µL
10X RP-C	200 mM Tris•HCl pH8
	100 mM $MgCl_2 \cdot 6H_2O$
	50 mM dithiothreitol
	600 µM each of dGTP, dTTP, dATP
	25% glycerol
bovine serum albumin (BSA)	1 µg/µL
[α-^{32}P]dCTP	10 µCi/µL
labelling grade Klenow fragment	2 Units/µL

7.5 M $NH_4OAcetate$

95% ethanol

TE pH 7.6	50 mM Tris•HCl pH 7.6
	1 mM EDTA
2X SSC	0.3 M NaCl
	35 mM Na_3 citrate

2X SSC/1% SDS

Procedure

1. Place membranes in tubes of hybridization incubator. Add a minimum of hybridization solution to each tube, just enough to leave a small amount of liquid in the tube, anywhere from 8 mL to 25 mL.
2. Boil herring sperm DNA for 5 minutes and add 50 µL to each tube.
3. Prehybridize membranes in incubator at least 2 hours at 38°C-40°C.
4. Digest DNA for probes if it is not already linear. Prepare probes by adding together 1 µL random hexamers, 5 µL sterile water and 2.5 µL DNA.
5. Boil DNA and hexamers for 5 minutes.
6. Add to each probe the following:
   1.25 µL 10X RP-C
   1.5 µL BSA
   1.0 µL ^{32}P-dCTP
   0.5 µL Klenow fragment
7. Incubate probes at room temperature for 5 minutes then at 37°C for 40 minutes.
8. Add ½ volume $NH_4OAcetate$ and 2 volumes ethanol to probes. Vortex and centrifuge for 5 minutes.
9. Remove supernatant and check probes using Gieger counter.
10. Resuspend probes in approximately 100 µL TE.
11. Boil probes for 5 minutes.
12. Add probes to membranes and hybridize for one to two hours for Southern blots and overnight for dot blots. Hybridize at 38°C-40°C.
13. Pour out hybridization solution and rinse membranes twice in room temperature 2X SSC.
14. Wash membranes in preheated 2X SSC/1% SDS at 65°C-70°C for one hour.
15. Remove membranes from tubes and rinse in 2X SSC.
16. Expose membranes to X-ray film.

Extraction of Haloarchaeal Genomic DNA for Pulsed-Field Gel Electrophoresis

Materials

FMC SeaPlaque GTG agarose dissolved in 1 M NaCl

lysis buffer	1 M Tris pH 7.6
	0.5 M EDTA
	1% N-lauroylsarcosine
TE pH 7.6	50 mM Tris•HCl pH 7.6
	1 mM EDTA

TE/1 M NaCl

Procedure

1. Grow cells in appropriate medium until saturation. Incubate 4 to 6 mL at 55°C for 10 minutes and mix with an equal volume of 1% agarose at 55°C.
2. Pour mixture onto a clean flat surface (glass or acrylic plates work well) and allow to harden.
3. Slice agarose and suspended cells with a coverslip into rectangular plugs and deposit plugs into a 50 mL screw capped tube.
4. Lyse cells by filling the tube with lysis buffer and incubating at 55°C for one hour.
5. Remove lysis buffer and add TE/1 M NaCl. Incubate on ice for at least two hours.
6. Repeat step 5 three times. Incubations should be for as long as practical. At least one overnight incubation on ice is highly recommended.
7. Remove buffer and fill tube with TE. Incubate on ice as above and repeat once.
8. Store plugs in TE at 4°C.

GeneClean® Kit for the Isolation of DNA Suspended in Agarose

Materials

6 M NaI	
glassmilk	
NEWwash	NaCl
	Tris
	EDTA
	ethanol
	water
TE pH 7.6	50 mM Tris•HCl pH 7.6
	1 mM EDTA

Procedure

1. Cut DNA band from agarose gel taking as little agarose as possible.
2. Place DNA in agarose in a 1.5 mL polypropylene tube and add 1 mL NaI. Place tube in a 45°C water bath and incubate for 5 minutes. Mix tube once or twice during incubation.
3. Add 5 µL glassmilk and mix. Incubate on ice for 5 to 10 minutes depending on the expected yield of DNA. Mix tube every 2 minutes.
4. Spin tube in a microcentrifuge for 5 seconds and remove supernatant.
5. Add 600 mL of ice-cold NEWwash solution. Vortex to resuspend glassmilk.
6. Repeat steps 4 and 5 twice.
7. Resuspend pellet in 10 µL of TE and incubate at 45°C for 3 minutes.
8. Centrifuge at 12 000 rpm for 30 seconds and transfer DNA containing supernatant to a fresh tube.
9. Add 5 µL of TE to glassmilk pellet and use a pipette tip to resuspend.
10. Incubate at 45°C for 3 minutes, centrifuge for 30 seconds and combine supernatant with previously transferred 10 µL. Store DNA at -20°C.

APPENDIX B

The randomize() function of compagen.dll

Following is a copy of the function prototype found in compagen.dll.

```
int __stdcall randomize(int lowerBound, int upperBound, int numberOfSets, char
*outputfile);
```

Any replacement of compagen.dll must contain a function with this prototype.

Parameters

COMPAGEN passes the following four parameters to the randomize function in compagen.dll:

int lowerBound	Designates the lower bound of the range of numbers to randomize.
int upperBound	Designates the upper bound of the range of numbers to randomize.
int numberOfSets	Designates the number of random permutations to generate.
char *outputfile	Designates the output file name. This can include a path.

Return value

COMPAGEN expects to receive an integer variable from randomize(). COMPAGEN doesn't do anything with this integer.

Calling Convention

COMPAGEN uses the standard calling convention (stdcall) to call randomize(). Because of this, any exported function written in C++ must also specify stdcall as its calling convention. The specifier '__stdcall' in the above function prototype is the Microsoft® Visual C++ version 4.0 keyword and may differ on different compilers.

Output

The randomize() function creates or opens and truncates the file specified by outputfile. COMPAGEN requires a very precise format for this file to be read properly so the file generated by compagen.dll should be studied carefully. Of note are the spaces present at the end of each line and the fact that the last line does not have a carriage return.

Appendix C

Web Sites Relevant to Prokaryotic Genomics

American Society for Microbiology (ASM)	www.asmusa.org
E. coli Genetic Stock Center	cgsc.biology.yale.edu
Encyclopedia of *E. coli* genes and metabolism (EcoCyc)	www.ai.sri.com/ecocyc/ecocyc.html
European Molecular Biology Laboratories (EMBL)	www.embl-heidelberg.de
Multipurpose Automated Genome Project Investigation Environment (MAGPIE)	www.mcs.anl.gov/home/gaasterl/magpie.html
Mycoplasma genome database	kiev.physchem.kth.se/MycDB.html
National Center for Biotechnology Information (NCBI)	www.ncbi.nlm.nih.gov
Non-Redundant Database for *Bacillus subtilis* (NRSub)	ddbjs4h.genes.nig.ac.jp
Ribosomal Database Project (RDP)	rdp.life.uiuc.edu
The Institute for Genomic Research (TIGR)	www.tigr.org
WFCC World Data Centre for Microorganisms	www.wdcm.riken.go.jp

REFERENCES

1. **Adams, D.E., E.M. Shekhtman, E.L. Zechiedrich, M.B. Schmid, and N.R. Cozzarelli.** 1992. The role of topoisomerase IV in partitioning bacterial replicons and the structure of catenated intermediates in DNA replication. Cell **71**:277-288.

2. **Akhmanova, A.S., V.K. Kagramanova, and A.S. Mankin.** 1993. Heterogeneity of small plasmids from halophilic archaea. J. Bacteriol. **175**:1081-1086.

3. **Albertini, A.M., M. Hofer, M.P. Calos, and J.H. Miller.** 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell **29**:319-328.

4. **Allgood, N.D., and T.J. Silhavy.** 1988. Illegitimate recombination in bacteria, p. 309-330. *In* R. Kucherlapati, and G.R. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.

5. **Allgood, N.D., and T.J. Silhavy.** 1991. *Escherichia coli xonA* (*sbcB*) mutants enhance illegitimate recombination. Genetics **127**:671-680.

6. **Altamura, S., J.L. Sanz, R. Amils, P. Cammarano, and P. Londei.** 1988. The antibiotic sensitivity spectra of ribosomes from the *Thermoproteales*: phylogenetic depth and distribution of antibiotic binding sites. Syst. Appl. Microbiol. **10**:218-225.

7. **Anagnostopoulos, C., P.J. Piggot, and J.A. Hoch.** 1993. The genetic map of *Bacillus subtilis*, p. 425-461. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics*. American Society for Microbiology, Washington, D.C.

8. **Antón, J., P. López-García, J.P. Abad, C.L. Smith, and R. Amils.** 1994. Alignment of genes and *Swa*I restriction sites to the *Bam*HI genomic map of *Haloferax mediterranei*. FEMS Microbiol. Lett. **117**:53-60.

9. **Antón, J., R. Amils, C.L. Smith, and P. López-García.** 1995. Comparative restriction maps of the archael megaplasmid pHM300 in different *Haloferax mediterranei* strains. System. Appl. Microbiol. **18**:439-447.

10. **Aoyama, T., and M. Takanami.** 1988. Supercoiling response of *E. coli* promoters with different spacer lengths. Biochim. Biophys. Acta **949**:311-317.

11. **Arnold G.F., and I. Tessman.** 1988. Regulation of DNA superhelicity by *rpoB* mutations that suppress defective rho-mediated transcription termination in *Escherichia coli*. J. Bacteriol. **170**:4266-4271.

12. **Azevedo, V., E. Alvarez, E. Zumstein, G. Damiani, V. Sgaramella, S.D. Ehrlich, and P. Serror.** 1993. An ordered collection of *Bacillus subtilis* DNA segments cloned in yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA **90:**6047-6051.

13. **Bafna, V., and P.A. Pevzner.** 1995. Sorting by reversals: genome rearrangements in plant organelles and evolutionary history of X chromosome. Mol. Biol. Evol. **12:**239-246.

14. **Bagga, R., N. Ramesh, and S.K. Brahmachari.** 1990. Supercoil-induced unusual DNA structures as transcriptional block. Nucleic Acids Res. **18:**3363-3369.

15. **Bairoch, A.** 1992. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **20:**2013-2018.

16. **Balch, W.E., G.E. Fox, L.J. Magrum, C.R. Woese, and R.S. Wolfe.** 1979. Methanogens: reevaluation of a unique biological group. Microbiol. Rev. **43:**260-296.

17. **Balch, W.E., L.J. Magrum, G.E. Fox, R.S. Wolfe, and C.R. Woese.** 1977. An ancient divergence among the bacteria. J. Mol. Evol. **9:**305-311.

18. **Balke, V.L., and J.D. Gralla.** 1987. Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli.* J. Bacteriol. **169:**4499-4506.

19. **Bates, A.D., and A. Maxwell.** 1993. DNA topology. Oxford University Press, Inc., New York.

20. **Bayley, S.T., and R.A. Morton.** 1978. Recent developments in the molecular biology of extremely halophilic bacteria. CRC Crit. Rev. Microbiol. **6:**151-205.

21. **Beckwith J.R., E.R. Signer, and W. Epstein.** 1966. Transposition of the *Lac* region of *E. coli.* Cold Spring Harbor Symp. Quant. Biol. **31:**393-401.

22. **Ben-Mahrez, K., W. Sougakoff, M. Nakayama, and M. Kohiyama.** 1988. Stimulation of an alpha like DNA polymerase by v-*myc* related protein of *Halobacterium halobium*. Arch. Microbiol. **149:**175-180.

23. **Betlach, M., F. Pfeifer, J. Friedman, and H.W. Boyer.** 1983. Bacterio-opsin mutants of *Halobacterium halobium*. Proc. Natl. Acad. Sci. USA **80:**1416-1420.

24. **Birkenbihl R.P., and W. Vielmetter.** 1989. Complete maps of IS1, IS2, IS3, IS4, IS5, IS30 and IS150 locations in *Escherichia coli* K12. Mol. Gen. Genet. **220:**147-153

25. **Biserùić M., and H. Ochman.** 1993a. Natural populations of *Escherichia coli* and *Salmonella typhimurium* harbor the same classes of insertion sequences. Genetics **133:**449-454

26. **Biserùić M., and H. Ochman.** 1993b. The ancestry of insertion sequences common to *Escherichia coli* and *Salmonella typhimurium*. J. Bacteriol. **175:**7863-786.

27. **Blake, R.D., and W.P. Hinds.** 1984. Analysis of the codons bias in *E. coli* sequences. J. Biomol. Struct. Dynam. **2:**593-606.

28. **Blanchette, M., T. Kunisawa, and D. Sankoff.** 1996. Parametric genome rearrangement. Gene **172:**GC11-GC17.

29. **Blanck, A., D. Oesterhelt, E. Ferrando, E.S. Schegk, and F. Lottspeich.** 1989. Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor. EMBO J. **8:**3963-3971.

30. **Bliska, J.B., and N.R. Cozzarelli.** 1987. Use of site-specific recombination as a probe of DNA structure and metabolism *in vivo*. J. Mol. Biol. **194:**205-218.

31. **Bloch, G.A., C.K. Rode, V. Obreque, and K.Y. Russell.** 1994. Comparative genome mapping with mobile physical map landmarks. J. Bacteriol. **176:**7121-7125.

32. **Boone, D.R., S. Worakit, I.M. Mathrani, and R.A. Mah.** 1986. Alkaliphilic methanogens from high-pH lake sediments. System. Appl. Microbiol. **7:**230-234.

33. **Bork, P., C. Ouzounis, G. Casari, R. Schneider, C. Sander, M. Dolan, W. Gilbert, and P.M. Gillevet.** 1995. Exploring the *Mycoplasma capricolum* genome: a minimal cell reveals its physiology. Mol. Microbiol. **16:**955-967.

34. **Borodovsky, M., E.V. Koonin, and K.E. Rudd.** 1994. New genes in old sequence: a strategy for finding genes in the bacterial genome. Trends in Biochem. Sci. **19:**309-313.

35. **Borowiec, J.A., and J.D. Gralla.** 1987. All three elements of the *lac* p^S promoter mediate its transcriptional response to DNA supercoiling. J. Mol. Biol. **195:**89-97.

36. **Borowiec, J.A., L. Zhang, S. Sasse-Dwight, and J.D. Gralla.** 1987. DNA supercoiling promotes formation of a bent repression loop in *lac* DNA. J. Mol. Biol. **196:**101-111.

37. **Bourke, B., P. Sherman, H. Louie, E. Hani, P. Islur, and V.L. Chan.** 1995. Physical and genetic map of the genome of *Campylobacter upsaliensis*. Microbiology (Reading) **141:**2417-2424.

38. **Bouthier de la Tour, C., C. Portemer, R. Huber, P. Forterre, and M. Duguet.** 1991. Reverse gyrase in thermophilic eubacteria. J. Bacteriol. **173:**3921-3923.

39. **Bouthier de la Tour, C., C. Portemer, M. Nadal, K.O. Stetter, P. Forterre, and M. Duguet.** 1990. Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. J. Bacteriol. **172:**6803-6808.

40. **Boutrou, R., D. Thuault, and C.M. Bourgeois.** 1995. Identification and characterization of *Streptococcus thermophilus* strains by pulsed-field gel electrophoresis. J. Appl. Bacteriol. **79**:454-458.

41. **Bracco, L., D. Kotlarz, A. Kolb, S. Diekmann, and H. Buc.** 1989. Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli*. EMBO J. **8**:4289-4296.

42. **Brahms, G., S. Brahms, and B. Magasanik.** 1995. A sequence-induced superhelical DNA segment serves as transcriptional enhancer. J. Mol. Biol. **246**:35-42.

43. **Bratley, P., B.L. Fox, and E.L. Schrage.** 1983. A guide to simulation. Springer Verlag, New York, NY.

44. **Brewer, B.J.** 1988. When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. Cell **53**:679-686.

45. **Brewer, B.J.** 1990. Replication and the transcriptional organization of the *Escherichia coli* chromosome, p. 61-83. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

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

47. **Broyles, S.S., and D.E. Pettijohn.** 1986. Interaction of the *Escherichia coli* HU protein with DNA. Evidence for formation of nucleosome-like structures with altered DNA helical pitch. J. Mol. Biol. **187**:47-60.

48. **Brunier, D., B.P.H. Peeters, S. Bron, and S.D. Ehrlich.** 1989. Breakage-reunion and copy choice mechanisms of recombination between short homologous sequences. EMBO J. **8**:3127-3133.

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

50. **Burland, V., G. Plunkett, D.L. Daniels, and F.R. Blattner.** 1993. DNA sequence and analysis of 136 kilobases of the *Escherichia coli* genome: organizational symmetry around the origin of replication. Genomics **16**:551-561.

51. **Burland, V., G. Plunkett, H.J. Sofia, D.L. Daniels, and F.R. Blattner.** 1995. Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res. **23**:2105-2119.

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

53. **Campbell, A.M.** 1993. Genome organization in prokaryotes. Current Opinion Genet. Develop. **3**:837-844.

54. **Canard, B., B. Saint-Joanis, and S.T. Cole.** 1992. Genomic diversity and organization of virulence genes in the pathogenic anaerobe *Clostridium perfringens*. Mol. Microbiol. **6**:1421-1429.

55. **Carlson, C.R., and A.-B. Kolstø.** 1993. A complete physical map of a *Bacillus thuringiensis* chromosome. J. Bacteriol. **175**:1053-1060.

56. **Carlson, C.R., A. Grønstad, and A.-B. Kolstø.** 1992. Physical maps of the genomes of three *Bacillus cereus* strains. J. Bacteriol. **174**:3750-3756.

57. **Casari, G., M.A. Andrade, P. Bork, J. Boyle, A. Daruvar, C. Ouzounis, R. Schneider, J. Tamames, A. Valencia, and C. Sander.** 1995. Challenging times for bioinformatics. Nature (London) **376**:647-648.

58. **Casjens, S., M. Delange, H.L. Ley III, P. Rosa, and W.M. Huang.** 1995. Linear chromosomes of lyme disease agent spirochetes: Genetic diversity and conservation of gene order. J. Bacteriol. **177**:2769-2780.

59. **Cendrin, F., J. Chroboczek, G. Zaccai, H. Eisenberg, and M. Mevarech.** 1993. Cloning, sequencing, and expression in *Escherichia coli* of the gene coding for malate dehydrogenase of the extremely halophilic archaebacterium *Haloarcula marismortui*. Biochemistry **32**:4308-4313.

60. **Charlebois, R.L.** 1993. Physical mapping of genomes using the landmark strategy, p. 219-229. *In* H.A. Lim, J.W. Fickett, C.R. Cantor, and R.J. Robbins (ed.), *The second international conference on bioinformatics, supercomputing and complex genome analysis*. World Scientific, Singapore.

61. **Charlebois, R.L.** Physical map of *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB. *In* F.J. de Bruijn, J.R. Lupski, and G.Weinstock (ed.), *Bacterial genomes: physical structure and analysis*. Chapman & Hall, New York, N.Y. in press.

62. **Charlebois, R.L., and W.F. Doolittle.** 1989. Transposable elements and genome structure in halobacteria, p. 297-307. *In* M. Howe, and D. Berg (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.

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

64. **Charlebois, R.L., L.C. Schalkwyk, J.D. Hofman, and W.F. Doolittle.** 1991. A detailed physical map and set of overlapping clones covering the genome of the archaebacterium *Haloferax volcanii* DS2. J. Mol. Biol. **222**:509-524.

65. **Charlebois, R.L., and A. St. Jean.** 1995. Supercoiling and map stability in the bacterial chromosome. J. Mol. Evol. **41**:15-23.

66. **Chédin, F., E. Dervyn, R. Dervyn, S.D. Ehrlich, and P. Noirot.** 1994. Frequency of deletion formation decreases exponentially with distance between short direct repeats. Mol. Microbiol. **12**:561-569.

67. **Chen, D., R. Bowater, C.J. Dorman, and D.M.J. Lilley.** 1992. Activity of a plasmid-borne *leu-500* promoter depends on the transcription and translation of an adjacent gene. Proc. Natl. Acad. Sci. USA **89**:8784-8788

68. **Chuang, S.-E., D.L. Daniels, and F.R. Blattner.** 1993. Global regulation of gene expression in *Escherichia coli*. J. Bacteriol. **175**:2026-2036

69. **Churchill, G.A., D.L. Daniels, and M.S. Waterman.** 1990. The distribution of restriction enzyme sites in *Escherichia coli*. Nucleic Acids Res. **18**:589-597.

70. **Cline, S.W., and W.F. Doolittle.** 1987. Efficient transfection of the archaebacterium *Halobacterium halobium*. J. Bacteriol. **169**:1341-1344.

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

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

73. **Collin, R.G., H.W. Morgan, D.R. Musgrave, and R.M. Daniel.** 1988. Distribution of reverse gyrase in representative species of eubacteria and archaebacteria. FEMS Microbiol. Lett. **55**:235-240.

74. **Collins, J.** 1980. Instability of palindromic DNA in *Escherichia coli*. Cold Spring Harbor Symp. Quant. Biol. **45**:409-416.

75. **Condemine, G., and C.L. Smith.** 1990. Transcription regulates oxolinic acid-induced DNA gyrase cleavage at specific sites on the *E. coli* chromosome. Nucleic Acids Res. **18**:7389-7396

76. **Conover, R.K., and W.F. Doolittle.** 1990. Characterization of a gene involved in histidine biosynthesis in *Halobacterium* (*Haloferax*) *volcanii*: isolation and rapid mapping by transformation of an auxotroph with cosmid DNA. J. Bacteriol. **172:**3244-3249.

77. **Cook, D.N., G.A. Armstrong, and J.E. Hearst.** 1989. Induction of anaerobic gene expression in *Rhodobacter capsulatus* is not accompanied by a local change in chromosomal supercoiling as measured by a novel assay. J. Bacteriol. **171:**4836-4843.

78. **Cook, D.N., D. Ma, N.G. Pon, and J.E. Hearst.** 1992. Dynamics of DNA supercoiling by transcription in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **89:**10603-10607.

79. **Correia, A., J.F. Martín, and J.M. Castro.** 1994. Pulsed-field gel electrophoresis analysis of the genome of amino acid producing corynebacteria: chromosome sizes and diversity of restriction patterns. Microbiology (Reading) **140:**2841-2847.

80. **Coukell, M.B., and C. Yanofsky.** 1970. Increased frequency of deletions in DNA polymerase mutants of *Escherichia coli*. Nature (London) **228:**633-635.

81. **Craig, N.L., and N. Kleckner.** 1987. Transposition and site-specific recombination, p. 1054-1070. *In* F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), Escherichia coli *and* Salmonella typhimurium *cellular and molecular biology vol. 2*. American Society for Microbiology, Washington, D.C.

82. **Danchin, A.** 1995. Why sequence genomes? The *Escherichia coli* imbroglio. Mol. Microbiol. **18:**371-376.

83. **Daniels, C.J., A.H.Z. McKee, and W.F. Doolittle.** 1984. Archaebacterial heat-shock proteins. EMBO J. **3:**745-749.

84. **DasSarma, S.** 1989. Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations. Can. J. Microbiol. **35:**65-72.

85. **DasSarma, S., T. Damerval, J.G. Jones, and N. Tandeau de Marsac.** 1987. A plasmid-encoded gas vesicle protein gene in a halophilic archaebacterium. Mol. Microbiol. **1:**365-370.

86. **DeLong, E.F.** 1992. Archaea in coastal marine environments. Proc. Natl. Acad. Sci. USA **89:**5685-5689.

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

88. **Dempsey, J.O.F., A.B. Wallace, and J.G. Cannon. 1994.** The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosomes of the two mapped strains of the closely related species *N. gonorrhoeae*. J. Bacteriol. **177**:6390-6400.

89. **Derkacheva, N.I., V.K. Kagramanova, and S. Man'kin.** 1993. Genetic variability in halophilic archaebacteria (a review). Mol. Biol. **27**:287-295.

90. **Deschavanne, P., and J. Filipski.** 1995. Correlation of GC content with replication timing and repair mechanisms in weakly expressed *E. coli* genes. Nucleic Acids Res. **23**:1350-1353.

91. **DiGate, R.J., and K.J. Marians.** 1988. Identification of a potent decatenating enzyme from *Escherichia coli*. J. Biol. Chem. **263**:13366-13373.

92. **DiGate, R.J., and K.J. Marians.** 1989. Molecular cloning and DNA sequence analysis of *Escherichia coli topB*, the gene encoding topoisomerase III. J. Biol. Chem. **264**:17924-17930.

93. **DiNardo, S., K.A. Voelkel, and R. Sternglanz.** 1982. *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. Cell **31**:43-51.

94. **Doolittle, W.F., and C.J. Daniels.** 1985. Prokaryotic genome evolution: what we might learn from the archaebacteria, p. 31-44. *In* K.H. Schleifer, and E. Stackebrandt (ed.), *Evolution of prokaryotes*. Academic Press, London.

95. **Dorman, C.J.** 1995. DNA topology and the global control of bacterial gene expression: implications for the regulation of virulence gene expression. Microbiology (Reading) **141**:1271-1280.

96. **Dorman, C.J., G.C. Barr, N. Ní Bhriain, and C.F. Higgins.** 1988. DNA supercoiling and the anaerobic growth phase regulation of *tonB* gene expression. J. Bacteriol. **170**:2816-2826.

97. **Dorman, C.J., A.S. Lynch, N. Ní Bhriain, and C.F. Higgins.** 1989. DNA supercoiling in *Escherichia coli*: *topA* mutants can be suppressed by DNA amplifications involving the *tolC* locus. Mol. Microbiol. **3**:531-540.

98. **Dove, S.L., and C.J. Dorman.** 1994. The site-specific recombination system regulating expression of the Type 1 fimbrial subunit gene of *Escherichia coli* is sensitive to changes in DNA supercoiling. Mol. Microbiol. **14**:975-988.

99. **Drlica, K.** 1984. Biology of bacterial deoxyribonucleic acid topoisomerases. Microbiol. Rev. **48**:273-289.

100. **Drlica, K.** 1987. The nucleoid, p. 91-103. *In* F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), Escherichia coli *and* Salmonella typhimurium*: cellular and molecular biology, vol 1*. American Society for Microbiology, Washington, D.C.

101. **Drlica, K.** 1992. Control of bacterial DNA supercoiling. Mol. Microbiol. **6**:425-433.

102. **Drlica, K., G.J. Pruss, R.M. Burger, R.J. Franco, L.-S. Hsieh, and B.A. Berger.** 1990. Roles of DNA topoisomerases in bacterial chromosome structure and function, p. 195-204. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

103. **Drlica, K., and J. Rouvière-Yaniv.** 1987. Histonlike proteins of bacteria. Microbiol. Rev. **51**:301-319.

104. **Dubnau, D.** 1993. Genetic exchange and homologous recombination, p. 555-584. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics*. American Society for Microbiology, Washington, D.C.

105. **Dürrenberger, M., M.-A. Bjornsti, T. Uetz, J.A. Hobot, and E. Kellenberger.** 1988. Intracellular location of the histonelike protein HU in *Escherichia coli*. J. Bacteriol. **170**:4757-4768.

106. **Dürrenberger, M., A. La Teana, G. Citro, F. Venanzi, C.O. Gualerzi, and C.L. Pon.** 1991. *Escherichia coli* DNA-binding protein H-NS is localized in the nucleoid. Res. Microbiol. **142**:373-380.

107. **Dybvig, K.** 1993. DNA rearrangements and phenotypic switching in prokaryotes. Mol. Microbiol. **10**:465-471.

108. **Ebert, K., and W. Goebel.** 1985. Conserved and variable regions in the chromosomal and extrachromosomal DNA of halobacteria. Mol. Gen. Genet. **200**:96-102.

109. **Ebert, K., W. Goebel, A. Moritz, U. Rdest, and B. Surek.** 1986. Genome and gene structures in halobacteria. System. Appl. Microbiol. 7:30-35.

110. **Ebert, K., W. Goebel, and F. Pfeifer.** 1984. Homologies between heterogeneous extrachromosomal DNA populations of *Halobacterium halobium* and four new halobacterial isolates. Mol. Gen. Genet. **194**:91-97.

111. **Eickbush, T.H., and E.N. Moudrianakis.** 1978. The compaction of DNA helices into either continuous supercoils or folded fiber rods and toroids. Cell **13**:295-301.

112. **Eiglmeier, K., N. Honoré, S.A. Woods, B. Caudron, and S.T. Cole.** 1993. Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. Mol. Microbiol. 7:197-206.

113. **Elazari-Volcani, B.** 1957. Genus XII. *Halobacterium*, p. 207-212. *In* Breed, Murray, and Smith (ed.), *Bergey's manual of determinative bacteriology, 7th edition*. The Williams & Wilkins Co., Baltimore.

114. **Elhardt, D., and A. Böck.** 1982. An in vitro polypeptide synthesizing system from methanogenic bacteria: sensitivity to antibiotics. Mol. Gen. Genet. **188:**128-134.

115. **Ennis, D.G., S.K. Amundsen, and G.R. Smith.** 1987. Genetic functions promoting homologous recombination in *Escherichia coli*: a study of inversion in phage λ. Genetics **115:**11-24.

116. **Evguenieva-Hackenberg, E., and S. Selenska-Pobell.** 1995. Genome analysis of five soil bacterial isolates named formerly *Enterobacter agglomerans*. J. Appl. Bacteriol. **79:**49-60.

117. **Eyre-Walker, A.** 1995. The distance between *Escherichia coli* genes is related to gene expression levels. J. Bacteriol. **177:**5368-5369.

118. **Eyre-Walker, A.** 1996. The close proximity of *Escherichia coli* genes: consequences for stop codon and synonymous codon usage. J. Mol. Evol. **42:**73-78.

119. **Farabaugh, P.J., U. Schmeissner, M. Hofer, and J.H. Miller.** 1978. Genetic studies of the *lac* repressor. J. Mol. Biol. **126:**847-863.

120. **Ferrer, C., F.J.M. Mojica, G. Juez, and F. Rodríguez-Valera.** 1996. Differentially transcribed regions of *Haloferax volcanii* genome depending on the medium salinity. J. Bacteriol. **178:**309-313.

121. **Figueroa. N., and L. Bossi.** 1988. Transcription induces gyration of the DNA template in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **85:**9416-9420.

122. **Figueroa, N., N. Wills, and L. Bossi.** 1991. Common sequence determinants of the response of a prokaryotic promoter to DNA bending and supercoiling. EMBO J. **10:**941-949.

123. **Firrao, G., C.D. Smart, and B.C. Kirkpatrick.** 1996. Physical map of the western X-disease phytoplasma chromosome. J. Bacteriol. **178:**3985-3988.

124. **Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.-F. Tomb, B.A. Dougherty, J.M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J.D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J.M. Kelley, J.F. Weidman, C.A. Phillips, T. Spriggs, E. Hedblom, M.D. Cotton, T.R. Utterback, M.C. Hanna, D.T. Nguyen, D.M. Saudek, R.C. Brandon, L.D. Fine, J.L. Fritchman, J.L. Fuhrmann, N.S.M. Geoghagen, C.L. Gnehm, L.A. McDonald, K.V. Small, C.M. Fraser, H.O. Smith, and J.C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496-512.

125. **Fonstein, M., and R. Haselkorn.** 1995. Physical mapping of bacterial genomes. J. Bacteriol. **177**:3361-3369.

126. **Fonstein, M., E.G. Koshy, T. Nikolskaya, P. Mourachov, and R. Haselkorn.** 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**:1827-1841.

127. **Fonstein, M., T. Nikolskaya, and R. Haselkorn.** 1995. High-resolution alignment of a 1-megabase-long genome region of three strains of *Rhodobacter capsulatus*. J. Bacteriol. **177**:2368-2372.

128. **Forterre, P., N. Benachenhou-Lahfa, F. Confalonieri, M. Duguet, C. Elie, and B. Labedan.** 1993. The nature of the last universal ancestor and the root of the tree of life, still open questions. Biosystems **28**:15-32.

129. **Forterre, P., G. Mirambeau, C. Jaxel, M. Nadal, and M. Duguet.** 1985. High positive supercoiling *in vitro* catalyzed by an ATP and polyethylene glycol-stimulated topoisomerase from *Sulfolobus acidocaldarius*. EMBO J. **4**:2123-2128.

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

131. **François, V., J. Louarn, J.-E. Rebollo, and J.-M. Louarn.** 1990. Replication termination, nondivisible zones, and structure of the *Escherichia coli* chromosome, p. 351-359. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

132. **Franklin, N.C.** 1967. Extraordinary recombinational events in *Escherichia coli*. Their independence of the *rec*$^+$ function. Genetics **55**:699-707.

133. **Fraser, C.M., J.D. Gocayne, O. White, M.D. Adams, R.A. Clayton, R.D. Fleischmann, C.J. Bult, A.R. Kerlavage, G. Sutton, J.M. Kelley, J.L. Fritchman, J.F. Weidman, K.V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T.R. Utterback, D.M. Saudek, C.A. Phillips, J.M. Merrick, J.-F. Tomb, B.A. Dougherty, K.F. Bott, P.-C. Hu, T.S. Lucier, S.N. Peterson, H.O. Smith, C.A. Hutchison III, and J.C. Venter.** 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270**:397-403.

134. **Free, A., and C.J. Dorman.** 1994. *Escherichia coli tyrT* gene transcription is sensitive to DNA supercoiling in its native chromosomal context: effect of DNA topoisomerase IV overexpression on *tyrT* promoter function. Mol. Microbiol. **14**:151-161.

135. **Friedman, D.I.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545-554.

136. **Fsihi, H., and S.T. Cole.** 1995. The *Mycobacterium leprae* genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel *polA* locus associated with genomic variability. Mol. Microbiol. **16:**909-919.

137. **Fuhrman, J.A., K. McCallum, and A.A. Davis.** 1992. Novel major archaebacterial group from marine plankton. Nature (London) **356:**148-149.

138. **Fujita, M.Q., H. Yoshikawa, and N. Ogasawara.** 1989. Structure of the *dnaA* region of *Pseudomonas putida*: conservation among three bacteria, *Bacillus subtilis*, *Escherichia coli* and *P. putida*.. Mol. Gen. Genet. **215:**381-387.

139. **Gäher, M., K. Einsiedler, T. Crass, and W. Bautsch.** 1996. A physical and genetic map of *Neisseria meningitidis* B1940. Mol. Microbiol. **19:**249-259.

140. **Galas, D.J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109-162. *In* D.E. Berg and M.M. Howe (ed.), *Mobile DNA*. American Society for Microbiology, Washington, D.C.

141. **Gamper, H.B., and J.E. Hearst.** 1982. A topological model for transcription based on unwinding angle analysis of *E. coli* RNA polymerase binary, initiation and ternary complexes. Cell **29:**81-90.

142. **Garrett, R.A., C. Aagaard, M. Andersen, J.Z. Dalgaard, J. Lykke-Andersen, H.T.N. Phan, S. Trevisanato, L. Østergaard, N. Larsen, and H. Leffers.** 1994. Archaeal rRNA operons, intron splicing and homing endonucleases, RNA polymerase operons and phylogeny. System. Appl. Microbiol. **16:**680-691.

143. **Gartenberg, M.R., and D.M. Crothers.** 1991. Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. J. Mol. Biol. **219:**217-230.

144. **Gauthier, A., M. Turmel, and C. Lemieux.** 1991. A group I intron in the chloroplast large subunit rRNA gene of *Chlamydomonas eugametos* encodes a double-strand endonuclease that cleaves the homing site of this intron. Curr. Genet. **19:**43-47.

145. **Gerl, L., and M. Sumper.** 1988. Halobacterial flagellins are encoded by a multigene family. J. Biol. Chem. **263:**13246-13251.

146. **Ghosal, D., and H. Saedler.** 1979. IS2-*61* and IS2-*611* arise by illegitimate recombination from IS2-*6*. Mol. Gen. Genet. **176:**233-238.

147. **Giometti, C.S., S.L. Tollaksen, S. Mukund, Z.H. Zhou, K. Ma, X. Mai, and M.W.W. Adams.** 1995. Two-dimensional gel electrophoresis mapping of proteins isolated from the hyperthermophile *Pyrococcus furiosus*. J. Chromatogr. A **698:**341-349.

148. **Glickman, B.W., and L.S. Ripley.** 1984. Structural intermediates of deletion mutagenesis: a role for palindromic DNA. Proc. Natl. Acad. Sci. USA **81**:512-516.

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

150. **Goldstein, E., and K. Drlica.** 1984. Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. Proc. Natl. Acad. Sci. USA **81**:4046-4050.

151. **Gonda, D.K., and C.M. Radding.** 1983. By searching processively recA protein pairs DNA molecules that share a limited stretch of homology. Cell **34**:647-654.

152. **Gonnet, G.H., M.A. Cohen, and S.A. Benner.** 1992. Exhaustive matching of the entire protein sequence database. Science **256**:1443-1445.

153. **Göransson, M., B. Sondén, P. Nilsson, B. Dagberg, K. Forsman, K. Emanuelsson, and E. Uhlin.** 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. Nature (London) **344**:682-685.

154. **Gouy, M., and C. Gautier.** 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**:7055-7074.

155. **Grant, W.D., and H. Larsen.** 1989. Group III. Extremely halophilic Archaeobacteria order Halobacteriales ord. nov., p. 2216-2219. *In* J.T. Staley, M.P. Bryant, N. Pfennig, and J.G. Holt (ed.), *Bergey's manual of systematic bacteriology, vol. 3*. The Williams & Wilkins Co., Baltimore.

156. **Grant, W.D., and H.N.M. Ross.** 1986. The ecology and taxonomy of halobacteria. FEMS Microbiol. Rev. **39**:9-15.

157. **Grau, R., D. Gardiol, G.C. Glikin, and D. de Mendoza.** 1994. DNA supercoiling and thermal regulation of unsaturated fatty acid synthesis in *Bacillus subtilis*. Mol. Microbiol. **11**:933-941.

158. **Grayling, R.A., K. Sandman, and J.N. Reeve.** 1994. Archaeal DNA binding proteins and chromosome structure. System. Appl. Microbiol. **16**:582-590.

159. **Green, P., D. Lipman, L. Hillier, R. Waterston, D. States, and J.-M. Claverie.** 1993. Ancient conserved regions in new gene sequences and the protein databases. Science **259**:1711-1716.

160. **Groisman, E.A., M.H. Saier Jr., and H. Ochman.** 1992. Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. EMBO J. **11**:1309-1316.

161. **Grothues, D., and B. Tümmler.** 1991. New approaches in genome analysis by pulsed-field gel electrophoresis: application to the analysis of *Pseudomonas* species. Mol. Microbiol. **5**:2763-2776.

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

163. **Gutiérrez, M.C., M.T. García, A. Ventosa, J.J. Nieto, and F. Ruiz-Berraquero.** 1986. Occurrence of megaplasmids in halobacteria. J. Appl. Bacteriol. **61**:67-71.

164. **Guttman, D.S., and D.E. Dykhuizen.** 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266**:1380-1383.

165. **Hackett, N.R., Y. Bobovnikova, and N. Heyrovska.** 1994. Conservation of chromosomal arrangement among three strains of the genetically unstable archaeon *Halobacterium salinarium*. J. Bacteriol. **176**:7711-7718.

166. **Hackett, N.R., M.P. Krebs, S. DasSarma, W. Goebel, U.L. RajBhandary, and H.G. Khorana.** 1990. Nucleotide sequence of a high copy number plasmid from *Halobacterium* strain GRB. Nucleic Acids Res. **18**:3408.

167. **Hall, L.M.C.** 1994. Are point mutations or DNA rearrangements responsible for the restriction fragment length polymorphisms that are used to type bacteria? Microbiology (Reading) **140**:197-204.

168. **Hall, R.M. and H.W. Stokes.** 1993. Integrons: novel DNA elements which capture genes by site-specific recombination. Genetica **90**:115-132.

169. **Hartke, A., S. Bouche, J.-M. Laplace, A. Benachour, P. Boutibonnes, and Y. Auffray.** 1995. UV-inducible proteins and UV-induced cross-protection against acid, ethanol, H_2O_2 or heat treatments in *Lactococcus lactis* subsp. *lactis*. Arch. Microbiol. **163**:329-336.

170. **Hase, T., W. Wakabayashi, H. Matsubara, L. Kercher, D. Oesterhelt, K.K. Rao, and D.O. Hall.** 1977. *Halobacterium halobium* ferredoxin: a homologous protein to chloroplast-type ferredoxins. FEBS Lett. **77**:308-310.

171. **Hendrickson, W.G., T. Kusano, H. Yamaki, R. Balakrishnan, M. King, J. Murchie, and M. Schaechter.** 1982. Binding of the origin of replication of *Escherichia coli* to the outer membrane. Cell **30**:915-923.

172. **Henikoff, S. and J.G. Henikoff.** 1991. Automated assembly of protein blocks for database searching. Nucleic Acids Res. **19**:6565-6572.

173. **Henner, D.J., and J.A. Hoch.** 1980. The *Bacillus subtilis* chromosome. Microbiol. Rev. **44**:57-82.

174. **Higgins, C.F., C.J. Dorman, and N. Ni Bhriain.** 1990. Environmental influences on DNA supercoiling: a novel mechanism for the regulation of gene expression, p. 421-432. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

175. **Higgins, C.F., C.J. Dorman, D.A. Stirling, L. Waddell, I.R. Booth, G. May, and E. Bremer.** 1988. A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. Cell **52**:569-584.

176. **Higgins, C.F., J.C.D. Hinton, C.S.J. Hulton, T. Owen-Hughes, G.D. Pavitt, and A. Seirafi.** 1990. Protein H1: a role for chromatin structure in the regulation of bacterial gene expression and virulence? Mol. Microbiol. **4**:2007-2012.

177. **Higgins, N.P., D.A. Collier, M.W. Kilpatrick, and H.M. Krause.** 1989. Supercoiling and integration host factor change the DNA conformation and alter the flow of convergent transcription in phage Mu. J. Biol. Chem. **264**:3035-3042.

178. **Hill, C.W., and J.A. Gray.** 1988. Effects of chromosomal inversion on cell fitness in *Escherichia coli* K-12. Genetics **119**:771-778.

179. **Hill, C.W., and B.W. Harnish.** 1981. Inversions between ribosomal RNA genes of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **78**:7069-7072.

180. **Hill, C.W., and B.W. Harnish.** 1982. Transposition of a chromosomal segment bounded by redundant rRNA genes into other rRNA genes in *Escherichia coli*. J. Bacteriol. **149**:449-457.

181. **Hinnebusch, J., and K. Tilly.** 1993. Linear plasmids and chromosomes in bacteria. Mol. Microbiol. **10**:917-922.

182. **Hobot, J.A., M.-A. Bjornsti, and E. Kellenberger.** 1987. Use of on-section immunolabeling and cryosubstitution for studies of bacterial DNA distribution. J. Bacteriol. **169**:2055-2062.

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

184. **Holloway, B.W., S. Dharmsthiti, V. Krishnapillai, A. Morgan, V. Obeyesekere, E. Ratnaningsih, M. Sinclair, D. Strom, and C. Zhang.** 1990. Patterns of gene linkages in *Pseudomonas* species, p. 97-105. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

185. **Holmes, M.L., and M.L. Dyall-Smith.** 1991. Mutations in DNA gyrase result in novobiocin resistance in halophilic archaebacteria. J. Bacteriol. **173**:642-648.

186. **Holmes, M.L., S.D. Nuttall, and M.L. Dyall-Smith.** 1991. Construction and use of halobacterial shuttle vectors and further studies on *Haloferax* DNA gyrase. J. Bacteriol. **173:**3807-3813.

187. **Honeycutt, R.J., M. McClelland, and B.W.S. Sobral.** 1993. Physical map of the genome of *Rhizobium meliloti* 1021. J. Bacteriol. **175:**6945-6952.

188. **Honoré, N., S. Bergh, S. Chanteau, F. Doucet-Populaire, K. Eiglmeier, T. Garnier, C. Georges, P. Launois, T. Limpaiboon, S. Newton, K. Niang, P. del Portillo, G.R. Ramesh, P. Reddi, P.R. Ridel, N. Sittisombut, S. Wu-Hunter, and S.T. Cole.** 1993. Nucleotide sequence of the first cosmid from the *Mycobacterium leprae* genome project: structure and function of the Rif-Str regions. Mol. Microbiol. **7:**207-214.

189. **Hopwood, D.A., H.M. Kieser, and T. Kieser.** 1993. The chromosome map of *Streptomyces coelicolor* A3(2), p. 497-504. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics*. American Society for Microbiology, Washington, D.C.

190. **Horne, M., C. Englert, F. Pfeifer.** 1988. Two genes encoding gas vacuole proteins in *Halobacterium halobium*. Mol. Gen. Genet. **213:**459-464.

191. **Horwitz, M.S.Z.** 1989. Transcription regulation *in vitro* by an *E. coli* promoter containing a DNA cruciform in the '-35' region. Nucleic Acids Res. **17:**5537-5545.

192. **Hsieh, L.-S., R.M. Burger, and K. Drlica.** 1991a. Bacterial DNA supercoiling and [ATP]/[ADP] changes associated with a transition to anaerobic growth. J. Mol. Biol. **219:**443-450.

193. **Hsieh, L.-S., J. Rouvière-Yaniv, and K. Drlica.** 1991b. Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. J. Bacteriol. **173:**3914-3917.

194. **Hsu, L.M., J.K. Giannini, T.W.C. Leung, and J.C. Crosthwaite.** 1991. Upstream sequence activation of *Escherichia coli argT* promoter in vivo and in vitro. Biochemistry **30:**813-822.

195. **Huber, L, and S. Selenska-Pobell.** 1994. Pulsed-field electrophoresis-fingerprinting, genome size estimation and rrn loci number of *Rhizobium galegae*. J. Appl. Bacteriol. **77:**528-533.

196. **Huber, R., M. Kurr, H.W. Jannasch, and K.O. Stetter.** 1989. A novel group of abyssal methanogenic archaebacteria (*Methanopyrus*) growing at 110°C. Nature (London) **342:**833-834.

197. **Hulton, C.S.J., A. Seirafi, J.C.D. Hinton, J.M. Sidebotham, L. Waddell, G.D. Pavitt, T. Owen-Hughes, A. Spassky, H. Buc, and C.F. Higgins.** 1990. Histone-like protein H1 (H-NS), DNA supercoiling, and gene expression in bacteria. Cell **63:**631-642.

198. **Ikeda, H., K. Aoki, and A. Naito.** 1982. Illegitimate recombination mediated *in vitro* by DNA gyrase of *Escherichia coli*: structure of recombinant DNA molecules. Proc. Natl. Acad. Sci. USA **79:**3724-3728.

199. **Ishiura, M., N. Hazumi, T. Koide, T. Uchida, and Y. Okada.** 1989. A *recB recC sbcB recJ* host prevent *recA*-independent deletions in recombinant cosmid DNA propagated in *Escherichia coli*. J. Bacteriol. **171:**1068-1074.

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

201. **Itaya, M.** 1993. Physical map of the *Bacillus subtilis* 168 chromosome, p. 463-471. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics*. American Society for Mcirobiology, Washington, D.C.

202. **Itoh, T.** 1988. Complete nucleotide sequence of the ribosomal 'A' protein operon from the archaebacterium, *Halobacterium halobium*. Eur. J. Biochem. **176:**297-303.

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

204. **Jaworski, A., N.P. Higgins, R.D. Wells, and W. Zacharias.** 1991. Topoisomerase mutants and physiological conditions control supercoiling and Z-DNA formation *in vivo*. J. Biol. Chem. **266:**2576-2581.

205. **Johansson, M.-L., M. Quednau, G. Molin, and S. Ahrné.** 1995a. Randomly amplified polymorphic DNA (RAPD) for rapid typing of *Lactobacillus plantarum* strains. Lett. Appl. Microbiol. **21:**155-159.

206. **Johansson, M.-L., M. Quednau, S. Ahrné, and G. Molin.** 1995b. Classification of *Lactobacillus plantarum* by restriction endonuclease analysis of total chromosomal DNA using conventional agarose gel electrophoresis. Intern. J. System. Bacteriol. **45:**670-675.

207. **Johnson, R.C., M.F. Bruist, and M.I. Simon.** 1986. Host protein requirements for in vitro site-specific DNA inversion. Cell **46:**531-539.

208. **Jones, W.J., J.A. Leigh, F. Mayer, C.R. Woese, and R.S. Wolfe.** 1983. *Methanococcus jannaschii* sp. nov., an extremely thermophilic methanogen from a submarine hydrothermal vent. Arch. Microbiol. **136:**254-261.

209. **Joshi, P. and P.P. Dennis.** 1993. Characterization of paralogous and orthologous members of the superoxide dismutase gene family from genera of the halophilic archaebacteria. J. Bacteriol. **175:**1561-1571.

210. **Joshi, J.G., W.R. Guild, and P. Handler.** 1963. The presence of two species of DNA in some halobacteria. J. Mol. Biol. **6:**34-38.

211. **Jovanovich, S.B., and J. Lebowitz.** 1987. Estimation of the effect of coumermycin A_1 on *Salmonella typhimurium* promoters by using random operon fusions. J. Bacteriol. **169:**4431-4435.

212. **Jumas-Bilak, E., C. Naugard, S. Michaux-Charachon, A. Allardet-Servent, A. Perrin, D. O'Callaghan, and M. Ramuz.** 1995. Study of the organization of the genomes of *Escherichia coli*, *Brucella melitensis* and *Agrobacterium tumefaciens* by insertion of a unique restriction site. Microbiology (Reading) **141:**2425-2432.

213. **Kagramanova, V.K., N.I. Derckacheva., and A.S. Mankin.** 1989. Unusual nucleotide sequence heterogeneity of small multicopy pHSB plasmid from *Halobacterium* strain SB3, an archaebacterium. Can. J. Microbiol. **35:**160-163.

214. **Kamekura, M., and M.L. Dyall-Smith.** 1995. Taxonomy of the family Halobacteriaceae and the description of two new genera *Halorubrobacterium* and *Natrialba*. J. Gen. Appl. Microbiol. **41:**333-350.

215. **Kamekura, M., and Y. Seno.** 1992. Nucleotide sequences of 16S rRNA encoding genes from halophilic archaea *Halococcus morrhuae* NRC16008 and *Haloferax mediterranei* ATCC33500. Nucleic Acids Res. **20:**3517.

216. **Kandler, O.** 1985. Evolution of the systematics of bacteria, p. 335-361. *In* K.H. Schleifer, and E. Stackebrandt (ed.), *Evolution of prokaryotes*. Academic Press, London.

217. **Kaneko, T., A. Tanaka, S. Sato, H. Kotani, T. Sazuka, N. Miyajima, M. Sugiura, and S. Tabata.** 1995a. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map position 64% to 92% of the genome. DNA Res. **2:**153-166.

218. **Kaneko, T., A. Tanaka, S. Sato, H. Kotani, T. Sazuka, N. Miyajima, M. Sugiura, and S. Tabata.** 1995b. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map position 64% to 92% of the genome (Supplement). DNA Res. **2:**191-198.

219. **Karem, K., and J.W. Foster.** 1993. The influence of DNA topology on the environmental regulation of a pH-regulated locus in *Salmonella typhimurium*. Mol. Microbiol. **10:**75-86.

220. **Karlin, S., G.M. Weinstock, and V. Brendel.** 1995. Bacterial classifications derived from RecA protein sequence comparisons. J. Bacteriol. **177:**6881-6893.

221. **Katayama, S.-I., B. Dupuy, T. Garnier, and S.T. Cole.** 1995. Rapid expansion of the physical and genetic map of the chromosome of *Clostridium perfringens* CPN50. J. Bacteriol. **177:**5680-5685.

222. **Kato, J.-I., Y. Hishimura, R. Imamura, H. Niki, S. Hiraga, and H. Suzuki.** 1990. New topoisomerase essential for chromosome segregation in *E. coli*. Cell **63:**393-404.

223. **Kato, J.-I., H. Suzuki, and H. Ikeda.** 1992. Purification and characterization of DNA topoisomerase IV in *Escherichia coli*. J. Biol. Chem. **267:**25676-25684.

224. **Kececioglu, J., and D. Sankoff.** 1995. Exact and approximation algorithms for sorting by reversals, with application to genome rearrangement. Algorithmica **13:**180-210.

225. **Kellenberger, E.** 1990. Intracellular organization of the bacterial genome, p. 173-185. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

226. **Kellenberger, E., and B. Arnold-Schulz-Gahmen.** 1992. Chromatins of low-protein content: special feature of their compaction and condensation. FEMS Microbiol. Lett. **100:**361-370.

227. **Kessel, M., and F. Klink.** 1980. Archaebacterial elongation factor is ADP-ribosylated by diphtheria toxin. Nature (London) **287:**250-251.

228. **Khasanov, F.K., D.J. Zvingila, A.A. Zainullin, A.A. Prozorov, and V.I. Bashkirov.** 1992. Homologous recombination between plasmid and chromosomal DNA in *Bacillus subtilis* requires approximately 70 bp of homology. Mol. Gen. Genet. **234:**494-497.

229. **Kikuchi, A., and K. Asai.** 1984. Reverse gyrase—a topoisomerase which introduces positive superhelical turns into DNA. Nature (London) **309:**677-681.

230. **Kimura, M., E. Arndt, T. Hatakeyama, T. Hatakeyama, and J. Kimura.** 1989. Ribosomal proteins of halobacteria. Can. J. Microbiol. **35:**195-199.

231. **King, S.R., M.A. Krolewski, S.L. Marvo, P.J. Lipson, K.L. Pogue-Geile, J.H. Chung, and S.R. Jaskunas.** 1982. Nucleotide sequence analysis of in vivo recombinants between bacteriophage lambda DNA and pBR322. Mol. Gen. Genet. **186:**548-557.

232. **Kohara, Y., K. Akiyama, and K. Isono.** 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.

233. **Koo, H.-S., H.-Y. Wu, and L.F. Liu.** 1990. Effects of transcription and translation on gyrase-mediated DNA cleavage in *Escherichia coli.* J. Biol. Chem. **265:**12300-12305.

234. **Koonin, E.V., R.L. Tatusov, and K.E. Rudd.** 1995. Sequence similarity analysis of *Escherichia coli* proteins: functional and evolutionary implications. Proc. Natl. Acad. Sci. USA **92:**11921-11925.

235. **Kovalsky, O.I., S.A. Kozyavkin, and A.I. Slesarev.** 1990. Archaebacterial reverse gyrase cleavage-site specificity is similar to that of eubacterial DNA topoisomerase I. Nucleic Acids Res. **18:**2801-2805.

236. **Krawiec, S.** 1985. Concept of a bacterial species. Int. J. Syst. Bacteriol. **35:**217-220.

237. **Krawiec, S., and M. Riley.** 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54:**502-539.

238. **Krueger, C.M., K.L. Marks, and G.M. Ihler.** 1995. Physical map of the *Bartonella bacilliformis* genome. J. Bacteriol. **177:**7271-7274.

239. **Krug, P.J., A.Z. Gileski, R.J. Code, A. Torjussen, and M.B. Schmid.** 1994. Endpoint bias in large Tn*10*-catalyzed inversions in *Salmonella typhimurium.* Genetics **136:**747-756.

240. **Kunisawa, T.** 1995. Identification and chromosomal distribution of DNA sequence segments conserved since divergence of *Escherichia coli* and *Bacillus subtilis.* J. Mol. Evol. **40:**585-593.

241. **Kunst, F., A. Vassarotti, and A. Danchin.** 1995. Organization of the European *Bacillus subtilis* genome sequencing project. Microbiology (Reading) **141:**249-255.

242. **Kuspa, A., D. Vollrath, Y. Cheng, and D. Kaiser.** 1989. Physical mapping of the *Myxococcus xanthus* genome by random cloning in yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA **86:**8917-8921.

243. **Labedan, B., and M. Riley.** 1995a. Widespread protein sequence similarities: origins of *Escherichia coli* genes. J. Bacteriol. **177:**1585-1588.

244. **Labedan, B., and M. Riley.** 1995b. Gene products of *Escherichia coli*: sequence comparisons and common ancestries. Mol. Biol. Evol. **12:**980-987.

245. **Ladefoged, S.A., and G. Christiansen.** 1992. Physical and genetic mapping of the genomes of five *Mycoplasma hominis* strains by pulsed-field gel electrophoresis. J. Bacteriol. **174:**2199-2207.

246. **Lake, J.A.** 1989. Origin of the eukaryotic nucleus: eukaryotes and eocytes are genotypically related. Can. J. Microbiol. **35:**109-118.

247. **Lake, J.A., M.W. Clark, E. Henderson, S.P. Fay, M. Oakes, A. Scheinman, J.P. Thornber, and R.A. Mah.** 1985. Eubacteria, halobacteria, and the origin of photosynthesis: the photocytes. Proc. Natl. Acad. Sci. USA **82:**3716-3720.

248. **Lam, W.L., A. Cohen, D. Tsouluhas, and W.F. Doolittle.** 1990. Genes for tryptophan biosynthesis in the archaebacterium *Haloferax volcanii*. Proc. Natl. Acad. Sci. USA **87:**6614-6618.

249. **Lam, W.L., and W.F. Doolittle.** 1989. Shuttle vectors for the archaebacterium *Halobacterium volcanii*. Proc. Natl. Acad. Sci. USA **86:**5478-5482.

250. **Lam, W.L., S.M. Logan, and W.F. Doolittle.** 1992. Genes for tryptophan biosynthesis in the halophilic archaebacterium *Haloferax volcanii*: the *trpDFEG* cluster. J. Bacteriol. **174:**1694-1697.

251. **Lam, W.L. and W.F. Doolittle.** 1992. Mevinolin-resistant mutations identify a promoter and the gene for a eukaryote-like 3-hydroxy-3-methylglutaryl-coenzyme A reductase in the archaebacterium *Haloferax volcanii*. J. Biol. Chem. **267:**5829-5834.

252. **Lander, E.S., and M.S. Waterman.** 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. Genomics **2:**231-239.

253. **Langer, D., J. Hain, P. Thuriaux, and W. Zillig.** 1995. Transcription in Archaea: similarity to that in Eucarya. Proc. Natl. Acad. Sci. USA **92:**5768-5772.

254. **Langworthy, T.A.** 1985. Lipids of archaebacteria, p. 459-498. *In* C.R. Woese, and R.S. Wolfe (ed.), *The bacteria vol. 8. Archaebacteria*. Academic Press, Inc., New York.

255. **Larsen, H., and W.D. Grant.** 1989. Genus I. Halobacterium, p. 2219-2224. *In* J.T. Staley, M.P. Bryant, N. Pfennig, and J.G. Holt (ed.), *Bergey's manual of systematic bacteriology, vol. 3*. The Williams & Wilkins Co., Baltimore.

256. **Laurent-Winter, C., P. Lejeune, and A. Danchin.** 1995. The *Escherichia coli* DNA-binding protein H-NS is one of the first proteins to be synthesized after a nutritional upshift. Res. Microbiol. **146:**5-16.

257. **Le Bourgeois, P., M. Lautier, M. Mata, and P. Ritzenthaler.** 1992. Physical and genetic map of the chromosome of *Lactococcus lactic* subsp. *lactis* IL1403. J. Bacteriol. **174:**6752-6762.

258. **Le Bourgeois, P., M. Lautier, L. van den Berghe, M.J. Gasson, and P. Ritzenthaler.** 1995. Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: Comparison with that of *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. J. Bacteriol. **177:**2840-2850.

259. **Leblond, P., P. Demuyter, J.-M. Simonet and B. Decaris.** 1991. Genetic instability and associated genome plasticity in *Streptomyces ambofaciens*: pulsed-field gel electrophoresis evidence for large DNA alterations in a limited genomic region. J. Bacteriol. **173**:4229-4233.

260. **Leblond, P., G. Fischer, F.-X. Francou, F. Berger, M. Guérineau, and B. Decaris.** 1996. The unstable region of *Streptomyces ambofaciens* includes 210 kb terminal inverted repeats flanking the extremities of the linear chromosomal DNA. Mol. Microbiol. **19**:261-271.

261. **Leblond, P., M. Redenbach, and J. Cullum.** 1993. Physical map of the *Streptomyces lividans* 66 genome and comparison with that of the related strain *Streptomyces coelicolor* A3(2). J. Bacteriol. **175**:3422-3429.

262. **Lechner, J. and M. Sumper.** 1987. The primary structure of a procaryotic glycoprotein cloning and sequencing of the cell surface glycoprotein gene of halobacteria. J. Biol. Chem. **262**:9724-9729.

263. **Leffers, H., F. Gropp, F. Lottspeich, W. Zillig, and R.A. Garrett.** 1989. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. J. Mol. Biol. **206**:1-17.

264. **Lejeune, P., and A. Danchin.** 1990. Mutations in the *bglY* gene increase the frequency of spontaneous deletions in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **87**:360-363.

265. **Lezhava, A., T. Mizukami, T. Kajitani, D. Kameoka, M. Redenbach, H. Shinkawa, O. Nimi, and H. Kinashi.** 1995. Physical map of the linear chromosome of *Streptomyces griseus*. J. Bacteriol. **177**:6492-6498.

266. **Lilley, D.M.J., and C.F. Higgins.** 1991. Local DNA topology and gene expression: the case of the *leu-500* promoter. Mol. Microbiol. **5**:779-783.

267. **Liu, B., M.L. Wong, R.L. Tinker, E.P. Geiduschek, and B.M. Alberts.** 1993. The DNA replication fork can pass RNA polymerase without displacing the nascent transcript. Nature (London) **366**:33-39.

268. **Liu, L.F., and J.C. Wang.** 1987. Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. USA **84**:7024-7027.

269. **Liu, S.-L., A. Hessel, H.-Y.M. Cheng, and K.E. Sanderson.** 1994. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella paratyphi* B. J. Bacteriol. **176**:1014-1024.

270. **Liu, S.-L., A. Hessel, and K.E. Sanderson.** 1993a. Genomic mapping with I-Ceu I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. Proc. Natl. Acad. Sci. USA **90**:6874-6878.

271. **Liu, S.-L., A. Hessel, and K.E. Sanderson.** 1993b. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella typhimurium* LT2 determined by double digestion, end labelling, and pulsed-field gel electrophoresis. J. Bacteriol. **175:**4104-4120.

272. **Liu, S.-L., A. Hessel, and K.E. Sanderson.** 1993c. The *Xba*I-*Bln*I-*Ceu*I genomic cleavage map of *Salmonella enteritidis* shows an inversion relative to *Salmonella typhimurium* LT2. Mol. Microbiol. **10:**655-664.

273. **Liu, S.-L., and K.E. Sanderson.** 1992. A physical map of the *Salmonella typhimurium* LT2 genome made by using *Xba*I analysis. J. Bacteriol. **174:**1662-1672.

274. **Liu, S.-L., and K.E. Sanderson.** 1995a. Rearrangements in the genome of the bacterium *Salmonella typhi*. Proc. Natl. Acad. Sci. USA **92:**1018-1022.

275. **Liu, S.-L., and K.E. Sanderson.** 1995b. I-*Ceu*I reveals conservation of the genome of independent strains of *Salmonella typhimurium*. J. Bacteriol. **177:**3355-3357.

276. **Liu, S.-L., and K.E. Sanderson.** 1995c. Genomic cleavage map of *Salmonella typhi* Ty2. J. Bacteriol. **177:**5099-5107.

277. **Liu, S.-L., and K.E. Sanderson.** 1995d. The chromosome of *Salmonella paratyphi* A is inverted by recombination between *rrnH* and *rrnG*. J. Bacteriol. **177:**6585-6592.

278. **Lodge, J.K., T. Kazic, and D.E. Berg.** 1989. Formation of supercoiling domains in plasmid pBR322. J. Bacteriol. **171:**2181-2187.

279. **López-García, P., J.P. Abad, and R. Amils.** 1993. Genome analysis of different *Haloferax mediterranei* strains using pulsed-field gel electrophoresis. System. Appl. Microbiol. **16:**310-321.

280. **López-García, P., J.P. Abad, C. Smith, and R. Amils.** 1992. Genomic organization of the halophilic archaeon *Haloferax mediterranei*: physical map of the chromosome. Nucleic Acids Res. **20:**2459-2464.

281. **López-García, P., A. St. Jean, R. Amils, and R.L. Charlebois.** 1995. Genomic stability in the archaea *Haloferax volcanii* and *Haloferax mediterranei*. J. Bacteriol. **177:**1405-1408.

282. **Louarn, J.M., J.P. Bouché, F. Legendre, J. Louarn, and J. Patte.** 1985. Characterization and properties of very large inversions of the *E. coli* chromosome along the origin-to-terminus axis. Mol. Gen. Genet. **201:**467-476.

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

284. **Lück, P.C., J.H. Helbig, V. Drašar, N. Bornstein, R.J. Fallon, and M. Castellani-Pastoris.** 1995. Genomic heterogenicity amongst phenotypically similar *Legionella micdadei* strains. FEMS Microbiol. Lett. **126:**49-54.

285. **Lurz, R., M. Grote, J. Dijk, R. Reinhardt, and B. Dobrinski.** 1986. Electron microscopic study of DNA complexes with proteins from the Archaebacterium *Sulfolobus acidocaldarius*. EMBO J. **5:**3715-3721.

286. **Lynch, A.S., and J.C. Wang.** 1993. Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I. J. Bacteriol. **175:**1645-1655.

287. **Ma, D., D.N. Cook, N.G. Pon, and J.E. Hearst.** 1994. Efficient anchoring of RNA polymerase in *Escherichia coli* during coupled transcription-translation of genes encoding integral inner membrane polypeptides. J. Biol. Chem. **269:**15362-15370.

288. **Magrum, L.J., K.R. Luehrsen, and C.R. Woese.** 1978. Are extreme halophiles actually "bacteria"? J. Mol. Evol. **11:**1-8.

289. **Mahajan, S.K.** 1988. Pathways of homologous recombination in *Escherichia coli*, p. 87-140. *In* R. Kucherlapati, and G.R. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.

290. **Mahajan, S.K., N.N. Pandit, and J.F. Sarkari.** 1984. Host functions in amplification and deamplification of Tn9 in *Escherichia coli* K-12: a new model for amplification. Cold Spring Harbor Symp. Quant. Biol. **49:**443-451.

291. **Mahan, M.J., A.M. Segall, and J.R. Roth.** 1990. Recombination events that rearrange the chromosome: barriers to inversion, p. 341-349. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

292. **Majumder, R., S. Sengupta, G. Khetawat, R.K. Bhadra, S. Roychoudhury, and J. Das.** 1996. Physical map of the genome of *Vibrio cholerae* 569B and localization of genetic markers. J. Bacteriol. **178:**1105-1112.

293. **Mankin, A.S., V.K. Kagramanova, N.L. Teterina, P.M. Rubtsov, E.N. Belova, A.M. Kopylov, L.A. Baratova, and A.A. Bogdanov.** 1985. The nucleotide sequence of the gene coding for the 16S rRNA from the archaebacterium *Halobacterium halobium*. *Gene* **37:**181-189.

294. **Marais, A., J.M. Bové, and J. Renaudin.** 1996. *Spiroplasma citri* virus SpV1-derived cloning vector: deletion formation by illegitimate and homologous recombination in a spiroplasmal host strain which probably lacks a functional *recA* gene. J. Bacteriol. **178:**862-870.

295. **Marsh, T.L., C.I. Reich, R.B. Whitlock, and G.J. Olsen.** 1994. Transcription factor IID in the Archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. Proc. Natl. Acad. Sci. USA **91**:4180-4184.

296. **Marshall, P., and C. Lemieux.** 1991. Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast large subunit rRNA-encoding gene of *Chlamydomonas eugametos*. Gene **104**:241-245.

297. **Martínez-Murcia, A.J., I.F. Boán, and F. Rodríguez-Valera.** 1995. Evaluation of the authenticity of haloarchaeal strains by random-amplified polymorphic DNA. Lett. Appl. Microbiol. **21**:106-108.

298. **Marvo, S.L., S.R. King, and S.R. Jaskunas.** 1983. Role of short regions of homology in intermolecular illegitimate recombination events. Proc. Natl. Acad. Sci. USA **80**:2452-2456.

299. **Masters, M., P.D. Moir, R. Spiegelberg, J.H. Pringle, and C.W. Vermeulen.** 1985. Is the chromosome of *E. coli* differentiated along its length with respect to gene density or accessibility to transcription?, p. 335-343. *In* M. Schaechter, F. Neidhardt, J. Ingraham, and N. Kjelgaard (ed.), *The molecular biology of bacterial growth*. Jones & Bartlett, Boston.

300. **May, B.P., P. Tam, and P.P. Dennis.** 1989. The expression of the superoxide dismutase gene in *Halobacterium cutirubrum* and *Halobacterium volcanii*. Can. J. Microbiol. **35**:171-175.

301. **Maynard Smith, J., C.G. Dowson, and B.G. Spratt.** 1991. Localized sex in bacteria. Nature (London) **349**:29-31.

302. **Maynard Smith, J., N.H. Smith, M. O'Rourke, and B.G. Spratt.** 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90**:4384-4388.

303. **McAllister, C.F., and E.C. Achberger.** 1988. Effect of polyadenine-containing curved DNA on promoter utilization in *Bacillus subtilis*. J. Biol. Chem. **263**:11743-11749.

304. **McAllister, C.F., and E.C. Achberger.** 1989. Rotational orientation of upstream curved DNA affects promoter function in *Bacillus subtilis*. J. Biol. Chem. **264**:10451-10456.

305. **McCloskey, J.A.** 1986. Nucleoside modification in archaebacterial transfer RNA. System. Appl. Microbiol. **7**:246-252.

306. **McGenity, T.J., and W.D. Grant.** 1995. Transfer of *Halobacterium saccharovorum*, *Halobacterium sodomense*, *Halobacterium trapanicum* NRC 34021 and *Halobacterium lacusprofundi* to the genus *Halorubrum* gen. nov., as *Halorubrum saccharovorum* comb. nov., *Halorubrum sodomense* comb. nov., *Halorubrum trapanicum* comb. nov., and *Halorubrum lacusprofundi* comb. nov. System. Appl. Microbiol. **18**:237-243.

307. **McNairn, E., N. Ní Bhriain, and C.J. Dorman.** 1995. Overexpression of the *Shigella flexneri* genes coding for DNA topoisomerase IV compensates for loss of DNA topoisomerase I: effect on virulence gene expression. Mol. Microbiol. **15**:507-517.

308. **Médigue, C., I. Moszer, A. Viari, and A. Danchin.** 1995. Analysis of a *Bacillus subtilis* genome fragment using a co-operative computer system prototype. Gene **165**:GC37-GC51.

309. **Médigue, C., T. Rouxel, P. Vigier, A. Hénaut, and A. Danchin.** 1991. Evidence for horizontal gene transfer in *Escherichia coli* speciation. J. Mol. Biol. **222**:851-856.

310. **Médigue, C., A. Viari, A. Hénaut, and A. Danchin.** 1993. Colibri: a functional data base for the *Escherichia coli* genome. Microbiol. Rev. **57**:623-654.

311. **Ménard, C., and C. Mouton.** 1995. Clonal diversity of the taxon *Porphyromonas gingivalis* assessed by random amplified polymorphic DNA fingerprinting. Infect. Immun. **63**:2522-2531.

312. **Méndez-Alvarez, S., V. Pavón, I. Esteve, R. Guerrero, and N. Gaju.** 1995. Genomic heterogeneity in *Chlorobium limicola*: chromosomic and plasmidic differences among strains. FEMS Microbiol. Lett. **134**:279-285.

313. **Menzel, R., and M. Gellert.** 1983. Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. Cell **34**:105-113.

314. **Menzel, R., and M. Gellert.** 1987. Fusions of the *Escherichia coli gyrA* and *gyrB* control regions to the galactokinase gene are inducible by coumermycin treatment. J. Bacteriol. **169**:1272-1278

315. **Messing, J.** 1983. New M13 vectores for cloning. Methods Enzymol. **101**:20-78.

316. **Mevarech, M., and R. Werczberger.** 1985. Genetic transfer in *Halobacterium volcanii*. J. Bacteriol. **162**:461-462.

317. **Milkman, R., and M. McKane.** 1995. DNA sequence variation and recombination in *E. coli*, p. 126-142. *In* S. Baumberg, J.P.W. Young, S.R. Saunders, and E.M.H. Willington (ed.), *Society for general microbiology symposium 52 population genetics of bacteria.* Cambridge University Press, Cambridge.

318. **Milkman, R., and M. McKane Bridges.** 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. Genetics **126:**505-517.

319. **Milkman, R., and A. Stoltzfus.** 1988. Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. Genetics **120:**359-366.

320. **Miller, W.G., and R.W. Simons.** 1993. Chromosomal supercoiling in *Escherichia coli*. Mol. Microbiol. **10:**675-684.

321. **Mishra, R.K., and D. Chatterji.** 1993. Mechanism of initiation of transcription by *Escherichia coli* RNA polymerase on supercoiled template. Mol. Microbiol. **8:**507-515.

322. **Mojica, F.J.M., C. Ferrer, G. Juez, and F. Rodríguez-Valera.** 1995. Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. Mol. Microbiol. **17:**85-93.

323. **Moore, R.L. and B.J. McCarthy.** 1969. Characterization of the deoxyribonucleic acid of various strains of halophilic bacteria. J. Bacteriol. **99:**248-254.

324. **Moszer, I., P. Glaser, and A. Danchin.** 1995. *SubtiList*: a relational database for the *Bacillus subtilis* genome. Microbiology (Reading) **141:**261-268.

325. **Mullakhanbhai, M.F., and H. Larsen.** 1975. *Halobacterium volcanii* spec. nov., a Dead Sea halobacterium with a moderate salt requirement. Arch. Microbiol. **104:**207-214.

326. **Murray, R.G.E.** 1968. Microbial structure as an aid to microbial classification and taxonomy. Spisy (Faculte des Sciences de l'Universite J. E. Purkyne, Brno) **43:**249-252.

327. **Naas, T., M. Blot, W.M. Fitch, and W. Arber.** 1994. Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. Genetics **136:**721-730.

328. **Nadal, M., G. Mirambeau, P. Forterre, W.D. Reiter, and M. Duguet.** 1986. Positively supercoiled DNA in a virus like particle of an archaebacterium. Nature (London) **321:**256-258.

329. **Nelson, K., and R.K. Selander.** 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. **174:**6886-6895.

330. **Nelson, K., T.S. Whittam, and R.K. Selander.** 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88:**6667-6671.

331. **Ng, W.-L. and S. DasSarma.** 1993. Minimal replication origin of the 200-kilobase *Halobacterium* plasmid pNRC100. J. Bacteriol. **175:**4584-4596.

332. **Ní Bhriain, N., C.J. Dorman, and C.F. Higgins.** 1989. An overlap between osmotic and anaerobic stress responses: a potential role for DNA supercoiling in the coordinate regulation of gene expression. Mol. Microbiol. **3:**933-942.

333. **Nickerson, C.A., and E.C. Achberger.** 1995. Role of curved DNA in binding of *Escherichia coli* RNA polymerase to promoters. J. Bacteriol. **177:**5756-5761.

334. **Nierlich, D.P.** 1996. Future directions for biomolecular databases. ASM News **62:**251-254.

335. **Nieuwlandt, D.T., E.S. Haas, and C.J. Daniels.** 1991. The RNA component of RNase P from the archaebacterium *Haloferax volcanii*. J. Biol. Chem. **266:**5689-5695.

336. **Nikolaichik, E.A., and A.G. Pesnyakevich.** 1995. A circular genetic map of *Erwinia carotovora* subsp. *atroseptica* 3-2. Genetika **31:**1052-1058.

337. **Nikolskaya, T., M. Fonstein, and R. Haselkorn.** 1995. Alignment of a 1.2-Mb chromosomal region from three strains of *Rhodobacter capsulatus* reveals a significantly mosaic structure. Proc. Natl. Acad. Sci. USA **92:**10609-10613.

338. **Norton, C.F.** 1992. Rediscovering the ecology of halobacteria. ASM News **58:**363-367.

339. **O'Byrne, C.P., N. Ní Bhriain, and C.J. Dorman.** 1992. The DNA supercoiling-sensitive expression of the *Salmonella typhimurium his* operon requires the *his* attenuator and is modulated by anaerobiosis and by osmolarity. Mol. Microbiol. **6:**2467-2476.

340. **O'Byrne, C.P., and C.J. Dorman.** 1994. Transcription of the *Salmonella typhimurium spv* virulence locus is regulated negatively by the nucleoid-associated protein H-NS. FEMS Microbiol. Lett. **121:**99-106.

341. **Ochman, H., and A.C. Wilson.** 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J. Mol. Evol. **26:**74-86.

342. **Ogasawara, N., Y. Fujita, Y. Kobayashi, Y. Sadaie, T. Tanaka, H. Takahashi, K. Yamane, and H. Yoshikawa.** 1995. Systematic sequencing of the *Bacillus subtilis* genome: progress report of the Japanese group. Microbiology (Reading) **141:**257-259.

343. **Ojaimi, C., B.E. Davidson, I. Saint Girons, and I.G. Old.** 1994. Conservation of gene arrangement and an unusual organization of rRNA genes in the linear chromosomes of the Lyme disease spirochaetes *Borrelia burgdorferi*, *B. garinii* and *B. afzelii*. Microbiology (Reading) **140:**2931-2940.

344. **Oppenheim, A.B., K.E. Rudd, I. Mendelson, and D. Teff.** 1993. Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli.* Mol. Microbiol. **10:**113-122.

345. **Oren, A.** 1991. Estimation of the contribution of Archaebacteria and Eubacteria to the bacterial biomass and activity in hypersaline ecosystems: novel approaches, p. 25-31. *In* F. Rodriguez-Valera (ed.), *General and applied aspects of halophilic microorganisms, NATO ASI Series A, vol. 201.* Plenum Press, New York.

346. **Oren, A., P. Gurevich, R.T. Gemmell, and A. Teske.** 1995. *Halobaculum gomorrense* gen. nov., sp. nov., a novel extremely halophilic archaeon from the Dead Sea. Intern. J. System. Bacteriol. **45:**747-754.

347. **Ornston, L.N., E.L. Neidle, and J.E. Houghton.** 1990. Gene rearrangements, a force for evolutionary change; DNA sequence arrangements, a source of genetic constancy, p. 325-334. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome.* American Society for Microbiology, Washington, D.C.

348. **Ouzounis, C., N. Kyrpides, and C. Sander.** 1995. Novel protein families in archaean genomes. Nucleic Acids Res. **23:**565-570.

349. **Owen-Hughes, T.A., G.D. Pavitt, D.S. Santos, J.M. Sidebotham, C.S.J. Hulton, J.C.D. Hinton, and C.F. Higgins.** 1992. The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. Cell **71:**255-265.

350. **Painbéni, E., E. Mouray, S. Gottesman, and J. Rouvière-Yaniv.** 1993. An imbalance of HU synthesis induces mucoidy in *Escherichia coli.* J. Mol. Biol. **234:**1021-1037.

351. **Pattee, P.A.** 1993. The genetic map of *Staphylococcus aureus*, p. 489-496. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics.* American Society for Mcirobiology, Washington, D.C.

352. **Pavitt, G.D., and C.F. Higgins.** 1993. Chromosomal domains of supercoiling in *Salmonella typhimurium*. Mol. Microbiol. **10:**685-696.

353. **Pérez-Martin, J., F. Rojo, and V. Lorenzo.** 1994. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol. Rev. **58:**268-290.

354. **Peterson, S.N., T. Lucier, K. Heitzman, E.A. Smith, K.F. Bott, P.-C. Hu, and C.A. Hutchison III.** 1995. Genetic map of the *Mycoplasma genitalium* chromosome. J. Bacteriol. **177:**3199-3204.

355. **Pettijohn, D.E.** 1988. Histone-like proteins and bacterial chromosome structure. J. Biol. Chem. **263:**12793-12796.

356. **Pettijohn, D.E., and Y. Hodges-Garcia.** 1990. Role of HU protein in transitional DNA coiling, p. 241-245. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

357. **Pettijohn, D.E., and O. Pfenninger.** 1980. Supercoils in prokaryotic DNA restrained *in vivo*. Proc. Natl. Acad. Sci. USA **77**:1331-1335.

358. **Pfeifer, F.** 1988. Genetics of halobacteria, p. 105-133. *In* F. Rodriguez-Valera (ed.), *Halophilic bacteria, vol. II*. CRC Press Inc., Boca Raton.

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

360. **Pfeifer, F., and U. Blaseio.** 1990. Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. Nucleic Acids Res. **18**:6921-6925.

361. **Pfeifer, F., K. Ebert, G. Weidinger, and W. Goebel.** 1982. Structure and functions of chromosomal and extrachromosomal DNA in halobacteria. Zentralbl. Bakteriol. Hyg. Abt. 1 Orig. C **3**:110-119.

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

363. **Pfeifer, F., M. Betlach, R. Martienssen, J. Friedman, and H.W. Boyer.** 1983. Transposable elements of *Halobacterium halobium*. Mol. Gen. Genet. **191**:182-188.

364. **Pfeifer, F., J. Griffig, and D. Oesterhelt.** 1993. The *fdx* gene encoding the [2Fe-2S] ferredoxin of *Halobacterium salinarium* (*H. halobium*). Mol. Gen. Genet. **239**:66-71.

365. **Pfeifer, F., G. Weidinger, and W. Goebel.** 1981a. Characterization of plasmids in halobacteria. J. Bacteriol. **145**:369-374.

366. **Pfeifer, F., G. Weidinger, and W. Goebel.** 1981b. Genetic variability in *Halobacterium halobium*. J. Bacteriol. **145**:375-381.

367. **Plaga, W., F. Lottspeich, and D. Oesterhelt.** 1992. Improved purification, crystallization and primary structure of pyruvate:ferredoxin oxidoreductase from *Halobacterium halobium*. Eur. J. Biochem. **205**:391-397.

368. **Pontiggia, A., A. Negri, M. Beltrame, and M.E. Bianchi.** 1993. Protein HU binds specifically to kinked DNA. Mol. Microbiol. **7**:343-350.

369. **Prozorov, A.A.** 1995. Genome structure of bacteria: uniformity or diversity? Genetika **31**:741-752.

370. **Pruss, G.J., and K. Drlica.** 1989. DNA supercoiling and prokaryotic transcription. Cell **56**:521-523.

371. **Pruss, G.J., S.H. Manes, and K. Drlica.** 1982. *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. Cell **31**:35-42.

372. **Pyle, L.E., T. Taylor, and L.R. Finch.** 1990. Genomic maps of some strains within the *Mycoplasma mycoides* cluster. J. Bacteriol. **172**:7265-7268.

373. **Rabb, F.T., S.R. Brummet, A. Bogert, K. Hujer, J. Krall, S. Domke, J. Szasz, K.M. Borges, M. Davis, C. Fuller, and J.W. Chase.** 1995. 'Eukaryotic' gene functions in the hyperthermophilic archaeon, *Pyrococcus furiosus*. Gen. Sci. Technol. **1**:P-46.

374. **Radding, C.M.** 1988. Homologous pairing and strand exchange promoted by *Escherichia coli* RecA protein, p. 193-229. *In* R. Kucherlapati, and G.R. Smith (ed.), *Genetic recombination*. American Society for Microbiology, Washington, D.C.

375. **Rahmouni, A.R., and R.D. Wells.** 1989. Stabilization of Z DNA in vivo by localized supercoiling. Science **246**:358-363.

376. **Rahmouni, A.R., and R.D. Wells.** 1992. Direct evidence for the effect of transcription on local DNA supercoiling *in vivo*. J. Mol. Biol. **223**:131-144.

377. **Rainey, P.B. and M.J. Bailey.** 1996. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. Mol. Microbiol. **19**:521-533.

378. **Ramirez, R.M., and M. Villarejo.** 1991. Osmotic signal transduction to *proU* is independent of DNA supercoiling in *Escherichia coli*. J. Bacteriol. **173**:879-885.

379. **Rebollo, J.-E., V. François, and J.-M. Louarn.** 1988. Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. Proc. Natl. Acad. Sci. USA **85**:9391-9395.

380. **Redenbach, M., H.M. Kieser, D. Denapaite, A. Eichner, J. Cullum, H. Kinashi, and D.A. Hopwood.** 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. Mol. Microbiol. **21**:77-96.

381. **Révet, B., S. Brahms, and G. Brahms.** 1995. Binding of the transcription activator NR_I (NTRC) to a supercoiled DNA segment imitates association with the natural enhancer: an electron microscopic investigation. Proc. Natl. Acad. Sci. USA **92**:7535-7539.

382. **Richardson, S.M.H., C.F. Higgins, and D.M.J.Lilley.** 1984. The genetic control of DNA supercoiling in *Salmonella typhimurium*. EMBO J. **3**:1745-1752.

383. **Richardson, S.M.H., C.F. Higgins, and D.M.J. Lilley.** 1988. DNA supercoiling and the *leu-500* promoter mutation of *Salmonella typhimurium*. EMBO J. 7:1863-1869.

384. **Riley, M.** 1993. Functions of the gene products of *Escherichia coli*. Microbiol. Rev. 57:862-952.

385. **Riley, M., and A. Anilionis.** 1978. Evolution of the bacterial genome. Annu. Rev. Microbiol. 32:519-560.

386. **Riley, M., and K.E. Sanderson.** 1990. Comparative genetics of *Escherichia coli* and *Salmonella typhimurium*, p. 85-95. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

387. **Riley, M., and S. Krawiec.** 1987. Genome organization, p. 967-981. *In* F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), Escherichia coli *and* Salmonella typhimurium *cellular and molecular biology vol. 2*. American Society for Microbiology, Washington, D.C.

388. **Rivera, M.C., and J.A. Lake.** 1996. The phylogeny of *Methanopyrus kandleri*. Intern. J. System. Bacteriol. 46:348-351.

389. **Robinow, C., and E. Kellenberger.** 1994. The bacterial nucleoid revisited. Microbiol. Rev. 58:211-232.

390. **Rode, C.K., V.H. Obreque, and C.A. Bloch.** 1995. New tools for integrated genetic and physical analyses of the *Escherichia coli* chromosome. Gene 166:1-9.

391. **Rodley, P.D., U. Römling, and B. Tümmler.** 1995. A physical genome map of the *Burkholderia cepacia* type strain. Mol. Microbiol. 17:57-67.

392. **Rodriguez-Valera, F., G. Juez, and D.J. Kushner.** 1983. *Halobacterium mediterranei* spec. nov., a new carbohydrate-utilizing extreme halophile. System. Appl. Microbiol. 4:369-381.

393. **Rohde, J.R., J.M. Fox, and S.A. Minnich.** 1994. Thermoregulation of *Yersinia enterocolitica* is coincident with changes in DNA supercoiling. Mol. Microbiol. 12:187-199.

394. **Römling, U., and B. Tümmler.** 1991. The impact of two-dimensional pulsed-field gel electrophoresis techniques for the consistent and complete mapping of bacterial genomes: refined physical map of *Pseudomonas aeruginosa* PAO. Nucleic Acids Res. 19:3199-3206.

395. **Römling, U., and B. Tümmler.** 1994. Bacterial genome mapping. J. Biotechnol. 35:155-164.

396. **Ronimus, R.S., and D.R. Musgrave.** 1995. A comparison of the DNA binding properties of histone-like proteins derived from representatives of the two kingdoms of the Archaea. FEMS Microbiol. Lett. **134:**79-84.

397. **Rosenshine, I., and M. Mevarech.** 1989. Isolation and partial characterization of plasmids found in three *Halobacterium volcanii* isolates. Can. J. Microbiol. **35:**92-95.

398. **Rosenshine, I., R. Tchelet, and M. Mevarech.** 1989. The mechanism of DNA transfer in the mating system of an archaebacterium. Science **245:**1387-1389.

399. **Rosenshine, I., T. Zusman, R. Werczberger, and M. Mevarech.** 1987. Amplification of specific DNA sequences correlates with resistance of the archaebacterium *Halobacterium volcanii* to the dihydrofolate reductase inhibitors trimethoprim and methotrexate. Mol. Gen. Genet. **208:**518-522.

400. **Roussel, Y., M. Pebay, G. Guedon, J.-M. Simonet, and B. Decaris.** 1994. Physical and genetic map of *Streptococcus thermophilus* A054. J. Bacteriol. **176:**7413-7422.

401. **Rouvière-Yaniv, J., E. Bonnefoy, O. Huisman, and A. Almeida.** 1990. Regulation of HU protein synthesis in *Escherichia coli*, p. 247-257.. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

402. **Rouvière-Yaniv, J., M. Yaniv, and J.-E. Germond.** 1979. *E. coli* DNA binding protein HU forms nucleosome-like structure with circular double-stranded DNA. Cell **17:**265-274.

403. **Rowlands, T., P. Baumann, and S.P. Jackson.** 1994. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. Science **264:**1326-1329.

404. **Ryter, A., and A. Chang.** 1975. Localization of transcribing genes in the bacterial cell by means of high resolution autoradiography. J. Mol. Biol. **98:**797-810.

405. **Ryu, S., and S. Garges.** 1994. Promoter switch in the *Escherichia coli pts* operon. J. Biol. Chem. **269:**4767-4772.

406. **Sambrook, J., E.F. Fritsch, and T. Maniatis.** 1989. Molecular cloning a laboratory manual vol. 1 2nd edition. Cold Spring Harbor Laboratory Press, New York.

407. **Sanderson, K.E., and M. Demerec.** 1965. The linkage map of *Salmonella typhimurium*. Genetics **51:**897-913.

408. **Sandman, K., J.A. Krzycki, B. Dobrinski, R. Lurz, and J.N. Reeve.** 1990. HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus*, is most closely related to histones. Proc. Natl. Acad. Sci. USA **87:**5788-5791.

409. **Sankoff, D., R. Cedergren, and Y. Abel.** 1990. Genomic divergence through gene rearrangement. Methods Enzymol. **183**:428-439.

410. **Sankoff, D., G. Leduc, N. Antoine, B. Paquin, B.F. Lang, and R. Cedergren.** 1992. Gene order comparisons for phylogenetic inference: evolution of the mitochondrial genome. Proc. Natl. Acad. Sci. USA **89**:6575-6579.

411. **Sanz, J.L., I. Marín, L. Ramirez, J.P. Abad, C.L. Smith, and R. Amils.** 1988. Variable rRNA gene copies in extreme halobacteria. Nucleic Acids Res. **16**:7827-7832.

412. **Sanz, J.L., I. Marín, D. Ureña, and R. Amils.** 1992. Functional analysis of seven ribosomal systems from extremely halophilic archaea. Can. J. Microbiol. **39**:311-317.

413. **Sanzey, B.** 1979. Modulation of gene expression by drugs affecting deoxyribonucleic acid gyrase. J. Bacteriol. **138**:40-47.

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

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

416. **Schaaper, R.M., B.N. Danforth, and B.W. Glickman.** 1986. Mechanisms of spontaneous mutagenesis: an analysis of the spectrum of spontaneous mutation in the *Escherichia coli lacI* gene. J. Mol. Biol. **189**:273-284.

417. **Schalkwyk, L.C., R.L. Charlebois, and W.F. Doolittle.** 1993. Insertion sequences on plasmid pHV1 of *Haloferax volcanii*. Can. J. Microbiol. **39**:201-206.

418. **Schmid, M.B., and J.R. Roth.** 1983. Selection and endpoint distribution of bacterial inversion mutations. Genetics **105**:539-557.

419. **Schmid, M.B., and J.R. Roth.** 1987. Gene location affects expression level in *Salmo lla typhimurium*. J. Bacteriol. **169**:2872-2875.

420. **Schmidt, K.D., B. Tümmler, and U. Römling.** 1996. Comparative genome mapping of *Pseudomonas aeruginosa* PAO with *P. aeruginosa* C, which belongs to a major clone in cystic fibrosis patients and aquatic habitats. J. Bacteriol. **178**:85-93.

421. **Schoop, G.** 1935. *Halococcus litoralis*, ein obligat halophiler Farbstoffbildner. Deut. Tierärztl. Wochenschr. **43**:817-820.

422. **Searcy, D.G., and D.B. Stein.** 1980. Nucleoprotein subunit structure in an unusual prokaryotic organism: *Thermoplasma acidophilum*. Biochim. Biophys. Acta **609**:180-195.

423. **Segall, A., M.J. Mahan, and J.R. Roth.** 1988. Rearrangement of the bacterial chromosome: forbidden inversions. Science **241**:1314-1318.

424. **Segall, A.M., and J.R. Roth.** 1989. Recombination between homologies in direct and inverse orientation in the chromosome of *Salmonella*: intervals which are nonpermissive for inversion formation. Genetics **122**:737-747.

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

426. **Serror, P., V. Azevedo, and S.D. Ehrlich.** 1993. An ordered collection of *Bacillus subtilis* DNA segments in yeast artificial chromosomes, p. 473-474. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics*. American Society for Mcirobiology, Washington, D.C.

427. **Sharp, P.M.** 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33**:23-33.

428. **Shimizu, H., H. Yamaguchi, and H. Ikeda.** 1995. Molecular analysis of *λbio* transducing phage produced by oxolinic acid-induced illegitimate recombination *in vivo*. Genetics **140**:889-896.

429. **Simon, R.D.** 1978. *Halobacterium* strain 5 contains a plasmid which is correlated with the presence of gas vacuoles. Nature (London) **273**:314-317.

430. **Sinden, R.R., and D.E. Pettijohn.** 1981. Chromosomes in living *Escherichia coli* cells are segregated into domains of supercoiling. Proc. Natl. Acad. Sci. USA **78**:224-228.

431. **Smith, C.L., J.G. Econome, A. Schutt, S. Klco, and C.R. Cantor.** 1987. A physical map of the *Escherichia coli* K12 genome. Science **236**:1448-1453.

432. **Smith, D.R., H.M. Lee, J. Dubois, D. Qui, W. Caubet, R. Bashirzadeh, P. Parenteau, J. Wierzbowski, X. Wang, S. Shimer, J. Nolling, and J. Reeve.** 1995. Microbial genome sequencing. Gen. Sci. Technol. **1**:P-48.

433. **Snyder, M., and K. Drlica.** 1979. DNA gyrase on the bacterial chromosome: DNA cleavage induced by oxolinic acid. J. Mol. Biol. **131**:287-302.

434. **Soppa, J., and D. Oesterhelt.** 1989. *Halobacterium* sp. GRB: a species to work with!? Can. J. Microbiol. **35**:205-209.

435. **Sorokin, A., E. Zumstein, V. Azevedo, S.D. Ehrlich, and P. Serror.** 1993. The organization of the *Bacillus subtilis* 168 chromosome region between the *spoVA* and *serA* genetic loci, based on sequence data. Mol. Microbiol. **10:**385-395.

436. **Spiridonova, V.A., A.S. Akhmanova, V.K. Kagramanova, A.K.E. Köpke, and A.S. Mankin.** 1989. Ribosomal protein gene cluster of *Halobacterium halobium*: nucleotide sequence of the genes coding for S3 and L29 equivalent ribosomal proteins. Can. J. Microbiol. **35:**153-159.

437. **Spirito, F., N. Figueroa-Bossi, and L. Bossi.** 1994. The relative contributions of transcription and translation to plasmid DNA supercoiling in *Salmonella typhimurium*. Mol. Microbiol. **11:**111-122.

438. **St. Jean, A., and R.L. Charlebois.** 1996. Comparative genomic analysis of the *Haloferax volcanii* DS2 and *Halobacterium salinarium* GRB contig maps reveals extensive rearrangement. J. Bacteriol. **178:**3860-3868.

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

440. **Stackebrandt, E.** 1985. Phylogeny and phylogenetic classification of prokaryotes, p. 309-334. *In* K.H. Schleifer, and E. Stackebrandt (ed.), *Evolution of prokaryotes*. Academic Press, London.

441. **Starich, M.R., K. Sandman, J.N. Reeve, and M.F. Summers.** 1996. NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. J. Mol. Biol. **255:**187-203.

442. **Steck, T.R., R.J. Franco, J.-Y. Wang, and K. Drlica.** 1993. Topoisomerase mutations affect the relative abundance of many *Escherichia coli* proteins. Mol. Microbiol. **10:**473-481.

443. **Steitz, J.A.** 1978. Methanogenic bacteria. Nature (London) **273:**10.

444. **Stettler, R., G. Erauso, and T. Leisinger.** 1995. Physical and genetic map of the *Methanobacterium wolfei* genome and its comparison with the updated genomic map of *Methanobacterium thermoautotrophicum* Marburg. Arch. Microbiol. **163:**205-210.

445. **Stöffler, G., and M. Stöffler-Meilicke.** 1986. Electron microscopy of archaebacterial ribosomes. System. Appl. Microbiol. **7:**123-130.

446. **Stoltzfus, A., J.F. Leslie, and R. Milkman.** 1988. Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. Genetics **120:**345-358.

447. **Sugino, A., and N.R. Cozzarelli.** 1980. The intrinsic ATPase of DNA gyrase. J. Biol. Chem. **255:**6299-6306.

448. **Sun, L., and J.A. Fuchs.** 1994. Regulation of the *Escherichia coli nrd* operon: role of DNA supercoiling. J. Bacteriol. **176:**4617-4626.

449. **Sutherland, K.J., M. Hashimoto, T. Kudo, and K. Horikoshi.** 1993. A partial physical map for the chromosome of alkalophilic *Bacillus* sp. strain C-125. J. Gen. Microbiol. **139:**661-667.

450. **Tabata, K., and T. Hoshino.** 1996. Mapping of 61 genes on the refined physical map of the chromosome of *Thermus thermophilus* HB27 and comparison of genome organization with that of *T. thermophilus* HB8. Microbiology (Reading) **142:**401-410.

451. **Tabata, K., T. Kosuge, T. Nakahara, and T. Hoshino.** 1993. Physical map of the extremely thermophilic bacterium *Thermus thermophilus* HB27 chromosome. FEBS Lett. **331:**81-85.

452. **Takao, M., T. Kobayashi, A. Oikawa, and A. Yasui.** 1989. Tandem arrangement of photolyase and superoxide dismutase genes in *Halobacterium halobium*. J. Bacteriol. **171:**6323-6329.

453. **Takayanagi, S., S. Morimura, H. Kusaoke, Y. Yokoyama, K. Kano, and M. Shioda.** 1992. Chromosomal structure of the halophilic archaebacterium *Halobacterium salinarium*. J. Bacteriol. **174:**7207-7216.

454. **Tan, J., L. Shu, and H.-Y. Wu.** 1994. Activation of the *leu-500* promoter by adjacent transcription. J. Bacteriol. **176:**1077-1086.

455. **Tanaka, I., K. Appelt, J. Dijk, S. White, and K. Wilson.** 1984. 3-Å resolution structure of a protein with histone-like properties in prokaryotes. Nature (London) **310:**376-381.

456. **Taylor, A.L., and M.S. Thoman.** 1964. The genetic map of *Escherichia coli* K-12. Genetics **50:**659-677.

457. **Taylor, D.E., M. Eaton, W. Yan, and N. Chang.** 1992. Genome maps of *Campylobacter jejuni* and *Campylobacter coli*. J. Bacteriol. **174:**2332-2337.

458. **Tenover, F.C., R.D. Arbeit, R.V. Goering, P.A. Mickelsen, B.E. Murray, D.H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233-2239.

459. **Tindall, B.J., H.N.M. Ross, and W.D. Grant.** 1984. *Natronobacterium* gen. nov. and *Natronococcus* gen. nov., two new genera of haloalkaliphilic archaebacterium. System. Appl. Microbiol. **5:**41-57.

460. **Thompson, R.J., J.P. Davies, G. Lin, and G. Mosig.** 1990. Modulation of transcription by altered torsional stress, upstream silencers, and DNA-binding proteins, p. 227-240. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

461. **Thornton, M., M. Armitage, A. Maxwell, B. Dosanjh, A.J. Howells, V. Norris, and D.C. Sigee.** 1994. Immunogold localization of GyrA and GyrB proteins in *Escherichia coli*. Microbiology (Reading) **140**:2371-2382.

462. **Tigges, E., and F.C. Minion.** 1994. Physical map of *Mycoplasma gallisepticum*. J. Bacteriol. **176**:4157-4159.

463. **Toda, T., and M. Itaya.** 1995. I-*Ceu*I recognition sites in the *rrn* operons of the *Bacillus subtilis* 168 chromosome: inherent landmarks for genome analysis. Microbiology (Reading) **141**:1937-1945.

464. **Torreblanca, M., F. Rodríguez-Valera, G. Juez, A. Ventosa, M. Kamekura, and M. Kates.** 1986. Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid composition, and description of *Haloarcula* gen. nov. and *Haloferax* gen. nov. Syst. Appl. Microbiol. **8**:89-99.

465. **Trieselmann, B.A., and R.L. Charlebois.** 1992. Transcriptionally active regions in the genome of the archaebacterium *Haloferax volcanii*. J. Bacteriol. **174**:30-34.

466. **Tsao, Y.-P., H.-Y. Wu, and L.F. Liu.** 1989. Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. Cell **56**:111-118.

467. **Tse-Dinh, Y.-C.** 1985. Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. Nucleic Acids Res. **13**:4751-4763.

468. **Tupper, A.E., T.A. Owen-Hughes, D.W. Ussery, D.S. Santos, D.J.P. Ferguson, J.M. Sidebotham, J.C.D. Hinton, and C.F. Higgins.** 1994. The chromatin-associated protein H-NS alters DNA topology *in vitro*. EMBO J. **13**:258-268.

469. **Ueguchi, C., and T. Mizuno.** 1993. The *Escherichia coli* nucleoid protein H-NS functions directly as a transcriptional repressor. EMBO J. **12**:1039-1046.

470. **Umeda, M., and E. Ohtsubu.** 1989. Mapping of insertion elements IS*1*, IS*2* and IS*3* on the *Escherichia coli* K-12 chromosome: role of the insertion elements in formation of Hfrs and F′ factors and in rearrangement of bacterial chromosomes. J. Mol. Biol. **208**:601-614.

471. **Umeda, M., and E. Ohtsubu.** 1990. Mapping of insertion element IS*5* in the *Escherichia coli* K-12 chromosome: chromosomal rearrangements mediated by IS*5*. J. Mol. Biol. **213**:229-237.

472. **Upton, M., P.E. Carter, G. Orange, and T.H. Pennington.** 1996. Genetic heterogeneity of M type 3 group A streptococci causing severe infections in Tayside, Scotland. J. Clin. Microbiol. **34:**196-198.

473. **Vandamme, P., Y. Glupczynski, A.P. Lage, C. Lammens, W.G.V. Quint, and H. Goossens.** 1995. Evaluation of random and repetitive motif primed polymerase chain reaction typing of *Helicobacter pylori*. System. Appl. Microbiol. **18:**357-362.

474. **Van Valen, L.M., and V.C. Maiorana.** 1980. The archaebacteria and eukaryotic origins. Nature (London) **287:**248-250.

475. **Van Workum, M., S.J.M. Van Dooren, N. Oldenburg, D. Molenaar, P.R. Jensen, J.L. Snoep, and H.V. Westerhoff.** 1996. DNA supercoiling depends on the phosphorylation potential in *Escherichia coli*. Mol. Microbiol. **20:**351-360.

476. **Vary, P.S.** 1993. The genetic map of *Bacillus megaterium*, p. 475-481. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics*. American Society for Mcirobiology, Washington, D.C.

477. **Ventosa, A., and A. Oren.** 1996. *Halobacterium salinarum* nom. corrig., a name to replace *Halobacterium salinarium* (Elazari-Volcani) and to include *Halobacterium halobium* and *Halobacterium cutirubrum*. Intern. J. System. Bacteriol. **46:**347.

478. **Vettakkorumakankav, N.N., and K.J. Stevenson.** 1992. Dihydrolipoamide dehydrogenase from *Haloferax volcanii*: gene cloning, complete primary structure, and comparison to other dihydrolipoamide dehydrogenases. Biochem. Cell Biol. **70:**656-663.

479. **Walker, J.E., P.K. Hayes, and A.E. Walsby.** 1984. Homology of gas vesicle proteins in cyanobacteria and halobacteria. J. Gen. Microbiol. **130:**2709-2715.

480. **Wang, J.C.** 1971. Interaction between DNA and an *Escherichia coli* protein ω. J. Mol. Biol. **55:**523-533.

481. **Wang, J.-Y., and M. Syvanen.** 1992. DNA twist as a transcriptional sensor for environmental changes. Mol. Microbiol. **6:**1861-1866.

482. **Ward-Rainey, N., F.A. Rainey, E.M.H. Wellington, and E. Stackebrandt.** 1996. Physical map of the genome of *Planctomyces limnophilus*, a representative of the phylogenetically distinct planctomycete lineage. J. Bacteriol. **178:**1908-1913.

483. **Watterson, G.A., W.J. Evens, T.E. Hall, and A. Morgan.** 1982. The chromosome inversion problem. J. Theor. Biol. **99:**1-7.

484. **Weidinger, G., G. Klotz, and W. Goebel.** 1979. A large plasmid from *Halobacterium halobium* carrying genetic information for gas vacuole formation. Plasmid **2:**377-386.

485. **Welker, N.E.** 1993. The genetic map of *Bacillus stearothermophilus* NUB36, p. 483-487. *In* A.L. Sonenshein, J.A. Hoch, and R. Losick (ed.), Bacillus subtilis *and other gram-positive bacteria biochemistry, physiology, and molecular genetics.* American Society for Mcirobiology, Washington, D.C.

486. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18:**7213-7218.

487. **Wenzel, R., E. Pirkl, and R. Herrmann.** 1992. Construction of an *Eco*RI restriction map of *Mycoplasma pneumoniae* and localization of selected genes. J. Bacteriol. **174:**7289-7296.

488. **Whitson, P.A., W.-T. Hsieh, R.D. Wells, and K.S. Matthews.** 1987. Supercoiling facilitates *lac* operator-repressor-pseudooperator interactions. J. Biol. Chem. **262:**4943-4946.

489. **Whoriskey, S.K., V.-H. Nghiem, P.-M. Leong, J.-M. Masson, and J.H. Miller.** 1987. Genetic rearrangements and gene amplification in *Escherichia coli*: DNA sequences at the junctures of amplified gene fusions. Gene Dev. **1:**227-237.

490. **Whoriskey, S.K., M.A. Schofield, and J.H. Miller.** 1991. Isolation and characterization of *Escherichia coli* mutants with altered rates of deletion formation. Genetics **127:**21-30.

491. **Williams, J.G., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. **18:**6531-6535.

492. **Williamson, R.M., J. Hetherington, and J.H. Jackson.** 1993. Detection of fundamental principles and a level of order for large-scale gene clustering on the *Escherichia coli* chromosome. J. Mol. Evol. **36:**347-360.

493. **Woese, C.R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221-271.

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

495. **Woese, C.R., and R. Gupta.** 1981. Are archaebacteria merely derived 'prokaryotes'? Nature (London) **289:**95-96.

496. **Woese, C.R., R. Gupta, C.M. Hahn, W. Zillig, and J. Tu.** 1984. The phylogenetic relationships of three sulfur-dependent archaebacteria. System. Appl. Microbiol. **5:**97-105.

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

498. **Woese, C.R., L.J. Magrum, and G.E. Fox.** 1978. Archaebacteria. J. Mol. Evol. **11**:245-252.

499. **Wong, K.K., and M. McClelland.** 1992. A *Blp*I restriction map of the *Salmonella typhimurium* LT2 genome. J. Bacteriol. **174**:1656-1661.

500. **Worcel, A., and E. Burgi.** 1972. On the structure of the folded chromosome of *Escherichia coli*. J. Mol. Biol. **71**:127-147.

501. **Worcel, A., and E. Burgi.** 1974. Properties of a membrane-attached form of the folded chromosome of *Escherichia coli*. J. Mol. Biol. **82**:91-105.

502. **Wu, H.-Y., S. Shyy, J.C. Wang, and L.F. Liu.** 1988. Transcription generates positively and negatively supercoiled domains in the template. Cell **53**:433-440.

503. **Wu, S.-R., S.L. Hillier, and K. Nath.** 1996. Genomic DNA fingerprint analysis of biotype 1 *Gardnerella vaginalis* from patients with and without bacterial vaginosis. J. Clin. Microbiol. **34**:192-195.

504. **Wu, Y., and P. Datta.** 1995. Influence of DNA topology on expression of the *tdc* operon in *Escherichia coli* K-12. Mol. Gen. Genet. **247**:764-767.

505. **Yamada, H., S. Muramatsu, and T. Mizuno.** 1990. An *Escherichia coli* protein that preferentially binds to sharply curved DNA. J. Biochem. **108**:420-425.

506. **Yamaguchi, H., T. Yamashita, H. Shimizu, and H. Ikeda.** 1995. A hotspot of spontaneous and UV-induced illegitimate recombination during formation of *λbio* transducing phage. Mol. Gen. Genet. **248**:637-643.

507. **Yamamoto, N., and M.L. Droffner.** 1985. Mechanisms determining aerobic or anaerobic growth in the facultative anaerobe *Salmonella typhimurium*. Proc. Natl. Acad. Sci. USA **82**:2077-2081.

508. **Yang, C.-F., and S. DasSarma.** 1990. Transcriptional induction of purple membrane and gas vesicle synthesis in the archaebacterium *Halobacterium halobium* is blocked by a DNA gyrase inhibitor. J. Bacteriol. **172**:4118-4121.

509. **Yang, C.-F., J.-M. Kim, E. Molinari, and S. DasSarma.** 1996. Genetic and topological analyses of the *bop* promoter of *Halobacterium halobium*: stimulation by DNA supercoiling and non-B-DNA structure. J. Bacteriol. **178**:840-845.

510. **Yang, Y., and G.F.-L. Ames.** 1988. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. Proc. Natl. Acad. Sci. USA **85**:8850-8854.

511. **Yang, Y., and G.F.-L. Ames.** 1990. The family of repetitive extragenic palindromic sequences: interaction with DNA gyrase and histonelike protein HU, p. 211-225. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.

512. **Ye, F., F. Laigret, and J.M. Bové.** 1994. A physical and genomic map of the prokaryote *Spiroplasma melliferum* and its comparison with the *Spiroplasma citri* map. C. R. Acad. Sci. Paris, Sciences de la vie, Biologie et génétique moléculaire **317**:392-398.

513. **Ye, F., F. Laigret, P. Carle, and J.M. Bové.** 1995. Chromosomal heterogeneity among various strains of *Spiroplasma citri*. Intern. J. System. Bacteriol. **45**:729-734.

514. **Young, G.M., and K. Postle.** 1994. Repression of *tonB* transcription during anaerobic growth requires Fur binding at the promoter and a second factor binding upstream. Mol. Microbiol. **11**:943-954.

515. **Yoshida, T., C. Ueguchi, H. Yamada, and T. Mizuno.** 1993. Function of the *Escherichia coli* nucleoid protein, H-NS: molecular analysis of a subset of proteins whose expression is enhanced in a *hns* deletion mutant. Mol. Gen. Genet. **237**:113-122.

516. **Zeigler, D.R., and D.H. Dean.** 1990. Orientation of genes in the *Bacillus subtilis* chromosome. Genetics **125**:703-708.

517. **Zillig, W., R. Schnabel, J. Tu, and K.O. Stetter.** 1982. The phylogeny of archaebacteria, including novel anaerobic thermoacidophiles, in the light of RNA polymerase structure. Naturwissenschaften **69**:197-204.

518. **Zhang, A., S. Rimsky, M.E. Reaban, H. Buc, and M. Belfort.** 1996. *Escherichia coli* protein analogs StpA and H-NS: regulatory loops, similar and disparate effects on nucleic acid dynamics. EMBO J. **15**:1340-1349.

519. **Zhilina, T.N.** 1986. Methanogenic bacteria from hypersaline environments. System. Appl. Microbiol. **7**:216-222.

520. **Zhu, Y.S., and J.E. Hearst.** 1988. Transcription of oxygen-regulated photosynthetic genes requires DNA gyrase in *Rhodobacter capsulatus*. Proc. Natl. Acad. Sci. USA **85**:4209-4213.

521. **Zuerner, R.L., J.L. Herrmann, and I. Saint Girons.** 1993. Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intraspecies heterogeneity. J. Bacteriol. **175**:5445-5451.

522. **Zuerner, R.L., and T.B. Stanton.** 1994. Physical and genetic map of the *Serpulina hyodysenteriae* B78^T chromosome. J. Bacteriol. **176**:1087-1092.

523. **Zyskind, J.W.** 1990. Priming and growth rate regulation: questions concerning initiation of DNA replication in *Escherichia coli*, p. 269-278. *In* K. Drlica, and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.