Pulsed-field gel electrophoresis studies of Rhizobiaceae

Bruno Walther S. Sobral

Iowa State University
INFORMATION TO USERS

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

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

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

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

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

UMI
University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600
Pulsed-field gel electrophoresis studies of *Rhizobiaceae*

Sobral, Bruno Walther S., Ph.D.

Iowa State University, 1989
Pulsed-field gel electrophoresis studies
of Rhizobiaceae
by
Bruno Walther S. Sobral

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
Major: Genetics

Approved:
Signature was redacted for privacy.

In Charge of Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1989
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Explanation of the Dissertation Format</td>
<td>1</td>
</tr>
<tr>
<td>A Brief History of Agarose Gel Electrophoresis of DNA</td>
<td>2</td>
</tr>
<tr>
<td>Why Use PFG to Study Rhizobiaceae?</td>
<td>27</td>
</tr>
<tr>
<td>Overall Goals and Specific Goals of These Studies</td>
<td>36</td>
</tr>
<tr>
<td><strong>SECTION I. PULSE TIME AND AGAROSE CONCENTRATION AFFECT THE ELECTROPHORETIC MOBILITY OF cccDNA DURING ELECTROPHORESIS IN CHEF AND FIGE</strong></td>
<td>38</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>40</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>41</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>44</td>
</tr>
<tr>
<td>Agarose Gel Electrophoresis</td>
<td>44</td>
</tr>
<tr>
<td>Circular and Linear DNA Ladders</td>
<td>45</td>
</tr>
<tr>
<td>Calculation of Electrophoretic Mobility</td>
<td>46</td>
</tr>
<tr>
<td>RESULTS</td>
<td>50</td>
</tr>
</tbody>
</table>
iii

Electrophoresis of cccDNAs in CHEF ............ 50
FIGE of cccDNAs ............................. 53

DISCUSSION ................................. 60

ACKNOWLEDGMENTS ........................... 65

SECTION II. APPLICATION OF LARGE DNA METHODS
TO ANALYSIS OF Rhizobiaceae GENOMES .... 66

ABSTRACT .................................... 68

INTRODUCTION ............................... 70

MATERIALS AND METHODS ..................... 74
Bacterial Strains, Plasmids, and Growth Conditions 74
Preparation and Restriction Digestion of Genomic DNA
in Agarose Plugs ............................ 75
PFG Equipment and Electrophoresis Conditions .... 79
Identification of Rare-Cutting Restriction
Endonucleases ............................... 81
Southern Hybridization ....................... 82
Image Analysis and Scanning Densitometry of PFG Gels 83
 Calculation of Sizes of Restriction Fragments on PFG
Gels .......................................... 83
GENERAL INTRODUCTION

Explanation of the Dissertation Format

The format of this dissertation is the alternate format. This format was selected for simplicity of organization. The dissertation is organized into three Sections. Each Section corresponds to a separate manuscript submitted to a scientific journal for publication.

The first Section describes studies in which the migration of circular, covalently closed (ccc) DNA molecules undergoing two forms of pulsed-field gel electrophoresis (PFG) was compared to the migration of linear DNA molecules in the same electrophoresis systems. The studies in Section I were to serve as a model system for predicting the appropriate PFG conditions to resolve circular DNAs. The resolution of large plasmids in PFG systems would be useful for mapping Rhizobium meliloti megaplasmids.

The second Section relates the application of PFG techniques to study genomic organization and structure in selected Rhizobiaceae. Methodologies developed for PFG studies in Rhizobiaceae bacteria are discussed. The second Section also describes PFG studies of genome structure and organization in (brady)rhizobia. R. J. Honeycutt, the second
author of this section, contributed by preparing bacteroids of bradyrhizobia for PFG analysis and by performing Southern hybridizations with \textit{nod} and \textit{nif} gene probes.

Section III describes a specific application of the PFG techniques, developed in Section II, to the study of a practical problem in \textit{Bradyrhizobium japonicum}, namely the genetic diversity in field isolates from serocluster USDA 123. M. J. Sadowsky (second author of the third Section) contributed by offering his personal field isolate collection and communicating unpublished data (not reported in the section) about these isolates.

A Brief History of Agarose Gel Electrophoresis of DNA

Bridging the 'resolution gap' between molecular cloning and genetic mapping

The last ten years has brought about considerable advances in many areas of molecular biology. One rapidly developing area is that of gene mapping and genome analysis, where the accumulation of mapping information continues at an exponential rate. This is especially true for the human gene mapping efforts (Collins, 1988).

Molecular cloning and gene mapping involve very different size scales. Conventional cloning techniques are only useful
up to approximately 50 kilobase pairs (kb), the maximum insert for cosmid cloning, although it is possible to map larger distances by linking overlapping cosmid clones (known as cosmid 'walking'). Cosmid walking techniques are impractical beyond distances of a couple hundred kb. Linkage analysis, in situ hybridization, and somatic cell methods are useful in the size range of approximately 5,000 kb to over 100,000 kb (Collins, 1988). Single genes can vary in size from less than 1 kb to over 2,000 kb. Because of the 'resolution gap' in the 100 to 5,000 kb range, it was not clear until recently that information obtained from mapping efforts might be useful to guide cloning efforts.

The last five years have yielded a variety of approaches that are useful in the size range required to fill the resolution gap. Pulsed-field gel electrophoresis (PFG) (Schwartz and Cantor, 1984; Carle and Olson, 1984; Carle et al., 1986; Chu et al., 1986), chromosome jumping (Collins and Weissman, 1984; Poustka et al., 1987), and yeast artificial chromosomes (YACs [Burke et al., 1987]) have provided means of bridging the gap and allowing the use of mapping data to direct searches for previously unclonable genes of interest.

The importance of PFG is underscored by the limitations of continuous field (CF) agarose gel electrophoresis of DNA molecules, which has a practical resolution limit of approximately 50 kb (Serwer, 1989). PFG, on the other hand,
has allowed the direct resolution of molecules in excess of 10 megabase pairs (Mb) in size (Vollrath and Davis, 1987; Orbach et al., 1988), and is one of the most promising breakthroughs in molecular genetics within this decade. PFG experiments have permitted the resolution of entire chromosomes of lower eucaryotes, such as a variety of fungi (Orbach et al., 1988; Carle and Olson, 1985; Schwartz and Cantor, 1984; Smith et al., 1987b; Vollrath and Davis, 1987; Snell and Wilkins, 1986; Brody and Carbon, 1989) and parasitic protozoa (Van der Ploeg et al., 1984). Techniques have been developed that allow the manipulation of DNAs from various procaryotic and eucaryotic organisms (Smith et al., 1988). PFG research has yielded complete physical maps for large regions of mammalian chromosomes (Shaw, 1986; Pohl et al., 1988; Fulton et al., 1989; Wallace et al., 1989), and for the entire chromosome of some procaryotes (Smith et al., 1987; Ventra and Weiss, 1989; Lee and Smith, 1988; Kauc et al., 1989; Canard and Cole, 1989; Bautsch, 1988; Suwanto and Kaplan, 1989a, 1989b; Ely and Gerardot, 1988; Bancroft et al., 1989). It is expected that PFG will yield even more information about genome organization and structure and that more complex organisms will be analyzed as the mechanisms of PFG and DNA electrophoresis become understood and the technology is refined.
A brief introduction to electrophoresis of DNA. PFG theory, principle and systems

The electrophoretic mobility ($\mu$) of a particle is primarily determined by its average electrical charge per surface area in the absence of a gel (Abramson et al., 1964; Shaw, 1969; Andrews, 1981). Electrophoretic mobility can be defined mathematically as $\mu = \frac{\text{velocity}}{\text{electrical potential gradient}}$ (E). In contrast to $\mu$ measured in gels, $\mu$ measured in the absence of a gel is known as $\mu_0$ (Serwer and Allen, 1984). For doubled-stranded DNA, $\mu_0$ is independent of DNA length ($L$), measured in base pairs (Olivera, et al., 1964; Serwer and Allen, 1984; Serwer, 1989). When agarose gel electrophoresis is used, $\mu$'s extrapolated to 0% agarose ($A$) are the same for open circular (oc), closed circular (ccc) and linear DNAs (Serwer and Allen, 1984). The extrapolated $\mu$ differs from $\mu_0$ by a constant, so the equality of the $\mu_0$ of linear, oc and ccc DNAs holds true (Serwer, 1989). This constant depends on the electroendosmosis (EEO) of the gel (Serwer, 1989), which is the electrical field-induced flow of the buffer (Abramson, 1964; Shaw, 1969; Johnson et al., 1980; Serwer, 1980; Serwer and Allen, 1984). By using $\mu = f(A)$ plots, Serwer (1980) showed that $\mu_0$ was independent of $L$ for DNAs from 2 to 40 kb. Therefore, all separations of double-stranded DNA by $L$ and conformation are caused by either steric
or hydrodynamic modification of μ by fibers of the gel matrix, a process known as sieving (Serwer, 1989).

When performing agarose gel electrophoresis on linear, double-stranded DNA molecules, the sieving of agarose usually causes μ to monotonically decrease in value as L increases (Serwer, 1980; McDonell et al., 1977; Hervert and Bean, 1987). Unfortunately, there are many exceptions to this rule. An inverted μ vs. L relationship occurs only in a very restricted range of values of A, E, and L (Serwer, 1989).

To estimate the L of linear DNAs, it is commonplace to perform agarose gel electrophoresis of standards in a lane adjacent to that of the DNA of unknown length. A semilogarithmic L = f(μ) plot of the standard DNA has a linearity range for 1- to 40-kb DNAs and therefore can be used to calibrate the gel for sizing of DNAs within this size range (Serwer, 1980; McDonell et al., 1977). Other fragment size estimation methods that are not based on semilogarithmic plots also can be used (Southern, 1979; Duggleby et al., 1981; Vandergraft, 1983). Decreasing A causes an increase in the linearity range (Serwer, 1989). As DNAs become longer than the DNA in the range of linearity, and as the voltage gradient and A increase, the values of μ become increasingly independent of L during electrophoresis (Serwer, 1980; McDonell et al., 1977). By appropriate manipulation of the values of A and E, the useable resolution by L can be
maintained for DNAs up to 100 to 170 kb (Serwer, 1980). However, both of these manipulations will lower the value of μ, thereby causing the electrophoresis time to be impractical for most experimental purposes (Serwer, 1980). Excessive lowering of A causes the subsequent manipulations of the agarose gel for photography, Southern analysis, etc., to become extremely difficult. In practice, then, the upper limit for CF electrophoresis is approximately 50 kb.

The first explanation for the loss of resolution by L, observed as L increased, was electrical field-induced elongation of the DNA (Dingman et al., 1972) based on its viscoelastic properties (Smith et al., 1986). This elongation was proposed to result in the end-first migration of the DNA, a process called "reptation" (de Gennes, 1971; Lumpkin et al., 1985). Reptation was originally applied successfully to situations involving diffusion, where no external electric field is present (de Gennes, 1971). The extension of reptation theories to situations where a uniform electric field is applied, such as in agarose gel electrophoresis of DNA, is not a mathematically straightforward process, although many groups have used mathematical approaches (Lerman and Frisch, 1982; Lumpkin and Zimm, 1982, Lumpkin et al., 1985; Slater and Noolandi, 1985, 1989; Slater et al., 1987; Viovy, 1987, 1989; Noolandi et al., 1989). The main problem is that the steady state statistics of the DNA molecule's
configurations are unknown (Deutsch and Madden, 1989). During reptation, DNA molecules are viewed as passing through an idealized "tube", composed of a series of gel pores, in a snakelike motion that is biased by the direction of the electric field. This process is known as "biased reptation"; these models are also referred to as tube models. Experimental observations from CF electrophoresis support tube theories, although the comparison was based only on values of $\mu$ (Hervert and Bean, 1987). At values of $A$ below those for which reptation is believed to lead to a loss in resolution by $L$ for linear DNAs, a semilogarithmic $\mu$ vs. $A$ plot (Ferguson plot, [Ferguson, 1964]) is linear (Serwer, 1980; Serwer and Allen, 1984; Stellwagen, 1985).

Deutsch (1988) proposed a different theory to explain the migration of DNA molecules in agarose gels. Based on numerical studies and computer simulations, he proposed that the application of an electric field would induce the DNA chain to contract on itself. Contraction is followed by "unwinding" of the DNA molecule into an extended configuration, in what Richards (1989) has called a "slinky" motion (because it is similar to a child's slinky toy, which is a squat helical spring, going down stairs). The kinks predicted to form in the DNA molecule as it migrates clearly violate tube models, in which kinks are predicted to be statistically extremely rare (Deutsch and Madden, 1989). The
predictions of Deutsch's simulations also are in agreement with experimental data from CF electrophoresis (Deutsch, 1988), just as tube theories. Tube theories, however, are not capable of satisfactorily explaining other forms of electrophoresis (see below), that can be explained by what I call the slinky theory. The slinky theory also predicts that the molecular motion of the DNA molecule undergoing electrophoresis would be quite different from the motion predicted by tube theories (Deutsch, 1988). Experimental observations of stained DNA molecules undergoing CF electrophoresis (Schwartz and Koval, 1989; Smith et al., 1989), conducted after Deutsch proposed his model, strongly support the slinky theory.

The above observations were made for electrophoresis performed by using an electrical field which is constant in space and time (CF electrophoresis). To increase the resolution by $L$ and overcome the limitation on the resolution of linear DNAs, the direction of the electric field has been periodically varied (pulsed-field gel electrophoresis, or PFG). It has been proposed that the alternation of the electrical fields causes the DNA molecules to continually orient and reorient as the field changes direction (Schwartz and Cantor, 1984; Serwer, 1989). Furthermore, the time required for orientation, disorientation and reorientation in the new field direction causes improved molecular weight
separation, provided that the switching times are correctly chosen to coincide with the orientation times of the DNA molecules under study (Serwer, 1989). The first PFG experiments were done by using an apparatus that had two orthogonally oriented sets of vertical electrodes (Schwartz and Cantor, 1984). To change the direction of the electric field, activation of one set of electrodes was stopped and interrupted by activation of the other set at defined intervals. The original PFG equipment is known by the acronym of orthogonal field alternating gel electrophoresis (OFAGE). The angle between successive field directions (ϕ) was initially designed to be 90° to prevent reptation (Schwartz and Cantor, 1984). This value for ϕ was later found to be ineffective (Southern et al., 1987; Serwer, 1987). Effective values for ϕ were ≤ 180° and > 90° (Southern et al., 1987; Serwer, 1987; Carle et al., 1986; Lalande et al., 1987). When ϕ = 180°, either time (Carle et al., 1986) or E (Lalande et al., 1987) of electrophoresis in the forward direction is greater than that used for the reverse direction (known as field inversion gel electrophoresis, or FIGE). FIGE systems are the easiest to build yet they suffer from nonmonotonic μ vs. L plots (Carle et al., 1986; Ellis et al., 1987; Bostock, 1988; Heller and Pohl, 1989; Violy, 1988); this nonmonotonic behavior can be reduced, but not eliminated, by increasing the pulse times (known as "ramping" [Ellis et al., 1987]).
Although FIGE is very simple to set up in the laboratory, the complex migratory behavior of the DNA molecules is difficult to explain theoretically (but see Viovy, 1988; Serwer, 1989), and makes size estimation more difficult. For all other values of $\phi$, the time of electrophoresis in each direction is the same. Some of the alternatives to conventional OFAGE boxes include gel boxes in which the electrodes are fixed and the box itself is rotated (Serwer, 1987; Sutherland et al., 1989); pulsed homogenous orthogonal field gel electrophoresis (PHOGE [Bancroft and Wolk, 1988]), in which the lane curvature is removed by using homogenous electric fields; contour clamped homogenous electric fields (CHEF [Chu et al., 1986]), in which homogenous electric fields ($\phi = 120^\circ$), and therefore straight DNA migration patterns between lanes, are produced by clamping voltage; and pulse oriented electrophoresis (POE [Schwartz and Koval, 1989]), which allows the effective angle of the electric field to be varied without electrode rearrangement. A unique variation in PFG systems, the transverse alternating field electrophoresis (TAFE) system (Gardiner et al., 1986; Gardiner and Patterson, 1988, 1989; Wert and Furst, 1989), uses a three-dimensional electrode design. In TAFE systems, the gel is placed vertically between two sets of electrodes arranged parallel to the gel faces, thereby generating fields that are transverse to the plane of the gel rather than parallel to it. Because of the upright
placement of the gel in the box, there are limitations on the A values which can be used. The electric field is equivalent for all lanes, therefore straight lines are produced. There is an electrical field gradient which decreases with distance down the gel. In TAFE systems, the E gradient is believed to be responsible for band sharpening (Gardiner and Patterson, 1989). The value of $\phi$ also is variable with distance down the gel; $\phi = 115^\circ$ at the top of the gel and decreases significantly midway down the gel. The variable field pattern produced in TAFE causes the proportion of the downward component of E to gradually increase relative to the lateral component through the upper half of the gel. In the lower half of the gel, the horizontal component of E increases until the fields become essentially horizontal, causing stacking and band compression to occur (Gardiner and Patterson, 1989).

In early studies, experiments were performed under conditions in which the angle between the electrodes varied throughout the gel box. It has been proposed that the difference in $\phi$ experienced as the DNA molecules migrate through the gel may be important for production of 'sharp' bands (i.e., superior resolution [Gardiner and Patterson, 1989]). Recent studies using a programmable, autonomously controlled electrode (PACE [Birren et al., 1988; Birren et al., 1989; Lai et al., 1989]) system have shown that resolution is almost the same for values of $\phi$ between 105° and
165°, for DNAs up to 1 Mb in size. Although the resolution is not altered, the DNA's migration velocity is significantly changed for the different $\phi$ values (Birren et al., 1988), with almost a three-fold difference in velocity depending on the value of $\phi$ (Birren et al., 1989). The smaller the $\phi$ value (but always $> 90^\circ$), the larger the velocity of migration of yeast chromosomes, especially the larger ones (Birren et al., 1989). In fact, **Schizosaccharomyces pombe** chromosomes have been separated in as little as 24 h by using $\phi = 94^\circ$ (Clark et al., 1988). Separations performed with $\phi$ values $\geq 120^\circ$ can cause DNAs to migrate in a size-independent manner (Carle and Olson, 1987). Experiments performed with the PACE system have shown that by decreasing the value of $\phi$ from $120^\circ$ to $110^\circ$ it is possible to restore the dependence of relative mobility on size (Birren et al., 1989).

The changes in velocity caused by changing the $\phi$ values resemble the changes in relative mobility caused by altering the pulse interval. A lowering of the value of $\phi$ seems to have an effect that is similar to lengthening the switch intervals. Decreasing the $\phi$ value and, at the same time, decreasing the switch interval causes an increase in the resolution and a lower separation time (Birren et al., 1989). Recent studies, using FIGE systems (Heller and Pohl, 1989), support the notion that the pulse time and E parameters are nearly exchangeable. These results are similar to those
obtained in other PFG systems (Mathew et al., 1988a, 1988b; Birren et al., 1988; Smith et al., 1987b).

Southern et al. (1987) proposed a model for separation of DNAs by crossed-field PFG systems ($90^\circ \leq \phi < 180^\circ$). Southern's model contrasts in its predictions about DNA mobility, when compared to predictions made by biased reptation models. Southern proposed that the number of pulses, rather than the pulse length, is the major factor governing the separation process during PFG in crossed fields. The effect of pulse length reported by earlier workers (Schwartz and Cantor, 1984; Carle and Olson, 1985) would really be an effect of the number of pulses, since the original PFG experiments considered gels which were run for the same amount of time (Southern et al., 1987). Under these conditions, the number of pulses is inversely proportional to the pulse length. This proportionality would explain the relationship between pulse length and separation observed by Schwartz and Cantor (1984).

Southern's model assumes that the extended orientation of a DNA molecule undergoing agarose gel electrophoresis is retained when the current is removed. Following this argument, it is concluded that the few seconds between pulses does not allow sufficient time for the large DNA molecules to undergo extensive reorientation (Southern et al., 1987). In Southern's model, the switching of field direction (for $\phi > 90^\circ$) causes the DNA molecule's new leading edge to most likely
be its old back edge, because other possibilities would require more work (Southern et al., 1987). This is known as the "ratcheting" mechanism for PFG resolution (Southern et al., 1987; Bancroft and Wolk, 1988). Southern et al. (1987) concluded that the separation between DNA molecules at the end of electrophoresis is proportional to the difference in their $L$ values (Southern et al., 1987). It is predicted by and consistent with Southern's model that $\phi$ values $< 90^\circ$ do not separate better than CF electrophoresis (Southern et al., 1987). The ineffectiveness of $\phi < 90^\circ$ is explained by Southern's model because the new leading edge of the DNA molecule would most likely be the old leading edge, as other possibilities would require more work. Therefore, the inhomogeneous electric fields of the original OFAGE system (Schwartz and Cantor, 1984) were necessary for the DNA molecules to experience values of $\phi > 90^\circ$ (Bancroft and Wolk, 1988). Both the ratcheting model and the reptation model are only useful in explaining separation in crossed fields; neither model is capable of explaining FIGE separations ($\phi = 180^\circ$; but see Noolandi et al., 1989).

The slinky model (Deutsch, 1987, 1988), derived for CF electrophoresis, has also been applied to PFG separations, especially to FIGE, where tube models are incapable of explaining the observed mobilities (Deutsch, 1988). The slinky model has far fewer assumptions than most tube models,
although it is still a very simplified model. Microscopic observations of individual DNA molecules (bacteriophage lambda cI857 Sam 7 DNA, which is 48.5 kb in size, and yeast chromosomes, which range from 0.2 to 2 Mb in size) undergoing electrophoresis in a CHEF system (Smith et al., 1989) or in a POE system (Schwartz and Koval, 1989) are in agreement with ratcheting (Southern et al., 1987) and slinky models (Deutsch, 1988). It appears that biased reptation models cannot satisfactorily explain the migration of DNA undergoing some forms of PFG (but see Deutsch, 1988; Noolandi et al., 1989). The ratcheting model (Southern et al., 1987) is also incapable of explaining separations observed in FIGE, whereas the slinky model (Deutsch, 1988) is in accord with most of the experimental observations seen in FIGE and other PFG systems. Tube models and the ratcheting model can be adapted to explain FIGE observations by assuming flexibility of the idealized segments that represent the DNA molecule (Noolandi et al., 1989; Southern et al., 1987), instead of the rigidity normally postulated by these two models. The mathematical deductions made for adaptations which allow flexibility of the DNA become increasingly complex, however.

The exact mechanisms involved in PFG separations are still unclear; however models are necessary to make predictions to increase resolution and decrease run time. It is generally agreed that factors which increase the velocity
of migration decrease the overall resolution of the gel (Birren et al., 1988; Lai et al., 1989). Factors which increase the velocity of migration include increasing the gel temperature, decreasing the magnitude of $\bar{E}$, increasing $E$, or decreasing $A$ (Birren et al., 1988, 1989; Lai et al., 1989). Transient electric birifringence has shown that there is significant orientation of the matrix itself (Stellwagen and Stellwagen, 1989b). Constant orientation and reorientation of the matrix domains in pulsed electric fields would give fluidity to the matrix, and therefore make it easier for very large molecules to migrate in response to $E$ (Stellwagen and Stellwagen, 1989a). The orientation of the matrix in an electric field also affects the mobility of DNA fragments in the gel (Stellwagen and Stellwagen, 1989b). By using linear dichroism spectroscopy, Akerman et al. (1989) showed that DNA/gel and DNA/DNA interactions also are important in the orientational dynamics of DNA molecules undergoing gel electrophoresis. These three studies, along with the gel hysteresis hypothesis of Serwer (1988) suggest an important and previously unsuspected role for the gel matrix during electrophoretic separations of DNA molecules subjected to the CF and PFG forms of agarose gel electrophoresis. Unfortunately, the statistics of agarose gels are not understood (Waki et al., 1982; West, 1987), making it difficult to incorporate gel heterogeneity as a factor even in
simulations (Deutsch and Madden, 1989).

The various factors that affect DNA migration during PFG are interrelated; therefore, changes in any given parameter will alter not only the rate, but also the relative mobility of DNA in a gel (Birren et al., 1989). It is necessary to introduce compensating changes in other parameters to maintain the resolution level desired while permitting the separation to occur faster. Selection of appropriate parameters, therefore, require an understanding of the interdependence of the factors that influence PFG separations (reviewed in Birren et al., 1989; Bostock 1988; Heller and Pohl, 1989).

Mathematical modelling (Noolandi et al., 1989; Slater et al., 1987; Viovy, 1987; Viovy, 1989; Willis et al., 1988; Deutsch, 1987, 1988; Deutsch and Madden, 1989), observation of single DNA molecules in agarose gels (Smith et al., 1989; Schwartz and Koval, 1989), linear and electric dichroism studies (Holzwarth et al., 1987; Akerman et al., 1989; Porsche, 1989), electric birifringence measurements (Stellwagen, 1985; Stellwagen and Stellwagen, 1989a, 1989b), scanning tunnel microscopy (Keller et al., 1989), observation of the movement of fluorescence patterns after photobleaching (Chu et al., 1989), and computer simulations (Lalande et al., 1987; Batoulis et al., 1989; Deutsch, 1988; Smith et al., 1989) are currently important areas of study and have increased our understanding of the CF and PFG electrophoretic processes (see
Together, these approaches should yield more precise methods for choosing appropriate experimental conditions to separate DNAs of different sizes and topologies, as well as a universal theory for the electrophoresis of DNA and for PFG separations (Heller and Pohl, 1989).

Overview of the applications of PFG

As in all scientific endeavors, the understanding we have of genome organization is closely tied to the references which we use to study it, namely genetic maps and physical maps. Genetic maps suffer from serious limitations because they represent apparent distances between genes or other scorable markers. Distances on genetic maps are determined by using relative frequencies of genetic events that occur between these genes or markers. It is known in many organisms that the frequency of genetic events such as recombination is not constant throughout the genome (for an example in *Zea mays*, see Dooner et al., 1985). Also, the resolution of genetic maps cannot, in many instances, achieve the detail desired because of the limitations on the number of matings that can be examined. In general, refinement of a genetic map requires detection of increasingly rarer events, which in turn requires analysis of a greater number of meiotic events (i.e., the progeny of matings). In mammals, such as man, even the best
mapped regions of the genome only achieve a resolution of approximately 1 Mb (Smith and Cantor, 1986). However, in smaller genomes, such as bacteriophages, the resolution of genetic maps can be refined to achieve the single base pair level. Finally, in organisms such as parasitic protozoa, no mating techniques exist to generate data for the construction of a genetic map.

In contrast to genetic maps, physical maps are a direct reflection of the structure of the DNA molecules that make up chromosomes. Before the advent of PFG technology, physical maps could be made either by examination of chromosomes with light microscopy, or by using molecular techniques. For most organisms, with the noteworthy exception of Drosophila melanogaster, chromosome maps are very limited in resolution. Molecular maps can be as complete as a full nucleotide sequence, or may be composed of restriction sites and physical distances (usually in kb) between these sites. Molecular maps containing information at the restriction site or nucleotide sequence level are generally smaller than 200 kb, and most are actually limited to approximately 50 kb.

An obvious application of PFG analysis is to link information generated by genetic analyses to information generated by molecular analyses. Many examples in the current literature illustrate the usefulness of PFG in bridging the gap between genetic or chromosome maps and molecular maps
(Fulton et al., 1989; Gardiner and Patterson, 1989; Gemmill et al., 1987; Michiels et al., 1987; Pohl et al., 1988; Poustka et al., 1987; Shaw, 1986; Smith et al., 1987a, 1987b; Smith and Cantor, 1986; Ventra and Weiss, 1989; Wallace et al., 1989).

The chromosomes of many yeasts and protozoa are in the size range that can be resolved by using current PFG technology. PFG can directly yield information about the karyotype, number, and size of chromosomes of organisms possessing relatively small genomes (Schwartz and Cantor, 1984). These electrophoretic karyotypes have the added benefit of providing separated chromosomes for subsequent experiments. Southern hybridization of cloned DNA fragments to blotted PFG gels can be used to assign these fragments to chromosomal locations (Schwartz and Cantor, 1984). In other words, PFG has created the novel technique of gene mapping by electrophoresis, which does not require a genetic phenotype or even genetic tools! Furthermore, electrophoretically separated chromosomes can be used as starting material for further mapping and for cloning, and chromosome-specific libraries can be prepared after PFG separation of chromosomes. These approaches have tremendous potential, and they have been applied to a number of organisms, including *Saccharomyces cerevisiae* (Schwartz and Cantor, 1984; Carle and Olson, 1985), *Schizosaccharomyces pombe* (Smith et al., 1987b), *Candida*
albicans (Snell and Wilkins, 1986; Lasker et al., 1989),
Aspergillus nidulans (Brody and Carbon, 1989), trypanosomes
(Van der Ploeg et al., 1984; Bernards et al., 1986),
leishmania (Giannini et al., 1986; Spithill and Samaras,
1985), Plasmodium falciparum (Corcoran et al., 1986; Van der
Ploeg et al., 1985), Tetrahymena thermophila (Conover and
Brunk, 1986), giardia (C. L. Smith, Department of Genetics,
Columbia University, NY, personal communication), and
Chlamydomonas reinhardii (Hall et al., 1989).

Most bacterial genomes contain a single, circular DNA
molecule. The electrophoretic mobility of large, circular DNA
molecules during PFG is under intense investigation (Hightower
et al., 1987, 1989; Levene and Zimm, 1987, 1989; Beverley,
1988, 1989; Mathew et al., 1988a; Hightower and Santi, 1989;
Simske and Scherer, 1989), but at present circular molecules
of the size of bacterial chromosomes (which are at least 1 to
2 Mb in size) apparently cannot be resolved by PFG systems.
However, bacterial chromosomes can be digested in situ with
appropriate rare-cutting restriction enzymes (McClelland et
al., 1987), thereby converting the circular DNA molecule into
a collection of relatively small linear molecules. If DNA
sequence information is available for the organisms under
investigation, statistical predictions can be used to
determine the appropriate restriction enzymes (McClelland et
al., 1987); otherwise, trial and error must be used. The
genomic fingerprint produced by digestion with rare-cutting restriction enzymes allows comparative studies involving closely related species, strains, or isolates (see Section III). In addition, it is possible to simply add together the sizes of the restriction fragments and compute the size of the entire genome. The bacterial chromosome can then be mapped by determination of the relative order of the fragments in the genome, much in the same way as is done for plasmid mapping. The macrorestriction map produced by PFG analysis is an important tool for genetic mapping, for monitoring genomic rearrangements, since deletions and insertions can be detected by changes in the sizes of certain restriction fragments (Smith et al., 1986), and for evolutionary studies. Genes can be assigned to fragments by hybridization with cloned probes to PFG blots digested with rare-cutting enzymes. Complete or near-complete physical maps have been generated for a few procaryotes so far, including Mycoplasma mobile (Bautsch, 1988), Borrelia burgdorferi (Ferdows and Barbour, 1989) Bacillus subtilis (Ventra and Weiss, 1989), Staphylococcus aureus (Patel et al., 1989), Haemophilus influenzae (Lee and Smith, 1988; Kauc et al., 1989), Chlamydia (Frutos et al., 1989), Rickettsiella (Frutos et al., 1989), Pororchlamydia (Frutos et al., 1989), Clostridium perfringens (Canard and Cole, 1989), Ureaplasma urealyticum (Cocks et al., 1989), Rhodobacter sphaeroides (Suwanto and Kaplan, 1989a, 1989b),
Anabaena sp. (Bancroft et al., 1989), Caulobacter crescentus (Ely and Gerardot, 1988), and Escherichia coli (Smith et al., 1987a).

PFG techniques also are capable of producing valuable information about much larger and more complex genomes, such as mammalian genomes, and those of higher plants. In these cases, the general approaches described above for smaller genomes can be applied to segments of more complex genomes. There are regions of mammalian genomes for which a large number of cloned probes are available; PFG analysis of such regions has made it possible to make restriction maps covering many megabases. In the case of the human genome, regions containing genes involved in major diseases have been analyzed, such as the major histocompatibility complex (Lawrence et al., 1986; Lawrence et al., 1987; Hardy et al., 1986; Tokunaga et al., 1989), the cystic fibrosis gene (Fulton et al., 1989; Rommens et al., 1989), the Duchenne muscular dystrophy gene (van Ommen et al., 1986; Kenwick et al., 1987; Burmeister and Lehrach, 1986), and the region involved in Huntington's disease (Pohl et al., 1988). Human chromosome 21, the smallest of the human chromosomes (current estimates suggest a size of 35 to 50 Mb), is an object of intense study in many laboratories. Smith and Cantor (Department of Genetics, Columbia University, NY, personal communication) have obtained an 8.5 Mb NotI restriction map of the short arm
of chromosome 21. It is expected that a complete NotI map of chromosome 21 will soon be completed (C. L. Smith, Department of Genetics, Columbia University, NY, personal communication). These studies not only highlight the practical importance of PFG technology, but also demonstrate the feasibility of using the technology to make physical maps of entire human chromosomes. In fact, many investigators are currently mapping other human chromosomes by using PFG. Complete physical maps of all the human chromosomes will greatly speed the human genome sequencing effort.

There are investigators applying PFG techniques to higher plants, especially those with smaller genomes, such as Arabidopsis thaliana (Guzman and Ecker, 1988), and various species of rice (J. R. Ecker, Plant Science Institute, University of Pennsylvania, PA, personal communication). Plants with larger genomes, such as Lycopersicon esculentum (Ganal and Tanksley, 1989; van Daelen et al., 1989), Glycine max (R. J. Honeycutt, Department of Genetics, Iowa State University, IA, personal communication), Triticum aestivum (Devos and Vercruysse-Dewitte, 1989), Daucus carota (Guzman and Ecker, 1988), and Z. mays (R. J. Honeycutt, Department of Genetics, Iowa State University, IA, personal communication) have begun to receive attention, although intact chromosomes of these important crop species will not be resolved by PFG before substantial advances in resolution capacity of PFG
separations are made. *A. thaliana* and *Oryza sativa* have relatively small genomes, when compared with other higher plants. *A. thaliana* has a genome size of 70 to 80 Mb (Bennet and Smith, 1976), and *O. sativa* has a genome size of approximately 450 to 600 Mb (Bennet and Smith, 1976). The relatively small genome size of these two plants make them excellent model systems for dicotyledonous and monocotyledonous plants, respectively. Both plants are currently being investigated by using PFG and yeast artificial chromosome (YAC) techniques (Guzman and Ecker, 1988; J. R. Ecker, Plant Science Institute, University of Pennsylvania, personal communication). It is expected that saturated restriction fragment length polymorphism (RFLP) maps will be linked by PFG technology within a few years. Also, rapidly evolving PFG technology should allow larger and more complex genomes to be analyzed with increasing ease in the near future. Modifications in PFG technology presently being investigated, such as secondary pulsing (C. R. Cantor, Department of Genetics, Columbia University, NY, personal communication), allow separation of DNA molecules larger than 10 Mb in size within approximately 30 hours.
Why Use PFG to Study Rhizobiaceae?

The importance of Rhizobiaceae

The Rhizobiaceae family includes members of Bradyrhizobium, Rhizobium, Agrobacterium, and Phyllobacterium genera (Jordan, 1984). The genus Agrobacterium includes important phytopathogenic microorganisms such as A. tumefaciens, the causative agent of crown gall (Jordan, 1984), and A. rhizogenes, the causative agent of hairy root disease (Jordan, 1984). The Bradyrhizobium and Rhizobium genera are composed of bacteria that are capable of fixing atmospheric dinitrogen in symbiosis with certain plants belonging to the Leguminosae family. Although estimates vary, biological dinitrogen fixation accounts for approximately $1.7 \times 10^8$ metric tons per year of fixed nitrogen (Hardy and Holsten, 1972). This is more than half of the total fixed nitrogen of the Earth's nitrogen balance. Within this context, the legume/(brady)rhizobia symbiosis is by far the most efficient system of biological nitrogen fixation, with estimated rates of 55-140 kg N$_2$ fixed ha$^{-1}$ yr$^{-1}$ (Delwiche, 1970). Manipulation of the (brady)rhizobia/legume interaction is of great agronomic importance because of its role in maintaining a favorable balance of nitrogen and of great biological importance because of its use as a model system for
understanding plant gene regulation (Gresshoff and Delves, 1986). (Brady)Rhizobia have been intensively studied for many years, resulting in the identification of genes involved in the nodulation and dinitrogen fixation processes (Djordjevic et al., 1983; Fischer and Hennecke, 1984; Schmidt et al., 1984; Schofield et al., 1983, 1984; Masterson et al., 1985). Fewer genetic techniques are available for studying (brady)rhizobia than for studying other bacterial groups, such as the enteric bacteria (Long, 1984), yet many fundamental techniques have been developed in many laboratories. Some important genetic techniques that have been developed include conjugation and transposon mutagenesis (Beringer et al., 1978; Meade et al., 1982), gene cloning by complementation (Ruvkun and Ausubel, 1981), transformation (Kiss and Kalman, 1982), development of broad host-range cloning vehicles (Jacob et al., 1976; Julliot and Boistard, 1979; Ditta et al., 1980; Friedman et al., 1982; Long et al., 1982; Cantrell et al., 1982; Ward and Grinsted, 1982; Kiss and Kalman, 1982; Tait et al., 1982; Johnston et al., 1983; Simon et al., 1983), clone bank construction and manipulation (Ditta et al., 1980; Ruvkun and Ausubel, 1981; Friedman et al., 1982), and electrotransformation (Bowen and Kossiak, 1989). Although these developments have allowed an accumulation of information about the symbiotic process, many aspects of symbiosis are still not well understood. The genetic techniques available
for fast-growing rhizobia are more well-developed than those available for the study of slow-growing bradyrhizobia (Long, 1984).

Evolutionary studies, using 35S ribosomal RNA and the conserved nitrogenase structural genes, have shown that members of the *Agrobacterium* and *Rhizobium* genera are more closely related to one another than either is to members of the *Bradyrhizobium* genus (Jordan, 1984; Hennecke et al., 1985). This is interesting because, in some instances, a member of the genus *Rhizobium* is capable of nodulating and fixing dinitrogen with the same legume host as a member of the genus *Bradyrhizobium* (Jordan, 1984). One such example is *R. fredii* and *B. japonicum*, both of which are capable of nodulating and fixing dinitrogen in symbiosis with certain genotypes of *G. max* (Keyser et al., 1982). Based on numerical taxonomy studies, Chen et al. (1988) proposed that *R. fredii* should be reclassified as *Sinorhizobium*, a new genus.

Many hypotheses have been forwarded to explain the evolutionary origin of the *Rhizobiaceae*, especially for (brady)rhizobia (Verma and Long, 1983). The evolutionary relatedness of different members of the *Rhizobiaceae* could be investigated with PFG, especially with regard to conserved regions of the genome. This would shed light on evolutionary hypotheses. An understanding of taxonomic and evolutionary considerations would provide essential background to the
understanding of how nodulation genes function, since many experiments are performed by the transfer of plasmids or cloned genes from one species of rhizobia into another (Long, 1984). The outcome of such experiments might well be influenced by the genetic relatedness of the strains used (Long, 1984).

The lack of detailed genetic maps in many important Rhizobiaceae

One of the problems investigators face when studying symbiosis is the lack of complete, high resolution genetic maps for (brady)rhizobia and host plants. Many (brady)rhizobial nodulation (nod), nitrogen fixation (nif and fix), and exopolysaccharide biosynthesis (exo) genes have been cloned and sequenced (for examples of reviews see Long, 1984; Gussin et al., 1986). Our understanding of the process of symbiotic dinitrogen fixation has increased from the study of these cloned genes. Molecular maps, which exist for nod and nif regions in many species of (brady)rhizobia, typically span no more than 50 kb (Long et al., 1982; Prakash et al., 1982; Schofield et al., 1983; Downie et al., 1983; Kondorosi et al., 1984; Masterson and Atherly, 1986). In most fast-growing rhizobia, many symbiotically important genes are localized on large extrachromosomal elements, known as megaplasmids
(reviewed in Denarie et al., 1981; Huguet et al., 1983). Many megaplasmids still are considered cryptic, mainly because of our lack of knowledge about the genes that they carry (Hynes et al., 1989). Interestingly, in some rhizobial species, such as *R. meliloti*, there are strains that possess megaplasmids that are extremely recalcitrant to curing (Eberhard, 1989; Huguet et al., 1983; Hynes et al., 1989). Resistance to curing has also been observed with some *Agrobacterium* large plasmids (Eberhard, 1989). Recently, genes encoding proteins of essential biosynthetic pathways, such as thiamine biosynthesis, have been localized to megaplasmids in *R. meliloti* (Finan et al., 1986). Based on electron microscope (EM) contour measurements of large samples of megaplasmids, Burkardt et al. (1987) suggested that variance in contour lengths of *R. meliloti* megaplasmids might be caused by DNA rearrangements. I believe that some of the large, extrachromosomal elements may, in fact, be essential for survival. If true, this would invalidate, by definition, the classification of these replicons as plasmids. Attempts to map the megaplasmids in *R. meliloti* by deletion analysis have failed because deletion of certain regions of these replicons appears to be lethal (S. R. Long, Department of Biology, Stanford University, personal communication). This would be expected if the replicons are essential for survival. Although *R. meliloti* is the most thoroughly studied of the
Rhizobiaceae at the genetic level, genetic maps (Meade and Signer, 1977; Casadeus and Olivares, 1979; Hooykaas et al., 1982; Martin and Long, 1984; Beringer et al., 1987) are of very low resolution for most regions of the genome. This is also true for R. leguminosarum (Kondorosi et al., 1980) and Agrobacterium (Hooykaas et al., 1982). In B. japonicum, the most thoroughly studied of the bradyrhizobia, investigators do not have a complete, circular genetic map of the chromosome, at any level of resolution. Some B. japonicum strains have large plasmids (150 to 300 kb), although symbiotic functions are not encoded by these replicons (Masterson et al., 1982; Russell and Atherly, 1981). Attempts to map some regions of the B. japonicum chromosome by cosmid cloning have failed because of the instability of certain regions in E. coli (M. Schell, Department of Microbiology, University of Tennessee, TN, personal communication).

PFG technology can be applied to Rhizobiaceae to further our knowledge about the genomic organization of the bacteria in this family, without the need for development of new techniques. The generation of chromosomal maps in Rhizobiaceae may only be possible by PFG analysis, since other techniques have generally failed. The uses of physical maps have been described in the previous sections.
Genome size in Rhizobiaceae

Size estimates for the genomes of Rhizobiaceae and other bacteria can be made by traditional methods, such as Feulgen microdensitometry (Feulgen and Rossenbeck, 1924), chemical extraction coupled with the diphenylamine reaction (Burton, 1956, 1968) and cell counts, or chemical extraction coupled with ultraviolet spectrophotometry (Sunderland and McLeish, 1961). When studying bacteria, total cellular DNA content, which is the quantity measured by Feulgen microdensitometry, is often erroneously confused with genome size (Herdman, 1985). Total DNA content is the genome size multiplied by the number of genome copies. In bacteria, the total number of genome copies is usually greater than one, and is dependent upon growth phase. There are many potential errors and problems associated with the use of Feulgen microdensitometry, most of which originate from the misuse of the method itself (Bennett and Smith, 1976). Standardization of every aspect of the method is crucial (reviewed in Bennett and Smith, 1976). Errors and problems can also arise from any or all of the following: incomplete staining, non-staining DNA (observed only in some eucaryotic tissues), variation in DNA density (it is known that the DNA amount decreases as the DNA density increases [Mittwoch, 1969]), and optical errors.

Because bacterial genomes usually are of low genetic
complexity, genome sizes are frequently measured by the kinetics of reassociation of denatured DNA as monitored spectrophotometrically by a decrease in absorbance (Wetmur and Davidson, 1968; Gillis et al., 1970; Gillis and DeLey, 1979). It is necessary to carefully control parameters such as ionic strength, temperature, and DNA fragment size and concentration, for the rate of renaturation to give a direct measurement of genetic complexity (Herdman, 1985).

Genome size estimates, produced by analysis of reassociation kinetics, are variable for many Rhizobiaceae. For *B. japonicum*, the genome is estimated at 10 to 12 Mb (Sober, 1970; Carlson et al., 1985; Chakrabarti et al., 1984). Estimates for most rhizobia and agrobacteria vary between 4.5 and 5.5 Mb (Sober, 1970; Crow et al., 1981; Herdman, 1985; Chakrabarti et al., 1984); these size estimates include megaplasmids, when present. *R. meliloti* 2011 megaplasmids have been physically sized by using electron microscopy (Burkardt and Burkardt, 1984; Burkardt et al., 1987). *R. meliloti* 2011 contains 2 megaplasmids; pRme2011a has a 400-μm contour length, and pRme2011b has a 560-μm contour length (Burkardt et al., 1987). A 1-mm contour length equals a molecular weight of $2 \times 10^9$, or approximately 3 Mb of DNA, assuming an average weight of 650 for a single base pair. The contour length measurements of *R. meliloti* megaplasmids imply that pRme2011a is approximately 1.2 Mb in size, whereas
pRme2011b is approximately 1.7 Mb in size. If the genome size of *R. meliloti* is 4.8 to 5.5 Mb (Sober, 1970; Crow et al., 1981; Chakrabarti et al., 1984; Herdman, 1985), the megaplasmid measurements would indicate that the so-called chromosome of *R. meliloti* must be between 1.9 to 2.6 Mb in size. Electron microscope measurements of the *R. meliloti* 2011 chromosome resulted in an estimated length of 1.8 mm, which implies that the chromosome is at least three times larger than the largest megaplasmid (Burkardt et al., 1987). Since pRme2011b is approximately 1.7 Mb in size (Burkardt et al., 1987), the chromosome should be at least 5.1 Mb in size, giving a total genome size of at least 7 Mb. Clearly, existing data about genome size in *R. meliloti* are not consistent. An accurate measure of the genome size is important for crucial manipulations such as clone bank construction, as well as for computing the number of clones that must be screened to identify a desired gene. As discussed in the previous section, current PFG technology has the potential of generating precise estimates of genome size for most procaryotic organisms, once the appropriate rare-cutting restriction enzymes have been identified.
Overall Goals and Specific Goals of These Studies

Overall goals of PFG analysis of Rhizobiaceae

The studies described herein are part of a project that will eventually provide a greater understanding of important plant microbes of the Rhizobiaceae family. The overall goals of the project are:

1. To apply techniques for preparing intact genomic DNA in agarose plugs from members of the Rhizobiaceae;
2. To identify rare-cutting restriction endonucleases and restriction endonuclease/methylase combinations that can be used in PFG analysis of the Rhizobiaceae;
3. To analyze whole genomic fingerprints in B. japonicum for major serogroups and for field isolates of the USDA 123 serocluster;
4. To identify the location of symbiotically-important genes on large DNA fragments for use in cloning and further mapping studies;
5. To construct yeast artificial chromosome (YAC) libraries for evolutionary studies and simplified mapping of cloned genes;
6. To use PFG to compare the genomes of free-living and bacteroid forms of B. japonicum, especially genome organization and structure;
7. To calculate the genome size of selected members of the **Rhizobiaceae**; and
8. To generate physical maps of *B. japonicum* and *R. meliloti*.

**Specific goals of these studies**

The specific goals addressed by the studies herein are described in items 1, 2, 3, 4, 6, and 7 of the previous section. Section I of this dissertation includes studies on the migration of supercoiled DNAs during PFG. One of the goals of the studies in Section I was to generate information that could be used in designing experiments to resolve megaplasmids from fast-growing rhizobia. Separation of megaplasmids by PFG experiments would be useful for the generation of physical maps of these replicons, as discussed previously. Goal number 3 above is addressed in Section III, whereas the other goals mentioned are addressed in Section II.
SECTION I. PULSE TIME AND AGAROSE CONCENTRATION AFFECT
THE ELECTROPHORETIC MOBILITY OF cccDNA DURING
ELECTROPHORESIS IN CHEF AND FIGE
TITLE: Pulse time and agarose concentration affect the electrophoretic mobility of cccDNA during electrophoresis in CHEF and in FIGE

AUTHORS: Bruno W. S. Sobral\textsuperscript{1,2*} and Alan G. Atherly\textsuperscript{1}

ADDRESS: Department of Genetics, Room 8, Curtiss Hall, Iowa State University, Ames, Iowa 50011\textsuperscript{1}.

CURRENT ADDRESS: California Institute of Biological Research, 11099 N. Torrey Pines Rd., La Jolla, CA 92037\textsuperscript{2}.

* Indicates corresponding author.

Circular DNAs have been shown to migrate in an unusual manner during field inversion gel electrophoresis (FIGE) and orthogonal field alternating gel electrophoresis (OFAGE). We studied the effect of varying pulse time and agarose concentration on the electrophoretic mobility of supercoiled (ccc) DNAs ranging from 2 kb to 16 kb during FIGE and contoured homogeneous electric fields (CHEF). Both supercoiled and linear molecules display a minimum mobility as a function of pulse time in a CHEF apparatus. Linear and cccDNAs of the same size are differently affected by pulse time. Pulse-time dependence was observed for cccDNAs in both systems. Pulse-time dependence in FIGE is very small at a 1.0% agarose concentration, but is pronounced in 0.8% or 1.2% gels.
INTRODUCTION

In continuous field (CF) agarose gel electrophoresis the electrophoretic mobility ($\mu$) of a DNA molecule is influenced by its size (Olivera et al., 1964; McDonell et al., 1977; Serwer, 1980; Hervert and Bean, 1987), topology (Dingman et al., 1972; Johnson and Grossman, 1977; Mickel et al., 1977; Serwer and Allen, 1984), applied field strength, $\epsilon$, (Hervert and Bean, 1987; Johnson and Grossman, 1977; Serwer and Allen, 1984; Stellwagen, 1985, Lumpkin et al., 1985), ionic environment (Olivera et al., 1964; Hervert and Bean, 1987; Johnson and Grossman, 1977; Mickel et al., 1977; Ross and Scruggs, 1964), agarose concentration, $A$ (Serwer, 1980; Hervert and Bean, 1987; Serwer and Allen, 1984), loading mass (Johnson et al., 1980), and temperature (Olivera et al., 1964; Hervert and Bean, 1987; Serwer and Allen, 1984; West, 1987). These parameters can be adjusted to give maximal resolution for DNAs up to 20 kilobase pairs (kb) in size. Pulsed-field gel electrophoresis (PFG) extends the separation range of agarose gel electrophoresis to larger DNAs (Schwartz and Cantor, 1984; Carle and Olson, 1984). During PFG, the mobilities of different-sized fragments of DNA are altered in a size-dependent manner in response to the variations in the electric field. PFG technology has introduced new factors that affect the migration of DNA molecules. Two new factors
introduced by PFG are pulse time and reorientation angle. It is not clearly understood how these new factors interact with factors previously studied in CF electrophoresis.

Few studies on the migration of circular DNAs during PFG have been reported (Hightower et al., 1987; Levene and Zimm, 1987; Mathew et al., 1988a; Beverley, 1988). Mathew et al. (1988a) suggested that reorientation times might be a function of both size and topology of the DNA molecule undergoing PFG. Electrophoretic studies involving different topological forms of DNA could be meaningful to theories about the migration of DNA in agarose during PFG. Using field-inversion gel electrophoresis (FIGE [Carle et al., 1986]), Levene and Zimm (Levene and Zimm, 1987) studied the migration of open circular (oc) DNAs at various field strengths. They showed that open circular (oc) DNAs from 2.9 kb to 56 kb, when subjected to a 1-s forward (P^) and 0.25-s reverse pulse (P_p), display greater electrophoretic mobilities than when the pulsing is omitted. Shorter pulses do not alter the mobilities of ocDNAs. Thus, varying pulse times can cause alterations in the value of \( \mu \) for one topological form of circular DNAs. Studies using orthogonal field alternating gel electrophoresis (OFAGE [Carle and Olson, 1984]) configurations suggest that changes in the mobility of circular topoisomers in relation to linear DNAs are caused by pulsing of the electric fields (Hightower et al., 1987; Mathew et al., 1988a; Beverley, 1988). In the
double inhomogeneous (di) OFAGE configuration, the mobility of supercoiled (ccc) DNAs is constant over a wide range of pulse times (Hightower et al., 1987). The degree of supercoiling also was shown to have a dramatic effect on plasmid migration. Beverley (1988) showed that a complex behavior is observed in OFAGE for cccDNAs of 30 kb and 85 kb. Directional and quantitative differences are observed for cccDNA mobilities when compared with larger yeast chromosomes, with which they co-migrate. There are no systematic reports about the electrophoretic mobility of circular DNAs undergoing PFG in electrophoresis using a contoured homogeneous electric field (CHEF [Chu et al., 1986]).

The present study investigates the influence of pulse time and agarose concentration on the electrophoretic mobility of cccDNAs undergoing FIGE and CHEF. We observed a pulse-time-dependent behavior in the migration of all sizes of cccDNAs studied. This pulse-time dependency was influenced by agarose concentration.
MATERIALS AND METHODS

Agarose Gel Electrophoresis

Agarose (Seakem LE, FMC) gels (11 cm x 14 cm x 3 mm) were prepared as described (Maniatis et al., 1982). A modified tris-borate buffer, TBEM (Mathew et al., 1988), a tris-acetate buffer, TAE (Maniatis et al., 1982), or a tris-borate buffer, TBE (Maniatis et al., 1982), was used at either full (1X) or half (0.5X) strength. The agarose solution was stirred gently and cooled to 55°C before being poured. The gels were allowed to polymerize in the casting tray overnight at room temperature before being loaded. A Saran Wrap cover was used to prevent excessive dehydration of the gels. Programmable pulse controllers or power supplies were used to control the duration of the runs. Constant voltage was maintained for all experiments. Current and temperature were monitored throughout the runs. For CF electrophoresis, an International Biotechnologies Incorporated (IBI, New Haven, CT) model MRH horizontal gel box was used. The interconnecting port was kept open to maintain uniform pH during electrophoresis. For CF gels the initial temperature was 22°C, and the equilibrium temperature was 27°C (±3°C). Temperature was maintained by recirculating the running buffer by using a peristaltic pump set at a flow rate of 25 ml/min. Actual voltage gradients (e)
were measured directly with a voltmeter. The voltmeter reading was divided by inter-electrode distance.

FIGE gels were run in either an MRH box or an SE600 (Hoefer Scientific Instruments, San Francisco, CA) vertical slab box. Either a PC750 (Hoefer Scientific Instruments, San Francisco, CA) or a Minipulse (IBI, New Haven, CT) pulse controller was used to control pulsing time. A constant forward pulse:reverse pulse (P_f:P_r) ratio of 3:1 was maintained for each experiment. Temperature and voltage gradients were measured in the same manner as in CF gels. Ramped FIGE gels (14 cm x 16 cm x 3 mm) were run in an SE600. All ramps increased linearly with time from the initial P_f value to the final P_r value. The temperature was maintained at 13°C.

Electrophoresis in CHEF was done by using a Pulsaphor Plus system (Pharmacia/LKB, Piscataway, NJ) and a hexagonal electrode kit. The voltage gradient was estimated by dividing the voltage read from the power supply by the distance between electrodes. A correction factor of 0.8 was used, to produce a more accurate estimation of the value of ε within the buffer. The running temperature was maintained at 13°C.

Circular and Linear DNA Ladders

The supercoiled ladder was purchased from BRL (Gaithersburg, MD) (size (kbp) = 16.14, 14.17, 12.14, 10.10,
8.07, 7.05, 6.03, 5.01, 3.99, 2.97, 2.07). The ladder was separated into aliquots upon arrival and used at a concentration of 200 ng/10 μl. To avoid accumulation of ocDNA in the ladder, the aliquots were frozen immediately and stored at -20°C until use. Aliquots were checked routinely on conventional gels for the presence of oc bands. Batches containing large amounts of ocDNAs were not used.

The following linear size standards were used for size estimation: 1) a lambda\textit{HindIII} digest, size (in kb) = 23.130, 9.416, 6.557, 4.361, 2.322, 2.027, 0.564, 0.125); 2) a high molecular weight ladder (BRL, Gaithersburg, MD, size (in kb) = 48.502, 38.416, 33.498, 29.942, 24.776, 22.621, 19.399, 17.057, 15.004, 12.220, 10.086, 8.612, 8.271); 3) lambda monomer, size (in kb) = 48.50; and 4) a 1-kb ladder (BRL, Gaithersburg, MD., size [in kb] = 12.216, 11.198, 10.180, 9.162, 8.144, 7.126, 6.108, 5.090, 4.072, 3.054, 2.036, 1.636, 1.018, 0.517, 0.506, 0.396, 0.344, 0.298, 0.220, 0.201, 0.154, 0.134, 0.075). All linear standards were aliquoted and stored at -20°C upon arrival. Linear standards were heated at 65°C for 10 min, then chilled on ice for 2 min before being loaded.

Calculation of Electrophoretic Mobility

After electrophoresis, gels were stained in ethidium bromide (100 ng/ml) overnight. Gels were destained in a 0.1%
aqueous solution of H\textsubscript{2}O\textsubscript{2} for 30 min. The migration (cm) was measured for circular and linear bands from photographic negatives. The electrophoretic mobility, $\mu$ (cm\textsuperscript{2} V\textsuperscript{-1} s\textsuperscript{-1}), was calculated for each supercoiled and linear DNA by dividing its migration (cm) by the product of $\epsilon$ (V cm\textsuperscript{-1}) and the duration of the run (s). Each $\mu$ represents an average of measurements from at least two gels. Graphs of $\mu = f($size$)$ were prepared for the supercoiled ladder and the linear standards of each gel. These graphs were used to determine the crossover point (in kb), which is the size at which linear and cccDNAs of the same size co-migrate.
FIGURE 1. Plots of the electrophoretic mobility of supercoiled and 1-kb ladders as a function of size for electrophoresis in CHEF at different pulse times. A) supercoiled ladder; B) 1-kb ladder. The insert on the upper right-hand corner contains the figure legends for both ladders, showing the curve that corresponds to each pulse time studied. Agarose concentration was 1%, and the runs lasted 16 h. Nominal voltage gradient was 6.4 V cm$^{-1}$, and the running buffer was 0.5 x TBE.
RESULTS

Electrophoresis of cccDNAs in CHEF

In all FIGE and CHEF gels, we found that the migration of the linear DNA ladders fit by using the SIZER program available on BIONET. SIZER considered all ladders a very good fit, with a standard error always below 0.280. SIZER strongly recommended against deleting any bands. SIZER-produced estimates of the cccDNA size by using the Southern (1979), spline (Vandergraft, 1983), or Duggleby (Duggleby et al., 1981) methods, invariably judged the larger cccDNAs off-scale because there were no linear standards that displayed migration distances within 10% of the larger cccDNAs.

In Figure 1A, $\mu_{ccc} = f(\text{size})$ plots are shown for the various pulse times studied with the cccDNAs. The lowest values for $\mu_{ccc}$ were produced by a pulse time of 10 s. The largest values of $\mu_{ccc}$ in electrophoresis using CHEF were produced by the 25-s pulse time. The 3-s, 5-s, and 15-s pulse times produced approximately equal values of $\mu$ for all cccDNAs studied. Figure 1B is a plot of $\mu = f(\text{size})$ for a 1-kbp linear DNA ladder. Both 1-s and 10-s pulse times slowed the migration velocity of linear DNAs, although the 1-s pulse time slowed the larger (greater than 5 kb) DNAs more than the smaller DNAs. The descending order of $\mu$ for the linear DNAs
FIGURE 2. Plots of electrophoretic mobility of supercoiled DNAs as a function of pulse time and of crossover point as a function pulse time for electrophoresis in CHEF. A) $\mu_{ccc}$ plotted as a function of pulse time for select cccDNAs ranging from 5 kbp to 16 kbp (the smaller supercoiled DNAs were omitted for clarity, although their behavior is similar); B) crossover point plotted as a function of pulse time. The insert below A) is the legend for the sizes of the cccDNAs that correspond to each curve. Gels were 1% agarose and were run for 16 h at a nominal voltage gradient of 6.4 V cm$^{-1}$. Electrophoresis was done in 0.5 x TBE.
Figure A: Plot of $\mu \times 10^{-6}$ against Pulse time (s) for different DNA lengths (16.21Kb to 5.01Kb).

Figure B: Plot of Crossover point (kbp) against Pulse time (s) for different DNA lengths (8.07Kb to 5.01Kb).
can be expressed as $25 \text{s} > 15 \text{s} > 5 \text{s} > 3 \text{s} > 10 \text{s} \approx 1 \text{s}$. For the 23.13-kb lambda/HindIII fragment, the sequence of decreasing values of $\mu$ was: $15 \text{s} \approx 25 \text{s} > 5 \text{s} \approx 3 \text{s} > 10 \text{s} > 1 \text{s}$. This sequence is similar to that observed for the 1-kb ladder. The running conditions for this gel produced band compression of the high molecular weight fragments of the linear ladder. Therefore, the high molecular weight fragments of the linear ladder were not plotted.

The response to pulse time of cccDNAs studied was similar for all sizes (Fig. 2A). The crossover point was plotted as a function of pulse time in Figure 2B. All crossover point values shown in Figure 2B were smaller than the value obtained from a gel subjected to continuous electric fields (not shown). In general, increasing the pulse time lowered the value of the crossover point. The response of the crossover point to increasing pulse time was similar to the response to increasing $\epsilon$ or $A$ (over 0.5%) observed by Serwer and Allen (1984).

**FIGE of cccDNAs**

We conducted studies in CF electrophoresis (not shown) for comparison with FIGE and CHEF data. As reported in other studies (Mickel et al., 1977; Mathew et al., 1988a), the crossover point was higher in TAE than in TBE running buffers.
Figure 3. Plots of electrophoretic mobility as a function of size for supercoiled and 1-kb ladders in FIGE at different agarose concentrations. A-C) supercoiled DNA ladder; D-F) 1-kb ladder. Inserts on the upper right-hand corner of each plot show the legend containing the pulse time which corresponds to each curve. Gels were run at a voltage gradient of 5.71 V cm$^{-1}$ for 7 (0.8% A), 8 (1.0% A), or 10 (1.2% A) h in 0.5 x TBE.
We found that the use of 0.5 x TBE or TBEM did not affect the crossover point in CF electrophoresis. The crossover point in TBEM at A = 0.8% was similar to the crossover point in TAE at A = 1.0%.

The effect of pulse time and agarose concentration on \( \mu \) vs. kb plots for ccc and linear DNAs of similar size is presented in Figure 3. For cccDNAs in 0.8% gels (Fig. 3A), \( \mu \) decreased in the following order: 0 s > 72 s \( \approx \) 120 s \( \approx \) 0.12 s > 1.2 s \( \approx \) 12 s > 36 s. At A = 1.0% (Fig. 3B), the decreasing order became: 0 s > 72 s \( \approx \) 120 s > 1.2 s > 0.12 s \( \approx \) 12 s \( \approx \) 36 s. For 1.2% gels (Fig. 3C), the behavior of \( \mu_{\text{ccc}} \) with respect to pulse time became more complex. At a 0.12-s pulse time, the larger (size > 7 kb) cccDNAs migrated even more quickly than in CF electrophoresis.

In Figure 3 D-F, \( \mu = f(\text{size}) \) is plotted for the linear DNA ladder. As the value of A was increased, the effect of pulse time on the mobilities of linear DNAs became more pronounced. There was a complex effect of A on \( \mu_{\text{linear}} \). At A = 0.8% (Fig. 3D) the lowest \( \mu \) for all sizes was produced by a 36-s pulse time, whereas a 0.12-s pulse time produced the lowest \( \mu \)'s at A = 1.0% (Fig. 3E). At A = 1.2% (Fig. 3F), a 0.12-s pulse time produced the lowest \( \mu \)'s for DNAs greater than 5 kb, whereas the smaller DNAs migrated slowest at a 12-s pulse time.

The plot of crossover point as a function of pulse time
for FIGE gels of different A values is shown in Figure 4. At all values of A studied, there was an increase in the crossover point for the 0.12-s pulse time. The crossover point for pulse times of 1.2 s and greater were similar at A = 1.0%. At A = 0.8% and 1.2%, the crossover point was more dependent upon the pulse time, and displayed a complex behavior. This complex behavior was even more pronounced at values of A lower than 0.8% (not shown).
Figure 4. Plot of crossover point as a function of pulse time in FIGE. The legend appears on the lower-right hand corner of the figure. For clarity, the lower left-hand corner of the plot shows an exploded view of the behavior of the crossover point at pulse times 0 s, 0.12 s, and 1.2 s. A constant 3:1 ratio was maintained for the forward and reverse pulse times, respectively. Gels were run at a voltage gradient of 5.71 V cm\(^{-1}\) for 8 h in 0.5 x TBE.
DISCUSSION

For the classes of DNAs studied, we have shown that the electrophoretic mobility of cccDNAs during PFG in FIGE and CHEF systems is influenced by pulse time. In addition, we have observed an interaction between pulse time and agarose concentration in FIGE. Depending on the agarose concentration, the influence of pulse time on \( \mu \) can be great or small. The \( \mu \)'s of small linear DNAs were also affected by these parameters, albeit in a different way. The mobilities of both supercoiled and linear molecules display a sharp minimum as a function of pulse time, in a CHEF apparatus (Fig. 2A and 2B).

The pulse-time-independent mobility of cccDNAs during FIGE in 1.0% TBEM gels observed here (Fig. 3B) is similar to that observed by Mathew et al. (1988a), Beverley (1988), and Hightower et al. (1987) in OFAGE. However, when \( A \) was varied, we show that the mobility of cccDNAs became increasingly pulse-time dependent, especially at higher values of \( A \) (Fig. 3C). Some pulsing regimes actually accelerate the larger supercoiled molecules (for example, a forward pulse time of 36 s, in Fig. 3C). This effect may be similar to that observed by Levene and Zimm (1987) for ocDNAs. Mathew et al. (1988b) found a sharp transition in \( \mu \) at approximately 1.1% agarose for linear DNAs in OFAGE. It is well established in CF
electrophoresis that $\mu_{ccc}$ is greatly affected by $A$ (Dingman et al., 1972; Serwer and Allen, 1984). Interestingly, studies showing pulse-time independence of $\mu_{ccc}$ were done at $A = 1.0\%$, which is within the transition region. Migration of ocDNAs in FIGE have also been shown to be pulse-time dependent (Levene and Zimm, 1987). Whether this field dependence of $\mu_{ccc}$ can be attributed solely to topology and $A$ or whether the reorientation angle of FIGE contributes to pulse-time dependence is unknown. Our CHEF results (Fig. 2A) also suggest pulse-time dependence of $\mu_{ccc}$, although a limited number of pulse times were investigated. The point at which ccc and linear DNAs of the same size co-migrate (crossover point) in FIGE was pulse-time independent at $A = 1.0\%$ (Fig. 4). The crossover point was constant at approximately 8 kb for $P_t$ times between 1.2 s and 120 s. The exact value of the crossover point observed was higher than the 5 kb value reported by Mathew et al. (1988a). However, the pulse-time independence is consistent with observations in OFAGE at $A = 1.0\%$ (Levene and Zimm, 1987; Beverley, 1988). Another observation consistent with OFAGE is that a very short $P_t$ (0.12 s) caused an increase in the crossover point. Mickel et al. (1977) found an inverse relationship between crossover point and $A$ in continuous field electrophoresis. Our data showed a pulse-time-dependent behavior of the crossover point at $A = 1.2\%$ and $A = 0.8\%$, which suggests that there may be a
pulse time/agarose interaction influencing the crossover point (Fig. 4). We also observed a pulse-time-dependent behavior of the crossover point in CHEF at A = 1.0% (Fig. 3B). This is in contrast to OFAGE (Mathew et al., 1988a) for the same value of A. Also in contrast to OFAGE, we found that the lowest values of \( \mu_{ccc} \) in CHEF were produced by a 10-s pulse time (Figs. 1A and B). A 10-s pulse time produced the largest values of \( \mu_{ccc} \) in OFAGE (Mathew et al., 1988a). A resonance-like effect may explain the behavior of \( \mu_{ccc} \) in OFAGE (Mathew et al., 1988a). Further studies are needed to understand the behavior of \( \mu_{ccc} \) in the various PFG systems studied.

Hightower et al. (1987) have shown that cccDNAs undergoing OFAGE display a different size-mobility relationship when compared with linear DNAs. It has been shown that the \( \mu \)'s of cccDNAs between 4 kb and 16 kb are similar to the \( \mu \)'s of linear DNAs between 260 kb and 2600 kb. Our data from ramped FIGE (not shown) show that cccDNAs between 10 kb and 16 kbp display \( \mu \)'s in the same range as linear DNA \( \mu \)'s of 300 kb to 400 kb. Together, these data support a different size-mobility relationship of cccDNAs during PFG as compared with linear DNAs. Clearly, PFG produces effects on cccDNAs that cannot be explained solely by their sizes. Serwer (1988) proposed a gel hysteresis hypothesis to explain why circular DNAs migrate more slowly during PFG than linear DNAs of the same size. According to
this hypothesis, if the reorientation angles are greater than 90° (but not greater than 180°), hair-like projections from the agarose gel fibers cause threading of all DNAs. Threading of linear DNAs is believed to be reversible. Reversible threading may be difficult for circular DNAs, however, and this may reduce the μ's of circular DNAs.

Within the range of conditions studied, a complex interaction is observed between pulse time and agarose concentration. Although generalizations should be viewed as tentative until more extensive studies can be completed, it seems that lower values of A can be used to slightly increase μ\text{ccc} and decrease running times. A pulse-time-independent behavior of μ\text{ccc} can be achieved by using A within the transition region (approximately 1.0 % agarose).

Studies using a wide range of pulse times and agarose concentrations, analysis of Ferguson plots (Ferguson, 1964), and possibly logit transformations (Willis et al., 1988) would be useful to design mathematical equations for cccDNA migration during PFG. Ferguson plots allow the analysis of free electrophoretic mobility (μ_0) and coefficients of retardation (K_x) at various pulse times for different topological isomers (Ferguson, 1964). The study of these variables has produced improvements in the understanding of CF electrophoresis. A recent study using logit and probit analyses has yielded a mathematical equation for the
electrophoretic mobility of linear DNA in CF agarose electrophoresis (Willis et al., 1988). It would be interesting to apply such mathematical methods to PFG gels and cccDNAs. These experiments could provide useful insight into the mechanisms of reorientation during PFG and practical parameters to obtain maximum resolution and gel performance.
B. W. S. Sobral was supported in part by a Brazilian National Research Council (CNPq) fellowship, number 200289/85-AG. Computer resources used to carry our studies were provided by the BIONET National Computer Resource for Molecular Biology which is funded by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health, Grant Number P41RR01685. We thank John Mayfield for the opportunity to interface with BIONET. We thank Renee Kossak and Rhonda Honeycutt for their invaluable suggestions and critical reading of this manuscript.
SECTION II. APPLICATION OF LARGE DNA METHODS TO ANALYSIS OF Rhizobiaceae GENOMES
TITLE: Application of Large DNA Methods to Analysis of 
Rhizobiaceae Genomes

AUTHORS: Bruno W. S. Sobral\textsuperscript{1,2\*}, Rhonda J. Honeycutt\textsuperscript{2}, and Alan 
G. Atherly\textsuperscript{2}.

ADDRESS: Department of Genetics, Iowa State University, Ames, 
Iowa 50011\textsuperscript{2}.

CURRENT ADDRESS: California Institute of Biological Research, 
11099 N. Torrey Pines Rd., La Jolla, California 92037\textsuperscript{1}.

RUNNING TITLE: PFG Studies of Rhizobiaceae Genomes

\*Corresponding author

______________________________________________________________________

Journal Paper No. J- of the Iowa Agriculture and Home 
Economics Experiment Station, Ames, Iowa. Project No. 2891.
The lack of high-resolution genetic or physical maps for the Rhizobiaceae limit our understanding of this agronomically important family. Pulsed-field gel electrophoresis (PFG) in combination with rare-cutting restriction enzymes are powerful tools for generating physical maps. We statistically analyzed Rhizobiaceae DNA sequences from GENBANK release 59.0 to predict frequencies of rare-cutting restriction enzyme sites. Comparisons between various statistical predictors and experimental observations were made. We prepared intact genomic DNA from Rhizobiaceae in agarose plugs and used PFG techniques to compare the genomes of free-living and bacteroid forms of bradyrhizobia, and to investigate the stability of genomic fingerprints produced by digestion with rare-cutting enzymes. No differences were observed between genomic fingerprints of free-living and two developmental stages of bacteroid forms of B. japonicum, when analyzed by field inversion gel electrophoresis (FIGE). B. japonicum genomic fingerprints were unaltered by the nodulation process or after maintenance on rich growth media for up to 100 generations. These data suggest that large-scale DNA rearrangements do not occur during bacteroid development, after bacteroid dedifferentiation, or during maintenance on rich growth media. The genome sizes of representative R. meliloti and B.
japonicum strains were approximately 4.7 Mb and 5.7 Mb, respectively. The results of our work will be useful for the application of PFG techniques to evolutionary studies and physical mapping of Rhizobiaceae genomes.
Physical mapping may be the only tool available for generating maps of regions of *Rhizobiaceae* genomes that are essential to survival of the bacterium or unstable when cloned in *Escherichia coli*. In many fast-growing rhizobia, many symbiotically important genes are localized on large extrachromosomal elements known as megaplasmids (reviewed in Denarie et al., 1981; Huguet et al., 1983). In some rhizobial species, such as *Rhizobium meliloti*, there are strains possessing megaplasmids that are extremely recalcitrant to curing (Huguet et al., 1983; Eberhard, 1989). Some *Agrobacterium* large plasmids are also difficult to cure (Eberhard, 1989). *R. meliloti* is the most thoroughly studied of the *Rhizobiaceae* at the genetic level; however, their genetic maps (Meade and Signer, 1977; Casadeus and Olivares, 1979; Hooykaas et al., 1982; Martin and Long, 1984; Beringer et al., 1987) are of low resolution for most regions of the genome. In *Bradyrhizobium japonicum*, the most thoroughly studied of the bradyrhizobia, there is no complete, circular genetic map of the chromosome at any level of resolution. Molecular maps exist for *nod* and *nif* regions in most species of (brady)rhizobia, yet they typically span no more than 50 kb (Long et al., 1982; Prakash et al., 1982; Schofield et al., 1983; Downie et al., 1983; Kondorosi et al., 1984; Masterson
and Atherly, 1986). Physical maps would be useful to investigators studying Rhizobiaceae because they could be used to assign cloned genes to specific genomic regions as well as to study DNA rearrangements (Smith et al., 1986).

Techniques have been developed that allow the manipulation of intact genomic DNA from various procaryotic and eucaryotic organisms (reviewed in Smith et al., 1988). Research using pulsed-field gel electrophoresis (PFG) has yielded complete physical maps for large regions of mammalian chromosomes (Fulton et al., 1989; Pohl et al., 1988; Shaw, 1986; Wallace et al., 1989), has permitted the resolution of entire chromosomes of lower eucaryotes, such as a variety of fungi (Orbach et al., 1988; Carle and Olson, 1985; Schwartz and Cantor, 1984; Smith et al., 1987b; Vollrath and Davis, 1987; Snell and Wilkins, 1986; Brody and Carbon, 1989) and parasitic protozoa (Van der Ploeg et al., 1984), and has produced physical maps of the entire chromosome of some procaryotes (Smith et al., 1987a; Ventra and Weiss, 1989; Lee and Smith, 1988; Kauc et al., 1989; Canard and Cole, 1989; Bautsch, 1988; Suwant and Kaplan, 1989a, 1989b; Bancroft et al., 1989; Ely and Gerardot, 1988). PFG can also be used to monitor genetic rearrangements (Smith et al., 1986).

The process of bacteroid differentiation and development may occur via DNA rearrangements. The structure and organization of the bacteroid genome, in relation to the
process of bacteroid development, has been an object of great interest and some dispute (Bergersen, 1958; Dilworth and Williams, 1967; Reijnders et al., 1975; Bisseling et al., 1977; Gresshoff et al., 1977; Paau et al., 1977; Tsien et al., 1977; Gresshoff and Rolfe, 1978; Zhou et al., 1985; McDermott et al., 1987; Price et al., 1987; Vierny and Iaccarino, 1989). There is some uncertainty about the relative amount of DNA in bacteroids as compared to their free-living counterparts (Bergersen, 1958; Dilworth and Williams, 1967; Reijnders et al., 1975; Bisseling et al., 1977; Paau et al., 1977; Vierny and Iaccarino, 1989).

We have adapted techniques for the preparation of intact genomic Rhizobiaceae DNA in agarose plugs (Sobral and Atherly, 1989). We now report the analysis of Rhizobiaceae DNA sequences from GENBANK release 59.0. These DNA sequences were grouped by species and used to make predictions about rare-cutting restriction enzymes in Rhizobiaceae. Using statistical methods (Almagor, 1983; Phillips et al., 1987a, 1987b; McClelland, 1985; McClelland et al., 1987), we calculated target site frequencies for 5 restriction endonucleases with AT-rich targets. Predictions were evaluated by digestion of intact Rhizobiaceae genomic DNA embedded in agarose plugs, followed by pulsed-field gel electrophoresis (PFG). At least one restriction endonuclease that cut the genome into ≤30 fragments was identified for each
of the four species studied. Genomic fingerprints produced by digestion with these rare-cutting enzymes were stable in B. japonicum USDA 123 during bacteroid development and dedifferentiation, as well as during free-living growth on rich media. Genomic fingerprints of free-living and bacteroid forms of B. japonicum USDA 122 were compared and no large-scale DNA rearrangements were seen. We computed the genome size for representative B. japonicum and R. meliloti strains by using PFG, optical image analysis, and scanning densitometry. These data will be used to guide physical mapping efforts and for evolutionary studies in Rhizobiaceae.

(Some of these results were presented at the Twelfth Annual North American Symbiotic Nitrogen Fixation Conference, Iowa State University, Ames, IA, July 1989.)
MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Growth Conditions

_{Bradyrhizobium japonicum}_ strains USDA 110, USDA 123, USDA 138, USDA 122, USDA 127, USDA 129, and their respective serotype-specific fluorescent antibodies (FAs) were a gift from Renee Kossak (Department of Genetics, Iowa State University, Ames). All other _B. japonicum_ strains were a gift from Michael Sadowsky (Soil Science Department, University of Minnesota, St. Paul). _Rhizobium meliloti_ strains 1021, AK631, bacteriophage N3 (Martin and Long, 1984), and _Rhizobium leguminosarum_ biovar _trifolii_ ANU 843, were gifts from Sharon Long (Department of Biological Sciences, Stanford University, California). The identity of _R. meliloti_ and _B. japonicum_ strains was confirmed, either by lysis with specific phage (Martin and Long, 1984) or by reaction with strain-specific FAs (Schmidt et al., 1986). _Agrobacterium rhizogenes_ A4 was a gift from James H. Zhou (Department of Genetics, Iowa State University, Ames, IA). Plasmids pEA4-19 (containing _nodD1_ from _Rhizobium fredii_ USDA 191 [Applebaum et al., 1988]) and pCHK12 (containing _nifHD_ from _R. meliloti_ 102F34 [Sadowsky et al., 1987]) were gifts from Edward Applebaum (Smith, Kline, and French, King of Prussia, PA) and Michael Sadowsky (Soil Science Department, University of Minnesota, St. Paul, MN),
respectively. *B. japonicum* was grown in AG medium (Sadowsky et al., 1987) and *R. meliloti*, *A. rhizogenes*, and *R. leguminosarum* biovar *trifolii* were grown in L broth (Maniatis et al., 1982). All *Rhizobiaceae* were grown at 28° C and maintained on agar slants.

Preparation and Restriction Digestion of Genomic DNA in Agarose Plugs

The number of bacteria necessary to give approximately 1 μg DNA per 100-μl agarose plug was calculated by using the following equation:

\[
\text{μg DNA} = (\text{CFU ml}^{-1}) \times (2.5 \text{ chr. cell}^{-1}) \times (1 \text{ μg DNA/X chr.}) \times (Y \text{ ml}),
\]

where X = number of chromosomes in 1 μg of genomic DNA, which is a function of the genome size, Y = number of ml of culture to be embedded, and chr. = chromosomes. We used an estimated genome size of 10 Mb for *B. japonicum* (Carlson et al., 1985) and of 5.5 Mb for *R. meliloti* (Herdman, 1985). Based on these assumptions, we calculated for *B. japonicum* that X = 9.125 x 10^7 chromosomes μg\(^{-1}\). To prepare the plugs, 100 ml of cells were harvested (4,000 x g for 20 min at 4° C) at 10^8 CFU ml\(^{-1}\) and resuspended in 100 ml of Pett IV buffer (PIV [Smith et al., 1988]). After washing and repelleting, the cells were
resuspended in the volume of PIV calculated to give the
correct bacterial density, based on assumed sizes of 5.5 Mb
(A. rhizogenes, R. meliloti, and R. trifolii), and 10 Mb (B.
japonicum). The cells were warmed to 37°C and mixed with an
equal volume of 2% low melting temperature agarose (SeaPlaque,
FMC, Rockland, ME) dissolved in sterile double distilled
water. The agarose was cooled to 40°C before mixing with the
bacteria, then immediately pipetted into 100-μl rectangular
molds (Pharmacia/LKB, Piscataway, NJ). The molds were placed
at 4°C and allowed to solidify for 30 min. Plugs were placed
directly into Escherichia coli (EC) lysis solution (Smith et
al., 1988) that lacked RNase and lysozyme and were incubated
for 24 h at 37°C with gentle agitation (80-100 rpm). EC lysis
solution was removed, and the plugs were placed in EDTA-
Sarkosyl-Pronase E solution (ESPe [Sobral and Atherly, 1989]).
The plugs were incubated in ESPe at 50°C for 48-72 h with
gentle agitation (80-100 rpm). After cell lysis, ESPe was
replaced with freshly made ESPe (1 mg ml⁻¹ pronase E) and the
plugs were stored at 4°C.

B. japonicum bacteroids were isolated from nodules as
described by Price et al. (1987). Glycine max cv. 'Williams'
plants were grown hydroponically as described by Ralston and
Imsande (1983) and were inoculated two weeks after germination
with a suspension of B. japonicum in late log phase (10⁹ CFU
ml⁻¹). The bacteroids were isolated 15 or 35 days after
inoculation (dai). Large peribacteroid units (LPBUs) were
diluted 1:20 in PIV and embedded directly in agarose as
described above for free-living bacteria. The typical yield
of LPBUs from 5 g of five-week-old nodules was sufficient to
make approximately 200 agarose plugs, each containing 1–2 μg
of intact genomic DNA. Bacteroid yields from 15–18 dai
nodules were smaller, so the bacteroids were diluted 1:3.
Small PBUs (SPBUs) were collected separately from the LPBUs
and diluted 1:1 with agarose and cast into molds. Because the
PBUs do not contain cell walls, the EC lysis solution step was
omitted. The quality of DNA in all plugs was tested by
continuous field (CF) electrophoresis at 6–10 V cm⁻¹ in 1%
agarose (Seakem LE, FMC) for 8–10 h; plugs were considered of
good quality if little or no DNA migrated out of the wells.
Lambda concatemers were prepared as described by Sobral and
Atherly (1989) and were used as molecular weight markers.
Yeast chromosomes, purchased from Beckman Instruments Inc.
(Palo Alto, CA) or FMC, were also used as molecular weight
markers.

Preceding restriction enzyme reactions, the plugs were
removed from ESPE and aseptically placed in sterile, 50-ml
disposable centrifuge tubes (Corning Glass Works, Corning, NY)
containing 2–5 ml plug⁻¹ of sterile 10 mM Tris-HCl, pH 8.0, 1
mM Na₂EDTA (TE). The plugs were washed thrice in TE at room
temperature for 2 h with gentle agitation (80–100 rpm). DNA
in agarose plugs could be stored in sterile TE at 4°C for months without observable degradation. To perform restriction enzyme reactions, one-third of a plug was aseptically placed in a sterile 1.7-ml microfuge tube, and 0.5 ml of the appropriate restriction enzyme buffer was added. The potassium glutamate (KGB) restriction buffer devised by McClelland et al. (1988) was used for all enzymes (see Table 4 for appropriate concentrations). Plugs were equilibrated in KGB for 30 min at room temperature. KGB was removed, and 200 μl of freshly prepared KGB, supplemented with 5 mM β-mercaptoethanol (Sigma Chemical Co., St. Louis, MO) and 100 μg ml⁻¹ of nuclease-free bovine serum albumin (Sigma), was added. Restriction enzyme was added at a concentration of 5-10 U μg⁻¹ of DNA. Restriction enzymes were purchased from New England Biolabs (Beverly, MA), Boehringer Mannheim Biochemicals (Indianapolis, IN), Bethesda Research Laboratories (Gaithersburg, MD), or International Biotechnologies Inc. (New Haven, CT). The plugs were incubated with restriction enzyme for 5 h or overnight at the temperature recommended by the manufacturer. Restriction enzymes were inactivated by incubation at 50°C for 10 min. Plugs were either loaded onto a gel immediately, or KGB was removed and 1 ml ESPE (containing 1 mg ml⁻¹ of pronase E) was added. Digested plugs were stored for up to 16 weeks without observable DNA degradation or restriction pattern alteration. Approximately
one-sixth to one-eighth of a plug was loaded in each well to prevent overloading.

**PFG Equipment and Electrophoresis Conditions**

Electrophoresis in contour-clamped homogeneous electric fields (CHEF [Chu et al., 1986]) was done by using a Pulsaphor Plus System equipped with a hexagonal electrode kit (Pharmacia/LKB). Field inversion gel electrophoresis (FIGE [Carle et al., 1986; Lalande et al., 1987]) was controlled by using either a PC750 Pulse Controller (Hoefer Scientific Instruments, San Francisco, CA), a Minipulse Controller (IBI, New Haven, CT), or a FIJI HV600 System (IBI). One percent agarose gels in modified 1 x or 0.5 x TBE (TBEM [Mathew et al., 1988b]) were run in vertical slab units. Sixteen-centimeter-long gels were run in SE600 units (Hoefer Scientific Instruments); SE620 units (Hoefer Scientific Instruments) were used for 32-cm-long gels. Voltage gradients were 8-13 V cm⁻¹. Gel boxes were cooled with 10°-15°C water circulating through a cooling manifold at a flow rate of approximately 1-2 l min⁻¹. Cooling maintained an equilibrium temperature of 13°-18°C during electrophoresis. Table 1 describes the parameters used in the pulsing programs for FIGE gels.
<table>
<thead>
<tr>
<th>Useful Resolution Window (in kbp)</th>
<th>Run Time (h)</th>
<th>PC750</th>
<th>FIJI HV600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IPF</td>
<td>IPR</td>
</tr>
<tr>
<td>A) 1 - 200</td>
<td>15</td>
<td>0.06</td>
<td>0.02</td>
</tr>
<tr>
<td>B) 200 - 800</td>
<td>20</td>
<td>0.30</td>
<td>0.10</td>
</tr>
<tr>
<td>C) 200 - 1600</td>
<td>24</td>
<td>1.20</td>
<td>0.40</td>
</tr>
<tr>
<td>D) 200 - 1000</td>
<td>48</td>
<td>0.30</td>
<td>0.10</td>
</tr>
</tbody>
</table>

*a Controller parameter calculations:

PC750: the following equation was used to calculate $r \ (h^{-1})$:

$$r = \frac{[(FPF/IPF) - 1]}{T},$$

where: $FPF$ is the final pulse forward (s), $IPF$ is pulse forward (s), and $T$ is the run duration (h).

FIJI HV600: the equations below were used:

$$AVT = \frac{[(FPF/IPF)/2] + IPF}{S},$$

where: $AVT$ is the average time per step.

$$S = \frac{t}{AVT},$$

where: $S$ is the number of steps, and $t$ is the run duration (s).

$$FI = \frac{(FPF/IPF)}{S},$$

where: $FI$ is the forward increment per step (s).

Other parameters: $IPR$ is the initial pulse in the reverse direction (s), $RI$ is reverse increment per step (s), and $N$ is the number of times the controller will through the routine before resetting pulse values to their original values.

*b TE means the routine was repeated until the end of the run.
s used in these studies and their useful resolution ranges

<table>
<thead>
<tr>
<th>Controller Parameters^a</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC750</strong></td>
</tr>
<tr>
<td>IPF</td>
</tr>
<tr>
<td>0.06</td>
</tr>
<tr>
<td>0.30</td>
</tr>
<tr>
<td>0.20</td>
</tr>
<tr>
<td>0.30</td>
</tr>
</tbody>
</table>

Calculations:
as used to calculate r (h^-1):
re: FPF is the final pulse forward (s), IPF is the initial run duration (h).
w were used:
where: AVT is the average time per step.
e number of steps, and t is the run duration (s).
FI is the forward increment per step (s).
itial pulse in the reverse direction (s), RI is the
and N is the number of times the controller will cycle tting pulse values to their original values.
repeated until the end of the run.
Identification of Rare-Cutting Restriction Endonucleases

The nucleotide sequences, restriction maps, mono-, di- and trinucleotide frequencies of all sequenced genes and noncoding regions from various Rhizobiaceae were analyzed and downloaded from GENBANK release 59.0 via BIONET. The sequences were then spliced according to species using the Intelligenetics (Palo Alto, CA) SPLICE program available within the SEQ program on BIONET. Splicing was done by uniting the individual sequences head-to-tail with spaces between genes to exclude nucleotide combinations which might not be present in Rhizobiaceae. Mono-, di-, and trinucleotide frequencies of the spliced sequences were calculated by using the Intelligenetics programs BASE-COMPOSITION, DINUCLEOTIDES, and TRINUCLEOTIDES, respectively. Calculation of tetranucleotide frequencies was done by using the SITES program (Intelligenetics). SITES was used to create imaginary restriction enzyme target sites that corresponded to the tetranucleotides whose frequencies we wished to determine.

Based on G+C content of Rhizobiaceae (Jordan, 1984; Herdman, 1985) and on sequence information, the rarity of five AT-rich restriction sites was calculated by using different statistical techniques (Almagor, 1983; Upholt, 1977; Phillips et al., 1987a, 1987b; McClelland, 1985; McClelland et al., 1987). These predictions were used as guidelines to select
the most rare-cutting enzymes. Restriction enzymes were also tested on the basis of trial and error and on differences in methylation sensitivities.

Southern Hybridization

The DNA was depurinated for 15 min at room temperature in 500 ml of 0.25 N HCl, denatured and neutralized as described (Maniatis et al., 1982), and blotted onto nylon membranes (Magnagraph, MSI, Westboro, MA) by using either a Transphor electroblotting unit (Hoefer Scientific Instruments), or by alkaline capillary blotting as recommended by the membrane manufacturer. DNA fragments used as radioactive probes were purified from their vectors on agarose gels (Seakem GTG, FMC) prepared in TAE buffer (Maniatis et al., 1982). Fragments were excised from the gel, purified by using GeneClean (BIO101, La Jolla, CA), and labelled with α-32P dCTP (New England Nuclear, Boston, MA, 3000 mCi mmol⁻¹) by using a Multiprime rapid hybridization system (Amersham Corporation, Arlington Heights, IL). Hybridization conditions, buffers, and washes were according to manufacturer's instructions. Unincorporated nucleotides were removed by chromatography through NickColumns (Pharmacia/LKB, Piscataway, NJ). Autoradiography was done at -70°C with two intensifying screens (Fisher Scientific, Pittsburgh, PA) and X-Omat film
(Kodak, Rochester, New York); typical exposure times were 4-24 h.

Image Analysis and Scanning Densitometry of PFG Gels

Image analysis was done on the photographic negatives of ethidium bromide stained FIGE and CHEF gels by using a densitometrical process employing a Zeiss SEM-IPS image analysis system (Zeiss-Kontron). The negatives were placed on a Chromapro IBAS Configuration Light Box and viewed with a Panasonic WV-CD50 video camera. The negatives were transilluminated by the light box. A program was written to measure and analyze the grey value of the individual DNA fragments on the negatives, one lane at a time. Grey values were used to indicate when multiple fragments were present.

Scanning densitometry of the photographic negatives was done on a Zeineh Soft Laser Scanning Densitometer, model SL-504-XL (Biomed Instruments, Inc., Fullerton, CA), by using a tungsten light source, with the following settings; 7.0 gain, and a 3.0 scan speed (in the high speed range).

Calculation of Sizes of Restriction Fragments on PFG Gels

At least three independent size measurements were taken, and the sizes were rounded to the nearest 5 kb before
summation. Fragments smaller than approximately 50 kb were not considered in our analyses. Molecular weight standards on PFG gels were plotted and regression analysis was done by using Sigma-Plot 3.10 (Jandel Scientific, Corte Madera, CA). Estimated sizes of restriction fragments were calculated by using a HP-41CV (Hewlett Packard, Corvallis, OR) scientific calculator, programmed for solving the polynomial equations generated by regression analysis.
Identification of rare-cutting restriction endonucleases

We analyzed **Rhizobiaceae** genomes by using DNA sequences available on GENBANK release 59.0 (Table 2). Because Rhizobiaceae are GC-rich bacteria (Jordan, 1984) the probability of target sites was calculated for five restriction endonucleases with AT-rich target sites. Predictions about the average size of restriction fragments produced by these five enzymes are shown in Table 3. Predictions were based on G+C molar ratios (Upholt, 1977), mono-, di-, tri-, and tetranucleotide frequencies (Almagor, 1983; Blaisdell, 1985; Phillips et al., 1987a, 1987b; McClelland, 1985; McClelland et al., 1987). Note the differences in predictions made by using different frequencies as generators. The five restriction enzymes shown in Table 3, together with others chosen because of differences in methylation sensitivity or because of previous observations in continuous field gels (not shown), were used to digest **Rhizobiaceae** DNA in agarose plugs. Table 4 shows results of digestion of **Rhizobiaceae** DNA plugs with various restriction enzyme and restriction enzyme plus methylase combinations. Restriction endonucleases which produced the least number of fragments per genome were: AseI, SpeI, and XbaI for A.
FIGURE 1A. Hybridization of nifHD to *R. meliloti* strain 1021 genomic DNA digested with various restriction enzymes and separated by FIGE; lanes 1) lambda concatemers, 2) AseI, 3) SpeI, 4) XbaI, 5) SspI, 6) DraI, 7) lambda concatemers, 8) NotI, 9) SfiI, 10) undigested, and 11) lambda concatemers. Parameter set B was used for this FIGE gel (see Table 1).
rhizogenses A4, R. meliloti 1021 and AK631, and R. leguminosarum bv. trifolii ANU 843; and AseI, DraI, and SpeI for R. japonicum (18 different strains and isolates tested).

To show that the digests produced by these enzymes were complete, and to localize the nod and nif gene regions to fragments produced by rare-cutting restriction enzymes, we hybridized with nodD1 and nifHD gene probes. Results from nifHD hybridizations with DNA from various strains of B. japonicum digested with rare-cutting restriction enzymes have been shown elsewhere (see Section III). Figure 1 shows hybridizations with nifHD (Figure 1A) and nodD1 (Figure 1B) for R. meliloti strain 1021.

Sequential M.ClaI + DpnI Digestions and 5'-GATC-3' Methylation of Rhizobiaceae DNA Plugs

We treated the plugs with M.ClaI followed by DpnI (Weil and McClelland, 1989) because of the very large predicted size of the fragments (Table 3). Since DpnI only is able to digest the sequence 5'-GATC-3' if N4-methyladenine is present, we first digested Rhizobiaceae DNA plugs with DpnI (Figure 2A). No DpnI-specific digestion was observed with B. japonicum, R. leguminosarum bv. trifolii, or R. meliloti DNA.

Interestingly, DpnI digestion of A. rhizogenses A4 DNA gave one large fragment (>100 kb) and a large number of smaller
FIGURE 1B. Hybridization of *nodD1* to *R. meliloti* strain 1021 genomic DNA digested with various restriction enzymes and separated by FIGE; lanes 1) lambda concatemers, 2) *AseI*, 3) *SpeI*, 4) *XbaI*, 5) *SspI*, 6) *DraI*, 7) lambda concatamers, 8) *NotI*, 9) *SfiI*, 10) undigested, and 11) lambda concatamers. Gel was 1% agarose in TBEM and parameter set B was used (see Table 1).
fragments (<50 kb), suggesting that a high degree of digestion occurred (Figure 2A, lane 5). To further investigate the methylation of 5'-GATC-3' in Rhizobiaceae, we digested the various DNA plugs with the restriction enzymes MboI and Sau3AI, two isoschizomers which differ in their methylation sensitivities (Nelson and McClelland, 1989). In B. japonicum USDA 123 and USDA 122, digestion with MboI produced more small fragments than did digestion with Sau3AI, indicating that MboI gave more complete digests (not shown). In B. japonicum USDA 122 and USDA 123 bacteroid DNA plugs, digestion with Sau3AI yielded partial digests, with some larger fragments (not shown). Partialis with Sau3AI were especially apparent for bacteroid plugs prepared from 35-days-after-inoculation (dai) bacteroids; 15-dai bacteroid plugs gave digests more like those from plugs prepared from free-living bradyrhizobia.

Finally, we tested incubation of Rhizobiaceae DNA plugs with M.ClaI alone, in the absence of Mg++ ions (Weil and McClelland, 1989). Surprisingly, under these conditions, some digestion was observed for DNA in plugs from B. japonicum strains and bacteroids, and from R. leguminosarum bv. trifolii (not shown). The digestion patterns produced by treatment with M.ClaI included a background smear and some defined bands, especially for R. leguminosarum bv. trifolii. These results suggest that M.ClaI is contaminated with other nucleases.

Sequential treatment of Rhizobiaceae plugs with M.ClaI
FIGURE 2A.  DpnI digestion of Rhizobiaceae DNA plugs; lanes 1) B. japonicum USDA 123 bacteroids (harvested 35 days after inoculation [dai]), 2) B. japonicum USDA 123, 3) R. meliloti AK631, 4) R. meliloti 1021, 5) A. rhizogenes A4, 6) R. leguminosarum bv. trifolii ANU 843, 7) B. japonicum USDA 122 35 dai bacteroids, 8) B. japonicum USDA 122 15 dai bacteroids, 9) B. japonicum USDA 122, 10) lambda concatemers, and 11) lambda digested with PstI. FIJI HV600 controller parameters for this FIGE gel were: IPF = 0.03 s, IPR = 0.01 s, RI = 0.01, S = 360, and N = 2. The gel was 1% agarose in TBEM and was run for 8 h at E = 12.5 V cm⁻¹.
and DpnI produced some specific, large DNA fragments superimposed on a smeared background (Figure 2B, lanes 3, 4, and 6; and Figure 2C lane 1). Hybridization using a nifHD gene probe to FIGE gels of sequential M.ClaI/DpnI digests of B. japonicum DNA plugs showed a single fragment. Since hybridization also was observed to the smeared region of the gel, the non-specific degradation produced by the treatment also was demonstrated. Unfortunately, we were unable to keep the amount of non-specific degradation low enough to use this strategy for generating genome size estimates.

Stability of Whole Genomic Fingerprints in Bradyrhizobium japonicum

Genome instability is common in some Rhizobiaceae, notably in R. leguminosarum biovar phaseoli (Soberon-Chaves et al., 1986; Flores et al., 1988; Romero et al., 1988; Soberon-Chaves and Najera, 1989). Furthermore, repeated DNA sequences exist in B. japonicum (Kaluza et al., 1985), R. fredii (Masterson and Atherly, 1986; Prakash and Atherly, 1984), and other Rhizobiaceae (Flores et al., 1987). Repeated sequences can lead to genomic instability or genetic rearrangements (Cantrell et al., 1982b; Haughland and Verma, 1981). For these reasons, the stability of genomic fingerprints produced with rare-cutting restriction enzymes was studied by
<table>
<thead>
<tr>
<th>Species</th>
<th>Number of strains</th>
<th>Number of sequences</th>
<th>Number of base pairs</th>
<th>%G+C</th>
<th>Rarest din.</th>
<th>(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agrobacterium sp.</td>
<td>12</td>
<td>27</td>
<td>71,342</td>
<td>51.2</td>
<td>TA (3.(\text{GT} (5.))))</td>
<td>(^b)</td>
</tr>
<tr>
<td>B. japonicium</td>
<td>2</td>
<td>13</td>
<td>19,101</td>
<td>59.2</td>
<td>TA (2.(\text{TT} (4.))))</td>
<td>(^b)</td>
</tr>
<tr>
<td>R. meliloti</td>
<td>5</td>
<td>31</td>
<td>41,880</td>
<td>57.6</td>
<td>TA (2.(\text{TT} (4.))))</td>
<td>(^b)</td>
</tr>
<tr>
<td>R. leguminosarum</td>
<td>1</td>
<td>5</td>
<td>6,886</td>
<td>54.4</td>
<td>TA (2.(\text{GT} (4.))))</td>
<td>(^b)</td>
</tr>
</tbody>
</table>
| biovar trifolii   | \(^{a}\) Retrieved from GENBANK version 59.0 using the Intelligenetics (Pa program, via BIONET.\(^b\) Rarest dinucleotides were calculated by using the Intelligenetics program. The number in parenthesis is the observed frequency (%) of the \(^c\) Rarest trinucleotides were calculated by using the Intelligenetics program. The number in parenthesis is the observed frequency (%) of the
sequence information available on GENBANK

<table>
<thead>
<tr>
<th>Number of sequences</th>
<th>Number of base pairs</th>
<th>%G+C</th>
<th>Rarest din.</th>
<th>Rarest trin.</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>71,342</td>
<td>51.2</td>
<td>TA (3.9)</td>
<td>TAG (0.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT (5.1)</td>
<td>GTA (0.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TAC (0.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CTA (0.9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TAA (1.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ACT (1.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AGT (1.0)</td>
</tr>
<tr>
<td>13</td>
<td>19,101</td>
<td>59.2</td>
<td>TA (2.2)</td>
<td>TTA (0.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT (4.2)</td>
<td>TAA (0.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT (4.7)</td>
<td>TAG (0.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AC (4.9)</td>
<td>GTA (0.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CTA (0.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA (0.6)</td>
</tr>
<tr>
<td>31</td>
<td>41,880</td>
<td>57.6</td>
<td>TA (2.6)</td>
<td>TAA (0.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TT (4.7)</td>
<td>TAG (0.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GTA (0.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TTA (0.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CTA (0.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ATA (0.7)</td>
</tr>
<tr>
<td>5</td>
<td>6,886</td>
<td>54.4</td>
<td>TA (2.6)</td>
<td>TTA (0.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>GT (4.9)</td>
<td>TAG (0.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GTA (0.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TAA (0.6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CTA (0.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TAC (0.7)</td>
</tr>
</tbody>
</table>

K version 59.0 using the Intelligenetics (Palo Alto, CA) QUEST

Dinucleotides and trinucleotides were calculated by using the Intelligenetics DINUCLEOTIDES and TRINUCLEOTIDES. Dinucleotide synthesis is the observed frequency (%) of the dinucleotide. Trinucleotide synthesis is the observed frequency (%) of the trinucleotide.
comparison of plugs made from *B. japonicum* USDA 123 inoculant, bacteroids harvested at 35 dai, streaks from crushed nodules, and free-living bacteria after approximately 20 and 100 generations of growth on AG medium (Figure 3 and data not shown). No differences were observed in the genomic fingerprints produced by digestion of the DNA plugs with the five enzymes listed in Table 3.

Previous studies done on various (brady)rhizobial species (Bergersen, 1958; Bisseling et al., 1977; Zhou et al., 1985) have shown that bacteroids may possess more, less, or the same amount of DNA as free-living (brady)rhizobia, depending on their developmental stage. To further compare genomic fingerprints of bacteroids and free-living *B. japonicum*, 15 and 35 dai bacteroid plugs were prepared from strains USDA 122 and USDA 123. Figure 4 shows a representative FIGE gel comparing the genomic fingerprints of *B. japonicum* USDA 123 and USDA 123 35 dai bacteroids. We observed no reproducible differences in the fingerprints produced with any of the rare-cutting restriction enzymes. This was also true for 15 dai bacteroids (not shown). Complete digests were not always easily obtainable with bacteroid DNA plugs of any developmental stage, especially with DraI (Fig. 4). XbaI was the only enzyme tested that was capable of consistently digesting bacteroid plugs without producing significant smearing (Fig. 4). The reasons for the greater amount of
FIGURE 2B. FIGE gel of sequential treatment of *B. japonicum* genomic DNA with *MmClaI* and *DpnI*; lanes 1) *Saccharomyces cerevisiae* strain 334 chromosomes, 2) lambda concatemers, 3) USDA 123 bacteroids (35 days after inoculation [dai]), 4) USDA 127, 5) USDA 129, and 6) USDA 123. The gel was 1% agarose in TBEM, and was run for 96 h at $E = 5 \text{ V cm}^{-1}$. PC750 pulsing parameters were $\text{IPF} = 6 \text{ s, IPR} = 2 \text{ s, and } r = 0.3 \text{ h}^{-1}$. 
smearing produced after digestion of bacteroid DNA in agarose plugs with \textit{AseI}, \textit{DraI}, and \textit{SpeI} are unknown.

\textbf{Genome Size in \textit{R. meliloti} and \textit{B. japonicum}}

To calculate the genome size of \textit{B. japonicum} and \textit{R. meliloti}, we used 32-cm-long FIGE gels for greater resolution, as well as electrophoresis in CHEF because of the monotonic relationship between migration distance and fragment size that is observed in CHEF systems (Chu et al., 1986). Figure 5A shows a FIGE gel of \textit{B. japonicum} DNA digested with rare-cutting restriction enzymes and Figure 5B shows digests of DNA from other \textit{Rhizobiaceae}. Figure 2C (lane 4) shows a representative CHEF gel of \textit{B. japonicum} USDA 123 DNA digested with \textit{SpeI}. In all gels, note the large number of fragments present in the 50 to 200 kb size range, especially for \textit{B. japonicum} digests (Figure 2C, lane 4; Figure 5A, lanes 2-5 and 8-11). Scanning densitometry was done by using the negatives of photographs of various gels to determine the number of bands present in the 50 to 200 kb region of the gel. The scans were complex and difficult to interpret (not shown), especially when multiple fragments were present. Therefore, image analysis was done to further clarify the number of fragments generated by \textit{SpeI} digestion of \textit{B. japonicum} and \textit{R. meliloti} DNA plugs. For FIGE gels, regression analysis of
TABLE 3. Predictions of average fragment size expected following digestion by restriction endonucleases that recognize AT-rich target sites and M.ClaI/DpnI treatment

<table>
<thead>
<tr>
<th>Species</th>
<th>Restriction endonuclease</th>
<th>Average fragment size (kb), as calculated by:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>G+C</td>
</tr>
<tr>
<td>Agrobacterium</td>
<td>AseI</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>DraI</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>SpeI</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>SspI</td>
<td>4.7</td>
</tr>
<tr>
<td></td>
<td>XbaI</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>M.ClaI/DpnI</td>
<td>1103.3</td>
</tr>
<tr>
<td>B. japonicum</td>
<td>AseI</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td>DraI</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td>SpeI</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>SspI</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td>XbaI</td>
<td>6.6</td>
</tr>
<tr>
<td></td>
<td>M.ClaI/DpnI</td>
<td>1807.4</td>
</tr>
<tr>
<td>R. meliloti</td>
<td>AseI</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>DraI</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>SpeI</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>SspI</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>XbaI</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>M.ClaI/DpnI</td>
<td>1601.1</td>
</tr>
<tr>
<td>R. leguminosarum</td>
<td>AseI</td>
<td>7.1</td>
</tr>
<tr>
<td>biovar trifolii</td>
<td>DraI</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>SpeI</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>SspI</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>XbaI</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>M.ClaI/DpnI</td>
<td>2012.4</td>
</tr>
</tbody>
</table>

\(^{a}\) Target sites for restriction enzymes are: AseI (ATTAAT), DraI (TTTA (ACTAGT), SspI (AATATT), XbaI (TCTAGA), M.ClaI/DpnI (ATCG\textsuperscript{m6}ATCGAT).

\(^{b}\) Fragment size (kb) = \(1/p\), where \(p\) is the probability of a cut site, described in Materials and Methods. Calculations using G+C molar ratios are based on Markov chain analysis of K. aerogenes by Phillips et al. (1987a, 1987b). NC = not calculated.

\(^{c}\) Observed restriction site frequencies were obtained by analysis of R. meliloti R. leguminosarum sequences as described in Materials and Methods. Frequencies were converted to the size of fragments as described in footnote b above. NO = not observed in this study.
of average fragment size expected following digestion with endonucleases that recognize AT-rich target sites and for sequential treatment

<table>
<thead>
<tr>
<th>Restriction Enzyme</th>
<th>Average Fragment Size (kb), as calculated by:</th>
<th>Observed in GENBANK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G+C</td>
<td>MONO</td>
</tr>
<tr>
<td>I</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>II</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>III</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>IV</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>V</td>
<td>4.3</td>
<td>4.3</td>
</tr>
<tr>
<td>DpnI</td>
<td>1103.3</td>
<td>1105.0</td>
</tr>
<tr>
<td></td>
<td>13.9</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>13.9</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
<td>6.7</td>
</tr>
<tr>
<td></td>
<td>13.9</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>6.6</td>
<td>6.7</td>
</tr>
<tr>
<td>DpnI</td>
<td>1807.4</td>
<td>1855.3</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>9.4</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>5.6</td>
</tr>
<tr>
<td>DpnI</td>
<td>1601.1</td>
<td>1472.9</td>
</tr>
<tr>
<td></td>
<td>7.1</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>7.1</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td></td>
<td>7.1</td>
<td>7.1</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>DpnI</td>
<td>2012.4</td>
<td>1302.4</td>
</tr>
</tbody>
</table>

* restriction enzymes are: AseI (ATTAAAT), DraI (TTTAAA), SpeI, XbaI (TCTAGA), M.ClaI/DpnI (ATCG^m ATCGAT).

b) = 1/p, where p is the probability of a cut site, calculated as methods. Calculations using G+C molar ratios are based on other calculations are based on Markov chain analysis, used for E. coli (1987a, 1987b). NC = not calculated.

c) Site frequencies were obtained by analysis of Rhizobiaceae in Materials and Methods. Frequencies were converted to average described in footnote b above. NO = not observed in the sequences
plots of the migration of lambda concatemers yielded third and fourth order polynomial equations that were used to estimate the sizes of the restriction fragments. For CHEF gels, regression analysis yielded first and second order polynomials. Table 5 summarizes the sizes of restriction fragments produced by digestion with SpeI, for B. japonicum strains USDA 424 (Sadowsky et al., 1987) and USDA 123, and for R. meliloti strains 1021 and AK631. The standard deviation for the USDA 424 genome size estimate, based on 4 repetitions of SpeI digests and FIGE analysis, was 501 kb which is less than 9% of the estimated total genome size. About 400 kb of the variability could be accounted for by variation in the estimated sizes of the 3 largest fragments. CHEF- and FIGE-generated estimates were compared for B. japonicum USDA 123 digested with SpeI. Both estimates agreed to within 400 kb of each other (not shown). Once again, the main source of variability was in estimates for the largest fragments in FIGE gels.
FIGURE 2C. Electrophoresis in contour-clamped homogenous electric fields (CHEF) of *B. japonicum* DNA plugs after treatment with SpeI or sequential treatment with M.ClaI and DpnI; lanes 1) USDA 110/M.ClaI+DpnI, 2) *Saccharomyces cerevisiae* strain YPH80 chromosomes, 3) lambda concatemers, and 4) USDA 123/SpeI. The gel was 1% agarose in 0.5 x TBEM and program parameters used for electrophoresis in CHEF were the same as parameter set B (see Table 1).
DISCUSSION

Using Rhizobiaceae DNA sequences available on GENBANK release 59.0 (see Table 2), we made predictions about the target site frequency of five restriction enzymes which have AT-rich target sequences (Table 3). These and other restriction enzymes were tested on agarose plugs containing intact genomic DNA from three different genera of Rhizobiaceae (Table 4). SpeI digested Rhizobiaceae genomes into the fewest number of fragments (Table 4). For B. japonicum strains, AseI was also a very rare-cutting restriction enzyme, as predicted by Markov chain analysis, and the lack of observed AseI sites in the approximately 20 kb of sequences analyzed (Table 3). In general, Markov chains accurately predicted which enzymes were the rarest-cutting. However, all statistical methods used underestimated the average fragment size at least two-fold (compare results shown in Tables 3, 4, and 5).

Generally, G+C molar ratios, and zero-order Markov chains, both of which are based on mononucleotide frequencies and do not consider nucleotide order within the target sequence, were not good predictors and grossly underestimated the average fragment size (Table 3). First-order Markov chains, which are based on dinucleotide frequencies, were better as predictors of average fragment size, as judged by comparison with the observed sites in GENBANK sequences (Table 3) or with actual
TABLE 4. Restriction endonucleases and methylases for PFG analysis of R Genomes

<table>
<thead>
<tr>
<th>Restriction enzyme</th>
<th>Temp. (°C)</th>
<th>Buffer(^b)</th>
<th>R. japonicum</th>
<th>R. trifolii</th>
<th>R. meliloti</th>
<th>A. rhizogenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AfrII</td>
<td>37</td>
<td>1X</td>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>AseI</td>
<td>37</td>
<td>1X</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>NT</td>
</tr>
<tr>
<td>BclI</td>
<td>50</td>
<td>1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>BspMIR(^c)</td>
<td>50</td>
<td>2X</td>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>BstBI</td>
<td>50</td>
<td>1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>BstXI(^c)</td>
<td>50</td>
<td>1X</td>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>2</td>
</tr>
<tr>
<td>ClaI</td>
<td>37</td>
<td>1X</td>
<td>1</td>
<td>1</td>
<td>NT</td>
<td>1</td>
</tr>
<tr>
<td>DraI(^c)</td>
<td>37</td>
<td>1X</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>EarI</td>
<td>37</td>
<td>1X</td>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>HpaI</td>
<td>37</td>
<td>1X</td>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>MClai + DpnI(^c)</td>
<td>30</td>
<td>1X + 1X</td>
<td>4</td>
<td>NT</td>
<td>4</td>
<td>NT</td>
</tr>
<tr>
<td>Mdam + ClaI</td>
<td>37</td>
<td>1X + 1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>Mdam + XbaI</td>
<td>37</td>
<td>1X + 1X</td>
<td>4</td>
<td>3</td>
<td>NT</td>
<td>3</td>
</tr>
<tr>
<td>MTagI + DpnI</td>
<td>50/30</td>
<td>1X + 1X</td>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>NotI</td>
<td>37</td>
<td>2X</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>NT</td>
</tr>
<tr>
<td>RsrlII</td>
<td>37</td>
<td>0.5X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>SacI</td>
<td>37</td>
<td>1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>SfiI</td>
<td>50</td>
<td>1X</td>
<td>NT</td>
<td>NT</td>
<td>1</td>
<td>NT</td>
</tr>
<tr>
<td>SmaI</td>
<td>25</td>
<td>1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>SpeI</td>
<td>37</td>
<td>1X</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>SspI</td>
<td>37</td>
<td>0.5X</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>TthlIII</td>
<td>50</td>
<td>1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>XbaI</td>
<td>37</td>
<td>1X</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>XmnI</td>
<td>37</td>
<td>1X</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
</tbody>
</table>

\(^a\) Key to numbers presented in table:
1, majority of fragments <50 kb; 2, majority of fragments <100 kb; 3, most fragments (<200 kb), but some larger fragments also; and 4, ≤30 fragments genome were observed. The enzymes in categories 3 and 4 are the most use mapping.

\(^b\) Composition of restriction buffers:
1X = 1 X KGB = 2 mM β-mercaptoethanol; 100 g ml\(^{-1}\) BSA; 100 mM potassium gel Tris-HCl (pH 7.4); 10 mM magnesium acetate.
1X + 1X = 1 X KGB + 1 X KGB without 10 mM magnesium acetate.
2X KGB for NotI = 2 X KGB supplemented with 0.01% Triton X-100.

\(^c\) Complete digests were difficult to obtain. For DraI, this could be purchasing the enzyme from Boehringer Mannheim Biochemicals (Indianapolis IN).
Cleases and methylases for PFG analysis of Rhizobiaceae

<table>
<thead>
<tr>
<th></th>
<th>B. japonicum</th>
<th>R. trifolii</th>
<th>R. meliloti</th>
<th>A. rhizogenes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>NT</td>
<td>4</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>NT</td>
<td>3</td>
<td>NT</td>
</tr>
<tr>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>1</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>2</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>3</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>4</td>
<td>NT</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

*Defined in table: 1, approximately 100 kb; 2, majority of fragments <100 kb; 3, mostly small, larger fragments also; and 4, ≤30 fragments per complete genome. Enzymes in categories 3 and 4 are the most useful for physical mapping.*

Buffer conditions:
- 50% ethanol; 100 mM Tris pH 8.0; 100 mM potassium glutamate; 0.5 mM magnesium acetate.
- Without 10 mM magnesium acetate.
- Supplemented with 0.01% Triton X-100.
- Difficult to obtain. For DraI, this could be avoided by obtaining enzymes from New England BioLabs (Indianapolis,
digests (Table 4), yet they did not always predict which enzymes would produce the fewest fragments. Second- and especially third-order Markov chains were better predictors of enzymes that produced the fewest fragments, but they generally underestimated the average fragment size at least ten-fold (Tables 3 and 4). Our results are in general agreement with similar analyses of other bacterial sequences (Phillips et al., 1987a, 1987b; McClelland et al., 1987), and reiterate the usefulness of Markov chain analysis in showing higher order sequence asymmetry (Elton, 1975; Algamor, 1983; Blaisdell, 1985, Rogerson, 1989) and in predicting site frequencies (McClelland et al., 1987; Phillips et al., 1987a, 1987b) in genomic DNAs.

Digestion of Staphylococcus aureus genomic DNA in agarose plugs with DpnI, after in vitro methylation with M.ClaI, has been used to produce very large DNA fragments (Weil and McClelland, 1989). As can be seen from Table 3, the predicted fragment sizes for Rhizobiaceae genomic DNA treated with M.ClaI/DpnI is quite large. Attempts to produce large fragments with this strategy revealed information about methylation patterns in Rhizobiaceae DNAs. For example, A. rhizogenes strain A4 DNA was digested by DpnI (lane 5, Fig. 2A), suggesting that N⁶-methyladenine in the sequence 5'-GATC-3' must occur frequently in this genome. 5'-G⁶ATC-3' must be rare in other Rhizobiaceae genomes. Different methylation
FIGURE 3. FIGE gel of *B. japonicum* USDA 123 genomic fingerprints; lane 1) lambda concatemers, lanes 4), 8), and 12) inoculum, lanes 3), 7), and 11) isolated from mature soybean nodules, lanes 2), 6), and 10) nodule isolate after 20 generations of growth on AG medium, lane 13) lambda concatemers, and lane 14) *Saccharomyces cerevisiae* strain 334 chromosomes. Restriction enzymes used to digest the DNA are indicated below the lanes. The gel was 1% agarose in TBEM and pulsing parameter set B was used (see Table 1).
Table 5. Sizes of SpeI-generated restriction fragments from *Bradyrhizobium japonicum* USDA 424 and USDA 123, and *R. meliloti* 1021 and AK631 genomic DNAs

<table>
<thead>
<tr>
<th>Fragment number</th>
<th>Fragment size (kb)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>USDA 424</td>
</tr>
<tr>
<td>1</td>
<td>805</td>
</tr>
<tr>
<td>2</td>
<td>620</td>
</tr>
<tr>
<td>3</td>
<td>575</td>
</tr>
<tr>
<td>4</td>
<td>450</td>
</tr>
<tr>
<td>5</td>
<td>270</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
</tr>
<tr>
<td>7</td>
<td>230</td>
</tr>
<tr>
<td>8</td>
<td>225</td>
</tr>
<tr>
<td>9</td>
<td>215</td>
</tr>
<tr>
<td>10</td>
<td>210</td>
</tr>
<tr>
<td>11</td>
<td>205</td>
</tr>
<tr>
<td>12</td>
<td>200</td>
</tr>
<tr>
<td>13</td>
<td>195</td>
</tr>
<tr>
<td>14</td>
<td>195</td>
</tr>
<tr>
<td>15</td>
<td>185</td>
</tr>
<tr>
<td>16</td>
<td>165</td>
</tr>
<tr>
<td>17</td>
<td>150</td>
</tr>
<tr>
<td>18</td>
<td>140</td>
</tr>
<tr>
<td>19</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>90</td>
</tr>
<tr>
<td>21</td>
<td>65</td>
</tr>
<tr>
<td>22</td>
<td>55</td>
</tr>
<tr>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>24</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>5665</td>
</tr>
<tr>
<td>26</td>
<td></td>
</tr>
</tbody>
</table>

* Fragments smaller than 50 kb were not included in these computations, and individual fragment sizes were rounded to the nearest 5 kb and were estimated as described in Materials and Methods.
patterns were also observed between free-living and bacteroid forms of *B. japonicum* (not shown). Methylation of cytosine during bacteroid development would explain our observations with *Mbo*I and *Sau*3AI digests of genomic DNA of free-living and bacteroid forms of *B. japonicum*. *Sau*3AI is insensitive to N⁶-methyladenine, yet cannot cut the DNA if the 3' cytosine is methylated at the N⁶ or N⁷ positions (Nelson and McClelland, 1989). We observed partial digestion of genomic DNA from bacteroids with *Sau*3AI, especially with DNA from 35 dai bacteroids. An alternative explanation is that bacteroid DNA plugs may be more difficult to digest than DNA plugs from free-living bacteria, due to unknown contaminants. Difficulty in digestion of bacteroid DNA plugs was also observed with *Ase*I, *Dra*I, and *Spe*I (Fig. 4). The methylation sensitivities of *Ase*I and *Dra*I have not been determined (Nelson and McClelland, 1989); however, it is known that *Spe*I cannot cut 5'-m⁶ACTAGT-3', and *Xba*I cannot cut 5'-TCTAGm⁵A-3', 5'-Tm⁵CTAGA-3', or Tm⁵CTAGA-3' (Nelson and McClelland, 1989). Because *Xba*I completely digested bacteroid DNA (Fig. 4), widespread cytosine methylation probably does not occur during bacteroid development, at least at the N⁵ position.

Bacteroid forms of *B. japonicum* were further compared to free-living forms by using FIGE and rare-cutting restriction enzymes to investigate a possible role for DNA rearrangements in bacteroid development and dedifferentiation. PFG analysis
FIGURE 4. Comparison of whole genomic fingerprints of free-living and bacteroid (35 days after inoculation [dai]) forms of B. japonicum USDA 123; lanes 1) bacteroids/DraI, 2) 123/DraI, 3) Saccharomyces cerevisiae strain 334 chromosomes, 4) lambda concatemers, 5) bacteroids/Spel, 6) 123/Spel, 7) bacteroids/XbaI, 8) 123/XbaI, 9) bacteroids/Asei, 10) 123/Asei, 11) lambda concatemers, and 12) lambda monomer. The gel was 1% agarose in 0.5 x TBEM and FIGE pulsing parameter set B was used (see Table 1).
of large restriction fragments is a powerful tool to detect rearrangements because of the relatively small number of fragments that need to be analyzed (Smith et al., 1986). Digestion of DNA plugs with AseI, DraI, SpeI, SspI, and XbaI did not indicate that large-scale genomic rearrangements had occurred during bacteroid development. We cannot rule out the possibility of relatively small rearrangements occurring within large restriction fragments, or even within smaller fragments (30 to 200 kb) in regions of the gel where a large number of fragments exist. We also cannot rule out simple inversions (reviewed in Craig, 1985). Southern hybridization studies with appropriate probes would be required to detect such rearrangements. Our present observations rule out large-scale DNA rearrangements as a mechanism of bacteroid development. However, changes in ploidy level, which would not be detected by our analyses, cannot be ruled out. Changes in ploidy during bacteroid development have been reported (Bergersen, 1958; Dilworth and Williams, 1967; Bisseling et al., 1977; Paau et al., 1977; Vierny and Iaccarino, 1989) and could conceivably play a role in bacteroid development, although other levels of developmental regulation, such as protein phosphorylation (Karr and Emerich, 1989), K+ ions (Gober and Kashket, 1987), and low oxygen tension (Gober and Kashket, 1987, 1989) probably control the developmental process.
Figure 5A. Representative 32-cm-long FIGE gel for genome size estimation of *B. japonicum*; lanes 1) lambda concatemers, 2) USDA 123/AsnI, 3) USDA 123/AseI, 4) USDA 123/SpeI, 5) USDA 123/DraI, 6) *Saccharomyces cerevisiae* strain YPH80 chromosomes, 7) lambda concatemers, 8) USDA 127/AsnI, 9) USDA 127/AseI, 10) USDA 127 SpeI, 11) USDA 127/DraI, 12) YPH80 chromosomes, 13) lambda concatemers. The gel was 1% agarose in TBEM and was run using FIGE parameter set D (see Table 1) at $E = 8 \text{ V cm}^{-1}$. 
We estimated the *R. meliloti* genome at approximately 4.7 Mb (Table 5). This value is in agreement with estimates based on reassociation kinetics (Crow et al., 1981; Chakrabarti et al., 1984; Herdman, 1985), yet it still leaves questions unresolved. Electron microscopy (EM) has been used to measure contour lengths of *R. meliloti* megaplasmids (Burkardt and Burkardt, 1984; Burkardt et al., 1987). These EM measurements estimate the size of the two megaplasmids present in strain 2011 (a sister strain of 1021) and the megaplasmid of strain MVII/1 at approximately 1.5 Mb each. The chromosome of *R. meliloti* 2011 was measured at more than 3 times the size of the megaplasmids (Burkardt et al., 1987). These EM measurements suggest a total genome size for *R. meliloti* 1021 that is much larger than values obtained by reassociation kinetics (Crow et al., 1981; Chakrabarti et al., 1984; Herdman, 1985) or from our results. We have not been able to unambiguously resolve intact, undigested megaplasmids from the *R. meliloti* 1021 genome, so we cannot estimate the size of each of the three replicons individually. We must conclude, however, that the size of the chromosome is approximately the same as the size of the megaplasmids, because we know from Southern hybridization data (Figure 4 and not shown) that pRm1021a and pRm1021b have been digested and that the summation of the fragment sizes yielded a total genome size of approximately 4.7 Mb (Table 5). Therefore, our size estimate
Figure 5B. Representative 32-cm-long FIGE gel for genome size estimation of \textit{R. meliloti}; lanes 1) lambda monomer plus lambda/PstI digest, 2) lambda concatemers, 3) \textit{R. meliloti} 1021/\textit{SpeI}, 4) \textit{R. meliloti} AK631/\textit{SpeI}, 5) \textit{A. rhizogenes} A4/\textit{SpeI}, 6) \textit{B. japonicum} USDA 122/\textit{SpeI}, 7) \textit{R. leguminosarum} bv. \textit{trifolii} ANU 843/\textit{SpeI}, 8) \textit{B. japonicum} USDA 123/\textit{SpeI}, and 9) lambda concatemers. The gel was 1\% agarose in TBEM and was run using FIGE parameter set D at E = 8 V cm$^{-1}$ (see Table 1).
for the *R. meliloti* chromosome (1.5 to 2.0 Mb) is in agreement with the size estimates suggested by reassociation kinetics studies (Chakrabarti et al., 1984; Herdman, 1985), when we use the EM measurements (Burkardt and Burkardt, 1984; Burkardt et al., 1987) as estimates for the sizes of the megaplasmids. Our FIGE-produced *R. meliloti* genome size estimate is not, however, in agreement with genetic data about Tn5 transfer frequencies into the chromosome as compared with the megaplasmids (Hynes et al., 1989), or with EM measurements of the chromosome of *R. meliloti* (Burkardt et al., 1987). EM measurements and genetic data are in agreement with each other and suggest that the *R. meliloti* chromosome is at least 4.5 Mb in size (Burkardt et al., 1987; Hynes et al., 1989). The genetic data assume a random insertion of Tn5 throughout the *Rhizobium* genome, which is generally accepted as valid although it has not been rigorously proven. If the chromosome is approximately the same size as the megaplasmids, as we are forced to conclude, why does it not migrate in the "in well lysis" gels (Huguet et al., 1983)? In fact, Hynes et al. (1985, 1986) observed the circular, covalently closed (ccc) form of the *Agrobacterium tumefaciens* chromosome in such gels. Also, because of the approximately equal sizes of the chromosome and megaplasmids, co-migration of these replicons could occur. Recent papers by Suwanto and Kaplan (1989a, 1989b) have clearly shown the existence of two unique
chromosomes in Rhodobacter sphaeroides 2.4.1. The authors argued that the smaller replicon (914 kb) could not be considered a large plasmid because it carried rRNA cistrons and the gene for glyceraldehyde-3-phosphate dehydrogenase, both of which are essential for normal growth. It is possible that complex genomic architecture may be a common feature of soil bacteria. This would certainly be interesting from an evolutionary standpoint.

The genome of B. japonicum USDA 424 was estimated as approximately 5.7 Mb in size, and USDA 123 was estimated as 4.9 Mb (Table 5). Our estimates are approximately half of the size of previous estimates made by reassociation kinetics (Chakrabarti et al., 1984), or as an unpublished observation (Carlson et al., 1985). We have no explanation for this deviation. Lee and Smith (1988), studying Haemophilus influenzae, showed that FIGE-derived genome size estimates were more reproducible than estimates derived by EM contour length measurement of the chromosome (MacHattie et al., 1965) or by determination of the amount of DNA per cell by colorimetric assays (Berns and Thomas, 1965; Zoon and Scocca, 1975). The estimates of the H. influenzae genome size produced by methods other than PFG varied by more than 100%. Also, the genome size of R. sphaeroides 2.4.1 is 4.4 Mb (Suwanto and Kaplan, 1989a), and the genome size of Anabaena sp. PCC 7120 is 6.4 Mb (Bancroft et al., 1989). Both of these
bacteria are capable of photosynthesis and nitrogen fixation, suggesting that their genomes may be some of the largest eubacterial genomes known. We therefore think it unlikely that *B. japonicum* would possess a genome that is almost twice that of *Anabaena*, as suggested by reassociation kinetics (Chakrabarti and Chakrabarty, 1984).
ACKNOWLEDGMENTS

B. W. S. Sobral was supported in part by a fellowship from the Brazilian National Research Council (CNPq), reference number 20089/85-3AG. Computer resources used to carry our studies were provided by the BIONET National Computer Resource for Molecular Biology which is funded by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health, Grant Number P41RR01685. We thank John Mayfield for access to BIONET, Carol Jacobson and Steve VanderWeil for image analysis, Marit Nielsen-Hamilton for help with scanning densitometry, Michael Weil and Michael McClelland for tips on M.ClaI/DpnI digestions, Peter Pattee for stimulating conversations, and Renee Kossak for FAs, for performing FA reactions, and for critically reading this manuscript.
SECTION III. GENOME ANALYSIS OF *Bradyrhizobium japonicum* SEROCLUSTER 123 FIELD ISOLATES BY USING FIELD INVERSION GEL ELECTROPHORESIS (FIGE)
TITLE: Genome Analysis of *Bradyrhizobium japonicum* Serocluster 123 Field Isolates by using Field Inversion Gel Electrophoresis (FIGE).

AUTHORS: Bruno W. S. Sobral*1,3, Michael J. Sadowsky2, and Alan G. Atherly1

ADDRESSES: Department of Genetics, Iowa State University, Ames, Iowa 500111, and Soil Science Department, University of Minnesota, St. Paul, Minnesota 551082.

CURRENT ADDRESS: California Institute of Biological Research, 11099 N. Torrey Pines Rd., La Jolla, California 920373.

*Corresponding author.

Eleven *Bradyrhizobium japonicum* field isolates belonging to serocluster 123 (which is composed of serogroups 123, 127, and 129) were analyzed by using Field Inversion Gel Electrophoresis (FIGE). Genomic fingerprints, which we called Field Inversion Gel Electrophoretic Types (FIGETs), produced by four rare-cutting restriction enzymes showed that the isolates were genetically diverse at the DNA level. Upon analysis of any restriction digest produced by rare-cutting enzymes, few isolates (30-50%) exhibited the same banding pattern as the USDA serogroup strain with which they are antigenically related. Similarities in FIGETs was not correlated with nodulation phenotypes on nodulation restrictive soybean hosts or with the geographical origin of the isolates. However, the isolates could be classified according to their FIGETs and FIGET classification was correlated with serological classification. Hybridizations with a *nifHD* gene probe revealed restriction fragment length polymorphisms (RFLPs) within members of serocluster 123 and provided further evidence of the relatedness between members of serogroups 123 and 127.
Pulsed-Field Gel Electrophoresis (PFG) has increased the resolution limit of agarose gel electrophoresis almost three orders of magnitude (Chu et al., 1986; Carle and Olson, 1984; Schwartz and Cantor, 1984). Field inversion gel electrophoresis (FIGE [Carle et al., 1986]) is a form of PFG that uses periodic inversion of the electric field and can be done by using conventional agarose gel electrophoresis apparatuses. The upper size limit for most FIGE separations is approximately 2 megabases (Mb [Heller and Pohl, 1989]). Since known bacterial genomes are smaller than 8 Mb (Herdman, 1985), the resolution limitations currently imposed by FIGE usually are not a constraint for FIGE analyses of large restriction fragments of bacterial genomes. FIGE has been used to generate size estimates for procaryotic genomes (Canard & Cole, 1989; Bautsch, 1988; Kauc et al., 1989; Lee and Smith, 1988) and, as with other PFG techniques, FIGE can be used to differentiate between different strains of the same species (Smith et al., 1986), or to monitor genetic rearrangements (Smith et al., 1986).

*Bradyrhizobium japonicum* is a Gram-negative soil bacterium capable of fixing atmospheric dinitrogen in symbiotic association with its legume host, *Glycine max* (L.) Merr. In most midwestern soils in the United States, members
of serocluster 123 are the most competitive in nodulating field-grown G. max cultivars (Cregan and Keyser, 1986; Hickey et al., 1987; Moawad et al., 1984; Sadowsky et al., 1987). A large amount of genetic diversity is present among serocluster 123 isolates (Hickey et al., 1987; Moawad et al., 1984; Sadowsky et al., 1987; Schmidt et al., 1986).

We used FIGE, in combination with rare-cutting restriction endonucleases, to fingerprint individual B. japonicum serocluster 123 isolates. Restriction enzymes that had fewer target sites per genome allowed a more easily discernible restriction fingerprint than commonly used enzymes that produce many bands per genome. Genomic fingerprints could be compared directly from FIGE gels, without the need for Southern hybridizations, and provided greater resolution than has been previously possible. The inter-isolate diversity was observed at the level of Field Inversion Gel Electrophoretic Types (FIGETs) indicated that genetic diversity within serocluster 123 may be greatly underestimated by conventional methodologies (Hickey et al., 1987; Moawad et al., 1984; Roberts et al., 1980; Sadowsky et al., 1987; Schmidt et al., 1986). Isolates from the same serogroup could be subdivided on the basis of FIGETs and, with some enzymes, on the basis of restriction fragment length polymorphism (RFLP) hybridization classes by using a nifHD gene probe. Both FIGET- and RFLP-based subdivisions correlated well with
serological relatedness of the isolates.

(Some of these results were presented at the Twelfth Annual North American Symbiotic Nitrogen Fixation Conference, Iowa State University, Ames, IA, July 1989.)
MATERIALS AND METHODS

Bacterial Strains, Media, and Growth Conditions

The B. japonicum isolates with USDA designations were described previously (Keyser and Cregan, 1987; Sadowsky et al., 1987) and were obtained from the culture collection of the U.S. Department of Agriculture, Beltsville, MD. Isolates designated by two-letter postal code abbreviations (IA-35, IA-44, IA-67, IN-78, IN-79, OH-6, and OH-13) were isolated from soybean root nodules obtained from field soils collected from 3 midwestern U.S. states; Iowa, Indiana, and Ohio. Glycine max cv. 'Williams' was the host, and plants were grown in pots as described (Cregan et al., 1989). B. japonicum were isolated from nodules as described (Vincent, 1970). Single colony nodule isolates were purified by streaking twice on yeast extract mannitol (YEM) agar medium (Vincent, 1970). The identities of all field isolates and serogroup strains was confirmed by using serogroup-specific fluorescent antibodies (FAs, kindly performed by Renee Kossak, Iowa State University, Ames), produced as described by Schmidt et al. (1986). The isolates were grown in arabinose-gluconate (AG) medium (Sadowsky et al., 1987) at 28°C and maintained on AG agar slants.
Plant assays were done in Monmouth fine sandy loam soil limed to a neutral pH with CaMg(CO₃)₂. This soil contained very low numbers of B. japonicum strains (<1 B. japonicum cell g⁻¹ soil) as determined using the most probable number plant infection method (Vincent, 1970). Plants were grown in 17.5-cm, surface-sterilized plastic pots containing 2.4 kg of soil (Cregan and Keyser, 1986). Pots were planted with four seeds each of G. max PI 377578 (a serogroup 123 nodulation restricting genotype [Cregan and Keyser, 1986; Keyser and Cregan, 1987]) and cv. 'Williams' (a non-restrictive host). Seeds were surface sterilized (Vincent, 1970) before planting and inoculated with 1 ml (about 10⁹ CFU seed⁻¹) of stationary phase B. japonicum cultures. All cultures were inoculated in triplicate. After inoculation, the seeds were covered and a 1-cm layer of sterile gravel was layered on the soil surface. Plants were grown in the greenhouse as described (Cregan and Keyser, 1986) and thinned three days after emergence to two seedlings of each genotype per pot. Plants were watered with nitrogen-free nutrient solution (Cregan and Keyser, 1986) as needed and harvested 35 days after inoculation. Nodule numbers were determined and isolates producing a nodule mass and number equal to that of strain USDA 123 were classified as restricted (R) for nodulation by the PI genotypes. Those with
nodule numbers and mass statistically equal to strain USDA 430 (previously designated strain MN1-1c [Cregan et al., 1989]) were designated as not restricted (U) for nodulation.

Preparation and Restriction Digestion of Intact Genomic DNA in Agarose Plugs

Fifty-milliliter cultures of *B. japonicum* were grown to a density of approximately $1 \times 10^8$ CFU ml$^{-1}$, were harvested by centrifugation at 8,000 x g for 10 min at 4°C. The cell pellet was suspended in 100 ml of Pett IV (PIV) buffer (Smith et al., 1988), sedimented as before, resuspended in 13.7 ml PIV, and warmed at 37°C for 10 min. An equal volume of a 2% (w/v) aqueous solution of low-melting agarose (SeaPlaque, FMC, Rockland, ME), cooled to 40°C, was mixed with the prewarmed cells; 100-μl portions were transferred to rectangular molds (Pharmacia/LKB, Piscataway, NJ). The molds were placed at 4°C and allowed to solidify for 30 min. The plugs were placed directly into *Escherichia coli* (EC) lysis solution (Smith et al., 1988), and incubated for 24 h at 37°C with gentle agitation (80-100 rpm). The EC lysis solution was removed, the plugs were placed in EDTA-Sarkosyl-Pronase E buffer solution (ESPe [Sobral and Atherly, 1989]), and incubated at 50°C for 48-72 h with gentle agitation (80-100 rpm). The plugs were stored at 4°C in ESPe (1 mg ml$^{-1}$ pronase E) until
restriction endonuclease digestion. The quality of agarose-DNA plugs was tested by continuous field (CF) electrophoresis at 6-10 V cm⁻¹ in 1% agarose (Seakem LE, FMC) for 8-10 h. They were considered of good quality if little or no DNA migrated from the wells into the gel.

Preceding restriction enzyme digestions, the plugs were removed from ESPE and aseptically placed in sterile 50-ml disposable centrifuge tubes (Corning Glass Works, Corning, NY) containing sterile 10 mM Tris-HCl, pH 8.0, 1 mM Na₂EDTA (TE) at a ratio of 2-5 ml TE buffer plug⁻¹. The plugs were washed thrice in TE at room temperature for 2 h with gentle agitation (80-100 rpm). Plugs could be stored in sterile TE at 4°C for months without observable degradation.

For restriction enzyme reactions, one-third of an agarose-DNA plug was aseptically placed in a sterile 1.7-ml microcentrifuge tube, and 0.5 ml of potassium glutamate (KGB [McClelland et al. 1988]) restriction enzyme buffer was added. Plugs were equilibrated in KGB for 30 min at room temperature. The KGB buffer was removed, and 200 μl of freshly-prepared KGB, supplemented with 5 mM β-mercaptoethanol (Sigma Chemical Co., St. Louis, MO) and 100 μg ml⁻¹ of nuclease-free bovine serum albumin (Sigma), was added. The restriction enzyme was added at a final concentration of 5 U μg⁻¹ of DNA. The restriction enzymes were purchased from New England Biolabs (Beverly, MA), Boehringer Mannheim Biochemicals (Indianapolis,
IN), Bethesda Research Laboratories (Gaithersburg, MD), or International Biotechnologies Inc. (New Haven, CT). The plugs were incubated with restriction enzymes for 5-8 h at temperatures recommended by the manufacturer. After the incubation period, the buffer was removed and 1 ml of ES was added, followed by incubation for 10 min at 50°C. The ES was removed, 1 ml of ESPe (containing 1 mg ml⁻¹ of pronase E) was added, and the digested plugs were stored for up to 16 weeks without observable DNA degradation or restriction pattern alteration. For FIGET analysis approximately one-sixth to one-eighth of a plug was loaded into each well. This amount was increased to one-third of a plug for Southern hybridizations.

FIGE Equipment and Electrophoresis Conditions

FIGE was controlled by using either a PC750 Pulse Controller (Hoefer Scientific Instruments, San Francisco, CA), a Minipulse Controller (IBI) or a FIJI HV600 System (IBI). The agarose gels were run in vertical slab units. Sixteen-centimeter-long gels were run in SE600 units (Hoefer Scientific Instruments), and cooled with 10°C water at a flow rate of approximately 1-2 l min⁻¹. This maintained the equilibrium temperature in the gel at 13°C during electrophoresis. The following program parameters were used
for the different controllers: A) FIJI HV600 and Minipulse; initial forward pulse $P_{fi} = 0.3$ s, with a forward pulse increment of 0.006 s, for a total of 4753 steps and a 20 h electrophoresis period; the initial reverse pulse ($P_{ri}$) was 0.1 s, and a constant 3:1 ratio ($P_{fi}:P_{ri}$) was maintained during electrophoresis. B) PC750 controller; $P_{fi}$ started at 0.3 s and a constant 3:1 ($P_{fi}:P_{ri}$) ratio was maintained, with a ramping factor ($r$) of 5 h$^{-1}$; all gels were composed of 1% agarose (Seakem LE, FMC), prepared in 1 x modified TBE (TBEM) (13) and run at a 10-V cm$^{-1}$ voltage gradient. These conditions resulted in good resolution of DNA fragments of up to approximately 0.8 Mb (see Section II). Lambda concatemers (Sobral and Atherly, 1989) and yeast chromosomes, purchased from Beckman Instruments Inc. (Palo Alto, CA) or FMC, were used as molecular weight markers.

Southern Hybridizations

The DNA in agarose gels was depurinated for 20 min at room temperature in 20 volumes of 0.25 N HCl, denatured and neutralized as described (Maniatis et al., 1982), and transferred onto nylon membranes (Magnagraph, MSI, Westboro, MA) by using a Transphor electroblotting unit (Hoefer Scientific Instruments), or by capillary blotting (Maniatis et al., 1982). DNA fragments used as radioactive probes were
purified twice from their vectors on agarose (Seakem GTG, FMC) gels in TAE buffer (Maniatis et al., 1982). The gels were stained briefly with ethidium bromide (100 ng ml⁻¹), and DNA fragments were visualized on a UV box (Fotodyne, New Berlin, WI) and then excised and purified by using a GeneClean kit (BIO101, La Jolla, CA). The probes were labelled with α⁻³²P dCTP (New England Nuclear, Boston, MA, 3000 mCi mmol⁻¹) by using a random-primer labelling kit (Boehringer Mannheim Biochemicals, Indianapolis, IN). Unincorporated nucleotides were removed by chromatography through NickColumns (Pharmacia/LKB, Piscataway, NJ). Approximately 12 ng of radiolabelled probe was added per 200 cm² of membrane surface area. After the membranes were prehybridized for 2-8 h in 80 μl cm⁻² of aqueous prehybridization buffer (Anderson and Young, 1985), the prehybridization buffer was replaced with 50 μl cm⁻² of freshly prepared aqueous hybridization solution (Anderson and Young, 1985) containing a radiolabelled probe. Hybridization continued for 1 x C₀t₁/₂ (Anderson and Young, 1985) or 24 h. Autoradiography was performed at room temperature by using X-Omat film (Kodak, Rochester, New York). Typical exposure times were 8-48 h. Membranes were stored for 6-8 months at 4°C without observable deterioration in hybridization quality.
Size Estimation of DNA Fragments

Size estimation of DNA fragments produced after restriction digestion and FIGE was performed through BIONET by using the Intelligenetics (Palo Alto, CA) SIZER program. DNA sequence data were downloaded from BIONET to analyze restriction sites in sequenced genes from \textit{B. japonicum}. Sequence analysis was performed by using the Intelligenetics SITE program.
RESULTS AND DISCUSSION

FIGET Grouping of Isolates

*B. japonicum* isolates were grouped by direct comparison of FIGETs produced by DNA restriction enzyme digestion with the rare-cutting enzymes AseI (Fig. 1A), DraI (Fig. 1B), SpeI (not shown), and XbaI (not shown). These enzymes cut the genome of *B. japonicum* in fewer than 30 fragments (see Section II). FIGETs produced by these four rare-cutting enzymes gave genomic fingerprints that were more easily interpreted and compared than DNA fingerprints produced by continuous-field (CF) electrophoresis of digests produced by using frequent-cutting enzymes (Bouzar and Moore, 1987; Stanley et al., 1985). This was due to the much smaller number of DNA fragments per genome that are produced by the former technique. The FIGETs could be easily analyzed and grouped. Serotype strains and field isolates belonging to the same serogroup can have either the same or a unique FIGET (Fig. 1, Table 1). At least two isolates per serogroup exhibited the same FIGET, regardless of which enzyme was used to digest the DNA (Table 1). For example, in serogroup 123, OH-13 was grouped with IN-79. In serogroup 127, isolates IA-67 and USDA 424 always produced the same FIGET, and in serogroup 129, USDA 129 and IN-78 were always classified together. With one
FIGURE 1A. FIGET classification of B. japonicum field isolates and strains after digestion of genomic DNA embedded in agarose plugs with AseI; lanes 1) uncut lambda DNA plus lambda cut with HindIII, 2) lambda concatemers, 3) OH-13, 4) IN-79, 5) IA-35, 6) OH-6, 7) USDA 434, 8) IN-78, 9) IA-67, 10) USDA 426, 11) USDA 424, 12) USDA 436, 13) IA-44, 14) USDA 438, 15) USDA 123. The numbers or letters that appear below the lanes indicate the FIGET class that the strain or isolate belongs to, as summarized in Table 1.
exception (USDA 438 digested with XbaI), serotype strains or isolates from one serogroup did not exhibit the same FIGET as an isolate from another serogroup.

Approximately one-half of the FIGETs produced by digestion with any given enzyme were unique; i.e., they were represented only once in the isolates analyzed (Table 1). Thus, FIGET analysis seems capable of demonstrating the large amount of genetic diversity present among isolates of the same serogroup, which is in agreement with studies using other methodologies (Hickey et al., 1987; Moawad et al., 1984; Roberts et al., 1980; Sadowsky et al., 1987). Furthermore FIGET analysis is capable of distinguishing isolates that are grouped in different serogroups, and rarely groups isolates from different serogroups in the same class. FIGET analyses therefore support the serologically based classification of *B. japonicum* isolates, although serological relatedness clearly was not a complete representation of the amount of genetic diversity present in soil populations.

Finally, a logical concern when using FIGETs as a tool to study genetic diversity is the stability of a strain's FIGET pattern over time or in a particular environment. Work in our laboratory has shown that FIGET patterns are stable over at least 100 generations under laboratory conditions and after symbiosis (see Section II). The question of FIGET stability
FIGURE 1B. FIGET classification of *B. japonicum* field isolates and strains after digestion of genomic DNA embedded in agarose plugs with *Dra*I; lanes 1) lambda concatemers, 2) IA-67, 3) USDA 426, 4) USDA 436, 5) USDA 434, 6) OH-6, 7) IN-78, 8) IA-44, 9) USDA 424, 10) IA-35, 11) OH-13, 12) USDA 438, 13) IN-79, 14) USDA 123. The numbers or letters that appear below the lanes indicate the FIGET class that the strain or isolate belongs to, as summarized in Table 1.
in the soil environment is much more difficult to address and will be examined in future investigations.

**Hybridization with nifHD**

Previous results (Sadowsky et al., 1987) indicated that hybridizations of a *nifHD* gene probe to *HindIII*-digested *B. japonicum* genomic DNA are capable of separating some of these isolates into restriction fragment length polymorphism (RFLP) classes that correspond to their nodulation phenotype on two *G. max* plant introductions (PIs) genotypes. The PI genotypes 371607 and 377578 restrict the nodulation of many serocluster 123 field isolates (Keyser and Cregan, 1987; Sadowsky et al., 1987). We tested whether *nifHD* hybridization to FIGE gels of digests produced by rare-cutting restriction enzymes would give RFLP classes that correlated with the nodulation phenotypes on these two PI genotypes. Table 2 shows the relationships between RFLP class and serology. None of the four rare-cutting restriction enzymes used were capable of producing RFLP classes that correlated with the nodulation phenotype of the isolates and serotype strains studied. In addition, no relationship of RFLP class and geographical origin of the isolate was observed with the *nifHD* probe.

When we compared the *nifHD* hybridization RFLP classes, an interesting relationship between serology and RFLP class was
TABLE 1. Relationship between FIGET and serological groupings of *B. japonicum* serocluster 123 isolates

<table>
<thead>
<tr>
<th>Serotype&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Isolate</th>
<th>AseI</th>
<th>DraI</th>
<th>SpeI</th>
<th>XbaI</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>USDA 123</td>
<td>1</td>
<td>1</td>
<td>U</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>OH-13</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>IN-79</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>OH-6</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>IA-35</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>USDA 438</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>2</td>
</tr>
<tr>
<td>127</td>
<td>USDA 127</td>
<td>U</td>
<td>2</td>
<td>U</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>IA-67</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>IA-44</td>
<td>2</td>
<td>2</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>USDA 424</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>129</td>
<td>USDA 129</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>IN-78</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>USDA 434</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>USDA 426</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>4</td>
</tr>
</tbody>
</table>

<sup>a</sup> Numbers 1-4 represent groupings that gave identical FIGET patterns when digested with the above restriction enzymes, whereas U represents unique FIGET patterns.

<sup>b</sup> Serological identity of strains and isolates was determined with fluorescent antibodies prepared as described by Schmidt et al. (1986).
observed with Asel, DraI, and SpeI digests (Table 2 and Fig. 2). Figure 2 shows a representative hybridization of the nifHD probe to Asel digests of the DNA-containing plugs. While serogroup 123 and 127 isolates shared a common 24-kb band, only serogroup 123 isolates had a 340-kb hybridizing band. Serogroup 129 isolates gave distinct RFLPs. In contrast, XbaI digests produced a single, conserved 23-kb band which hybridized with the nifHD probe in all of the isolates and serotype strains (Table 2). The RFLP classification of FIGE gels by hybridization with the nifHD probe therefore provided further genetic evidence that serogroups 123 and 127 isolates are more closely related to each other than to serogroup 129 isolates (Sadowsky et al., 1987). This is especially evident in Asel and DraI digests (Fig. 2 and Table 2). The RFLP classification presented here is in agreement with the serological relatedness of 123, 127, and 129 (Schmidt et al., 1986; Keyser and Cregan, 1987).

To further investigate the relationship between nifHD hybridization classes and serological groupings, we searched GENBANK release 59.0 for cloned B. japonicum nifHD sequences and used the SITE program to simulate digestion of these sequences with the enzymes used here. Cloned nifHD sequences are available in GENBANK from B. japonicum USDA 110 (Adams and Chelm, 1984; Kaluza and Hennecke, 1984). Analysis of these sequences revealed a DraI site in the promoter region of the
Figure 2. Representative nifHD hybridization to a FIGE gel of AseI-digested B. japonicum field isolates genomic DNA in plugs. The gel shown in Figure 1A was blotted and probed with nifHD, as described in Materials and Methods. Lane 1) lambda monomer plus lambda digested with HindIII, 2) lambda concatemers, 3) OH-13, 4) IN-79, 5) IA-35, 6) OH-6, 7) USDA 434, 8) IN-78, 9) IA-67, 10) USDA 426, 11) USDA 424, 12) USDA 436, 13) IA-44, 14) USDA 438, and 15) USDA 123.
nifD gene of strain USDA 110. An SspI site was located in the nifH gene (Adams and Chelm, 1984), but this enzyme gave too many small fragments per genome to be useful for FIGET studies (see Section II), even though it has an AT-rich target site. Results from hybridizations with the nifHD probe (Table 2) showed that for three of the serogroup 129 isolates two bands hybridized to DraI-digested DNA. In other isolates, from serogroups 123 and 127, this DraI site was not conserved, possibly suggesting a larger degree of evolutionary unrelatedness of serogroups 123 and 127 to serogroup 110. Interestingly, AseI-digested DNA from most isolates gave two bands when hybridized with nifHD sequences, although no AseI site exists in the USDA 110 nifHD sequences (Adams and Chelm, 1984; Kaluza and Hennecke, 1984). This AseI site is conserved in isolates from serogroups 123 and 127, but not present in serogroup 129 isolates (Table 2). An AseI site is found in the nifH sequence of Rhizobium sp. ORS571 (Norel and Emerich, 1987) and may imply evolutionary relatedness. Whatever the interpretation of the conservation of restriction sites in the nifHD gene sequences, they clearly reinforce the genetic relatedness of serogroups 123 and 127, and can differentiate serogroup 129 isolates from isolates of the two former serogroups.

This is the first application of FIGE to classify indigenous soil strains based on genomic fingerprints (FIGETs)
TABLE 2. Relationship between serological and FIGE-derived nifHD hybridization groupings of *B. japonicum* serocluster 123 isolates

<table>
<thead>
<tr>
<th>Serogroup&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Isolate</th>
<th>Asel</th>
<th>DraI</th>
<th>SpeI</th>
<th>XbaI</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>USDA 123</td>
<td>24, 340</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>OH-13</td>
<td>24, 340</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>IN-79</td>
<td>24, 340</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>OH-6</td>
<td>24, 340</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>IA-35</td>
<td>24, 340</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>USDA 438</td>
<td>24, 340</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>127</td>
<td>USDA 127</td>
<td>ND&lt;sup&gt;c&lt;/sup&gt;</td>
<td>ND</td>
<td>ND</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>IA-67</td>
<td>24, 150</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>IA-44</td>
<td>24, 320</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>USDA 424</td>
<td>24, 100</td>
<td>100</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>129</td>
<td>USDA 129</td>
<td>ND</td>
<td>ND</td>
<td>40, 70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>IN-78</td>
<td>100</td>
<td>50, 390</td>
<td>40, 70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>USDA 434</td>
<td>540</td>
<td>50, 300</td>
<td>40, 70</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>USDA 426</td>
<td>540</td>
<td>50, 300</td>
<td>40</td>
<td>23</td>
</tr>
</tbody>
</table>

<sup>a</sup> Serological identity of strains and isolates was determined using fluorescent antibodies (FAs) produced as described by Schmidt et al. (1986).

<sup>b</sup> Sizes for fragments that hybridized to the nifHD gene probe are in kilobase pairs (kb).

<sup>c</sup> ND = not determined.
produced by digestion with rare-cutting restriction enzymes. We believe the technique will have wide applications in ecological studies of soil microorganisms and may be useful for monitoring of engineered organisms in the soil.
ACKNOWLEDGMENTS

B. W. S. Sobral was supported in part by a Brazilian National Research Council (CNPq) Fellowship number 200289/85.3-AG. This research was also supported in part by a grant from the U. S. Department of Agriculture to M. J. Sadowsky. We thank Renee Kossak and Rhonda Honeycutt for their suggestions and critical reading of this manuscript and Harold Keyser for help in isolating some B. japonicum strains. Computer resources used to carry out our studies were provided by the BIONET National Computer Resource for Molecular Biology, which is funded by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health, Grant Number P41RR01685. We thank John Mayfield for the opportunity to interface with BIONET.
GENERAL SUMMARY

The studies conducted in Section I were initiated to determine the feasibility of resolving megaplasmids from Rhizobiaceae for physical mapping purposes. Although it now appears that resolution of large circular DNA molecules using various forms of pulsed-field gel electrophoresis (PFG) may be difficult (see General Introduction), some of the observations we made in Section I may apply to further the understanding of the mechanism of DNA migration during PFG.

The topology of circular, covalently closed (ccc) DNA molecules is quite different than that of linear DNAs, by definition. Such topological differences must cause differences in local rigidities and torsional strains during agarose gel electrophoresis. Also, the conformation that such molecules adopt within the gel matrix may be quite different from that of linear molecules. Once an electric field is applied, it is unknown how cccDNAs move, when compared to linear DNAs of approximately the same size. Observations in OFAGE systems (see General Introduction) show that superhelical density is an important influence on the electrophoretic mobility of cccDNAs during PFG. The effect of superhelical density may be a reflection of how the DNA molecule "sees" the gel, or how the DNA/gel system reaches steady-state dynamics during electrophoresis.
Pulse-time-independence has been observed for cccDNAs undergoing many forms of PFG (see Section I), yet my co-author and I (Section I) clearly show pulse-time dependence for cccDNAs undergoing field inversion gel electrophoresis (FIGE) and electrophoresis in contour-clamped homogeneous electric fields (CHEF). The mathematical complexity required of tube models to explain migration during FIGE make such models unattractive explanations for our observations and those of others. Mathematically simpler explanations are given by Deutsch's slinky model or Serwer's gel hysteresis hypothesis (see General Introduction). These models consider the agarose gel structure as an active component in the separation process. Considering our results, these models, and recent observations of individual DNA molecules undergoing electrophoresis in CF and PFG electrophoresis systems (see General Introduction), I believe that models such as Deutsch's and Serwer's are closer to reality than tube models or Southern's ratcheting model (see General Introduction). The latter model does not consider the gel structure carefully, although it predicts many experimental observations, just as do tube models.

The analysis of sequenced genes from Rhizobiaceae allowed my co-authors and I to make predictions about rare-cutting restriction enzymes for the genomes of these organisms. Such enzymes are essential to PFG techniques. Markov chain
analysis of the *Escherichia coli* genome (Phillips et al., 1987a, 1987b; Rogerson, 1989) has shown that there is a high degree of asymmetry in oligonucleotide abundance throughout the 4.7 Mb of DNA that make up the genome. Our analysis of *Rhizobiaceae* DNA sequences is in agreement with these observations, suggesting that the underrepresented DNA sequences in *Rhizobiaceae* may be of biological importance, as was found in some cases for *E. coli*. Specifically, in *E. coli*, palindromes were underrepresented, especially the Dam methylation site (Phillips et al., 1987a). Determination of the possible biological significance of rare DNA sequences in *Rhizobiaceae* (Section II, Table 2) awaits further investigation. It has been shown by many investigators that variation in di- and trinucleotide frequencies derives in part from codon usage, although codon usage alone is insufficient to explain the observed differences (Elton, 1975; Nussinov, 1980, 1981a, 1981b, 1984; Almagor, 1983; Blaisdell, 1985; Phillips et al., 1987a, 1987b). Markov chain analysis of frequencies of words in Western languages shows that first-order sequences using statistics of frequencies of letter doublets as a generator can be used to recognize the language from which the doublet frequencies have been taken (Abramson, 1964). The first-order correlations are strong enough to discriminate amongst several languages in the same family. Since DNA is also a language, it is clear that there are
strong constraints on doublet frequencies, which must reflect mechanisms rather than meanings.

Our estimates of genome size in *R. meliloti* and *B. japonicum* (see Section II) are important because they can be used as a guide to screen clone banks and for calculation of number of clones to cover the genome (see General Introduction). Interestingly, the size of the *B. japonicum* genome, as represented by strain USDA 424, was considerably smaller than has been suggested in the literature (see General Introduction and Section II). The reason for this discrepancy is unclear, however genome size estimates for other bacteria have been found to be erroneous once PFG techniques have been applied (see Section II). The differences in estimates may be a reflection of the inherent errors in using reassociation kinetics as an estimator (see General Introduction and Section II), or may stem from errors in sizing of PFG fragments, especially in the case of FIGE (see General Introduction), although the standard deviation of less than 9% for the whole genome of USDA 424 (see Section II) suggests that sizing errors are of acceptable magnitude. Even more interesting is our estimate of the size of the *R. meliloti* genome because I believe that our data, together with published data about the size of the megaplasmids in strain 1021 (a *Sm*² derivative of 2011), is suggestive of the existence of three chromosomes unique in *R. meliloti*. The three chromosome hypothesis
becomes especially attractive when one considers that it has been impossible to obtain megaplasmid-cured mutants of R. meliloti 1021 (or 2011 [see General Introduction]), and that genes for an essential biochemical pathway, the thiamine biosynthesis pathway, have been localized to a megaplasmid in R. meliloti (Finan et al., 1986). I believe that R. meliloti 1021 probably has three essential replicons in its genome, which should be called chromosomes, by definition, and that there is a large degree of genetic exchange between these replicons (Burkardt et al., 1987). Two recent papers by Suwanto and Kaplan (1989a, 1989b) have unambiguously demonstrated the existence of two unique chromosomes in Rhodobacter sphaeroides. The decision to call the large (914 kb) extrachromosomal element of R. sphaeroides a second chromosome was made based on the fact that it contained rRNA cistrons (rrnB and rrnC) as well as the gene for glyceraldehyde-3-phosphate dehydrogenase (gapB) (Suwanto and Kaplan, 1989b); these genes are essential for normal growth of the bacterium. If R. meliloti does in fact have three genetically distinct chromosomes, it will be only the second known case of a procaryotic organism with more than one unique chromosome. It may be a general feature of soil bacteria to possess complex overall genomic architecture, such as more than one chromosome per genome. This would certainly be of interest from an evolutionary standpoint.
Our finding that the restriction fragment patterns produced by rare-cutting enzymes is the same in bacteroids and free-living *R. japonicum* (see Section II) indicates that large-scale genetic rearrangements probably do not play a major role in bacteroid development or dedifferentiation. This view is indirectly supported by recent findings that show that DNA superhelicity (Gober and Kashket, 1989), protein phosphorylation (Karr and Emerich, 1989), and potassium ion levels (Gober and Kashket, 1987) are important to bacteroid functions. I feel that these mechanisms are better candidates as regulatory elements for bacteroid development than are alterations in DNA content or organization. Lack of major DNA rearrangements in large plasmids of *R. leguminosarum* bacteroids (Vierny and Iaccarino, 1989) also is in agreement with the stability of the (brady)rhizobial genome during symbiosis.

The techniques described within this dissertation should allow relatively simple and quick physical mapping of the genomes of *Rhizobiaceae* bacteria. A combination of partial digestion with restriction enzymes and PFG electrophoresis, using cloned genes as probes; cloning in yeast artificial chromosomes (known as YAC cloning: my calculations show that less than 100 YACs [with inserts on the order of 300 kb] are needed for any of these bacteria, in order to cover the entire genome at a 99% confidence level!); isolation of linking
probes that contain rare-cutting restriction enzyme target sites; and mutagenesis with engineered transposable elements that contain very rare target sites (such as the 12-mer specified by methylation with *M*.<sub>XbaI</sub> and subsequent cleavage with *DpnI* [M. McClelland, Department of Biochemistry, University of Chicago, personal communication]), should allow any of these bacteria to be physically mapped within a relatively short period of time.

In Section III, my co-authors and I described a practical application of PFG technology in *B. japonicum* field isolates from serocluster 123, which is known to be genetically diverse. Rather surprisingly, we found that a strong correlation existed between serological relatedness and RFLP or FIGET classification of field isolates, after digestion with rare-cutting restriction enzymes and electrophoresis in FIGE. This finding was surprising because it has been suggested that serological classification of field isolates may not represent genetic diversity well, since it is believed that major antigenic determinants probably represent the products of very few genes. Our findings (see Section III) show that a large amount of genetic diversity exists within the small sample of field isolates investigates, yet they also point to the genetic relatedness of serogroups 123 and 127, a fact that already had been implied at the phenotypic level by serological and SDS-PAGE studies (see Section III). FIGET
analysis may be used as a diagnostic tool in the future, because there is no need for Southern hybridizations to identify strains due to the small number of restriction fragments produced by the enzymes identified in Section II.


Casadeus, J., and J. Olivares. 1979. Rough and fine linkage


Craig, N. L. 1985. Site-specific inversion: enhancers,


Feulgen, R., and H. Rossenbeck. 1924. Mikrostopisch -


Hynes, M. F., R. Simon, P. Muller, K. Niehaus, M. Labes, and


Poustka, A., T. M. Pohl, D. P. Barlow, A.-M. Frischauf, and H.


Rogerson, A. C. 1989. The sequence asymmetry of the Escherichia coli chromosome appears to be independent of strand or function and may be evolutionarily conserved. Nucleic Acids Res. 17:5547-5563.


Sobral, B. W. S., and A. G. Atherly. 1989. A rapid and cost-effective method for preparing genomic DNA from Gram-


West, R. 1987. The electrophoretic mobility of DNA in


ACKNOWLEDGMENTS

I was supported by a fellowship from the Brazilian National Research Council (CNPq), reference number 20089/85-AG, and by a Genetics Department Research Assistantship. I sincerely thank these organizations for the opportunity and support to pursue my doctoral studies at Iowa State University; without their interest and willingness, it obviously would have been impossible.

I thank my family, Tania Lacerda Sobral, Paulo Lacerda Sobral, Juliana Lacerda Sobral, and Bruna Lacerda Sobral for enduring the many hours of absence from home that is required of molecular biologists; without their comprehension, I would not have reached this point in my career.

I thank my parents, Manoel Sobral Junior and Maria Cecilia Santos Sobral, for giving me the opportunity, support, and freedom of choice necessary to reach my own conclusions and pursue my own goals in my own time. If all parents could be so comprehensive and loving, certainly life would be much easier for all of us on what is left of planet Earth.

I thank my major professor, Alan Atherly, and his wife, Mary Atherly, for the warm reception I was given upon arrival in this country, and for their continued friendly support throughout my entire stay in Ames. It certainly makes life much easier for any graduate student to have such a good
friend in his or her major professor.

I thank my fellow graduate students for the long hours of discussion and discovery, over coffee or beer, in the middle of the night or in the mists of dawn. In particular, I thank Kim Engwall, for her help with teaching me the ropes when I arrived; Doug Walton and Rhonda Honeycutt for never-ending conversations about the squaring of $\sigma$ and other scientific motifs; John Luchansky for opening my mind to very important aspects about the attitude of a scientist; and Michele Antonelli, Geovanni di Guardo, and Laura Civardi for helping to keep a delicious touch of 'lovely latiny lifestyle' present in everyday midwestern life. Without such beautiful colleagues, it would never have been so rewarding and so interesting.

I thank Ms. Debora Reading, from the E.O. Building, for assorted advice about my future and for helping me keep the paperwork straight and my mind sane amidst the bureaucracy (which I never quite understood). For their unending patience and always prompt response to even the largest paperwork problems, as well as for putting up with my too-frequent memory losses, I warmly thank the Genetics Department secretaries, Ms. Mary Johnson, and Ms. Eileen Muff.

Finally, I thank Drs. Spartaco Astolfi Filho, Elza Maria Guimaraes, James Irvine, Wislon Marcelo da Silva, Carlos Floriano, Silvio Teixeira, John Imsande, Peter Pattee, David
Morris, Jette Foss, Brent Buckner, Benjamin Bowen, Renee Kosslak, and Michael McClelland for invaluable suggestions and conversations regarding my research, my career, and my life. It is the existence of such enthusiastic, fun-loving, and downright solid scientists that serves as a living testament for future scientists and that makes science such a fascinating and desirable career for all dreamers who pursue "Life As An Adventure" (for Harry Schurch in memorium).
Stated otherwise: Science Is The Drug, as John Imsande well knows....

The fire angel that turns slowly into the direction of the sunny heat;
asking what is the moment, how the flags will be flown...
Ask as you may, fire angel, yet carry the question one step further;
into the heart of the tired scientist, as he dwells over results unexplained, agonizing in the dawn of the drooling laboratory bench;
make the interrogation arrive with splendid exclamation, so that all may know the feeling of Eureka!

Carry the words and the feelings through the dirty streets that Man has created and set out to bake and freeze in
the twilight of City Life;
oh, fire angel, without horizons or skies, what is left for us
to fertilize;
how are we to create poetic beauty and warm life from the
cool, hard reality of concrete and rock?
But let the words ebb and flow throughout the corridors and
streets:
To grow in the light of Discovery,
to walk the crease that separates fact from conjecture,
to ask the previously unanswerable questions and focus all
efforts on their answers....
These are just a few of the delights offered to those who
journey through Science As An Adventure,
as a way to dream of Reality in such a deliciously rewarding
way;
a way to stretch the imagination and perception beyond the
boundaries that limit dull, gray, statistically
meaningless Everyday life!
A hop through the hole and you are out! Once there, like a
legendary traveller,
you are the Creator of Ideas that flow down through the rivers
of Time, for others to caress and tinker with;
a simple stretch of your imagination and you can reach your
bare feet into the landscapes of Tomorrow,
even if the skin on your feet is thin, you can look deep into
the eyes, the very retinas of tomorrow and,
while walking lightly, suddenly,
you are no longer afraid or alone:
You carry and can use the greatest tool of all:
Curiosity (and your life will never be the same again!).