AN ABSTRACT OF THE THESIS OF

Steven R. Strain for the degree of Doctor of Philosophy in Microbiology presented on October 12, 1993.

Title: Population Genetic Structure and Ecology of *Rhizobium leguminosarum*

Redacted for Privacy

Abstract approved: _____
Peter J. Bottomley

An understanding of the genetic structure of bacterial populations is prerequisite to asking ecologically and evolutionarily meaningful questions. To investigate population structure in *Rhizobium leguminosarum* bvs. *viciae* and *trifolii*, 682 strains from four populations (AV, AT, CV, CT) of the two biovars (V and T) on two field sites (A and C) in Oregon, and from populations of bv. *viciae* in Washington (WA) and the United Kingdom (UK) were analyzed for allelic variation at 13 enzyme-encoding loci by multilocus enzyme electrophoresis. With the exception of the WA population, genetic diversity was similar within each population (H=0.45-0.51). Only 4.2% of the genetic diversity in *R. leguminosarum* is attributable to geographical and biological separation, suggesting that migration is extensive and gene flow between the biovars is frequent. In this regard, the population structure of this species is more similar to *E. coli* than to that previously suggested for *Phaseolus*-nodulating rhizobia. Furthermore, the data indicate that low density populations of a bacterial species are not necessarily accompanied by low genetic diversity. The extent of disequilibrium (V_o/V_e) in

hierarchical combinations of populations decreased or remained unchanged as the geographical scale of the analyses increased providing further evidence of migration between the populations. Linkage disequilibrium analyses revealed significant multilocus structure in each individual population. Nine groups (A-I) of electrophoretic types (ETs) were defined in the populations based on a combination of cluster and disequilibrium analyses. Two groups (A and C) were found in all six populations, but their contributions to the number of isolates and ETs in the populations were not uniform. Although the other groups (B, D, E, F, G, H, I) were recovered from only certain of the populations, group B contributed significantly to the AV population, whereas group E made a substantial contribution to the AT, CT, and WA populations. Disequilibrium was significantly greater in the WA population than in any other individual population except AV and CT, and was least evident in the UK population. Although the WA and UK populations shared major groups of ETs (A and C), the difference in disequilibrium resulted from the amount of diversity within the groups. Since disequilibrium was differentially partitioned in individual rhizobial populations, the forces creating population structure can not be uniform. This is consistent with recent studies on soil-borne (*Bacillus*) and pathogenic (*Neisseria*) bacterial species.

Population Genetic Structure and Ecology of *Rhizobium leguminosarum*

by

Steven R. Strain

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Completed October 12, 1993

Commencement June 1994

APPROVED:

Redacted for Privacy

Professor of Microbiology, and Crop and Soil Science in charge of major

Redacted for Privacy

Chairman of department of Microbiology

Redacted for Privacy

Dean of Graduate School

Date thesis is presented October 12, 1993

Thesis presented and typed by Steven R. Strain

Acknowledgements

I wish to dedicate this thesis to Joy, without whose love and support none of this would have been possible. Through the good and the rough times, you were always there. There's a new metaphor to add to the story - "Joy and Steve, in Corvallis". To the rest of my family, you always said I could do it, although I occassionally had my doubts. Thanks for believing in me.

Uncountable thanks go to Peter, whose guidance in the course of my career as a graduate student has been invaluable. I consider myself lucky to have been associated with his lab. Remember, computers are our friends! I also extend my appreciation to the other members of my committee, Drs. Giovannoni, Loper, Myrold, and Mattson.

To Kam, Francis, Tom, Khrys and all the others who have come and gone from the lab, thanks for your friendship.

To Dr. Tom Whittam and Dr. Conrad Istock, I extend my sincerest thanks for taking the time to explain many population biology concepts to someone who can be as dense as a rock at times, and for answering the phone even when they knew it was me. I am grateful to Dr. Franz de Bruijn for generating the REP/ERIC PCR data presented in this thesis, and for giving me the opportunity to get some hands-on experience in his lab, and to Maria Schneider for leading me around by the hand while I was there. Their contribution to the thesis is greatly appreciated.

I would also like to acknowledge financial support through a Predoctoral Fellowship from the National Science Foundation, and from the Crop and Soil Science Department.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (CONTINUED)

Population Genetic Structure and Ecology of *Rhizobium leguminosarum*

Chapter I

Introduction to the Thesis

Relevant aspects of *Rhizobium* biology

Members of the genus *Rhizobium* possess the ability to enter into symbiotic associations with specific leguminous host plants where they fix atmospheric nitrogen to ammonia to the benefit of the host (Evans & Burris 1992). Although a substantial body of literature exists that deals with various aspects of *Rhizobium* biology, several issues are not resolved.

Early on it was found that when rhizobial inoculants were introduced into a soil harboring native populations of *Rhizobium*, the response of the host to the inoculation could be variable (Bottomley 1992). In this regard, two scenarios have been repeatedly observed: the introduced rhizobial inoculants were prevented from nodulating the host by the native populations; or an inoculation response was seen in the host during the season following inoculation, but this response rapidly dissipated over time (Bottomley 1992). The variability seen in these studies indicates that there is significant interaction between indigenous rhizobial populations and introduced inoculant strains. However, the biological factors influencing the outcome of competition between native and introduced strains remain unclear. Furthermore, considering the tremendous amount of genetic diversity that has been observed in bacterial populations in general (Selander, Caugant & Whittam 1987), and specifically in *Rhizobium* populations (Bottomley 1992), it becomes imperative that a thorough understanding of the genetic structure of indigenous rhizobial populations be attained, and the role that different subpopulations play in rhizobial ecology be established, if ecologically meaningful questions are to be posed.

Population structure and diversity in *Rhizobium leguminosarum*

One of the most important tools in the study of the genetic structure of bacterial populations has been multilocus enzyme electrophoresis (Selander *et al.* 1986). Because this technique is key to many of previous studies I will be discussing, as well as to the work in this thesis, a brief description of the method is in order.

Multilocus enzyme electrophoresis

Although multilocus enzyme electrophoresis (MLEE) has long been used extensively for the study of eucaryotic population genetics (Selander *et al.* 1986), only in the last 20 years has it become widely used as a tool for investigations into the population genetic structure and diversity of bacterial species (Selander *et al.* 1986, 1987). MLEE enjoys several advantages which make it useful for studies of this nature: i) it allows the rapid survey of the large number of bacterial strains required to make meaningful inferences about genetic diversity and population structure; ii) electrophoretic mobility variants can be directly related to alleles at distinct genetic loci; iii) genetic variability at loci for which no DNA sequence data is available may be assessed contingent only on the development of a specific assay; iv) protein allelic polymorphisms are most likely selectively neutral, or nearly so, making evolutionary convergence unlikely.

MLEE is based on the differential mobilities of soluble proteins having different net electrical charges in an electrical field. The differences in mobility are primarily a result of amino acid substitutions in which a residue is replaced by one having a different electrostatic

charge on the side chain. Although it has been estimated that electrophoresis can only detect between one-fourth and one-third of amino acid substitutions based on electrostatic charge differences (Nei 1987; Pasteur *et al.* 1988), substitutions which do not change the net charge of the protein may still be detectable by virtue of changes in the secondary and tertiary structure of the molecule (Pasteur *et al.* 1988). Following electrophoresis of the proteins, individual protein bands are visualized by the application of specific histochemical stains (Selander *et al.* 1986), which selectively reveal the positions of specific enzymatic activities.

MLEE has been used in a number of capacities in studies on bacterial species including: (i) investigations into the genetic structure and diversity of natural populations of bacteria, which have resulted in a greater understanding of the factors influencing the evolution and generation of diversity in a bacterial species; (ii) studies on the genetic structure and diversity within populations of pathogenic species; (iii) as a source of preliminary data for directing further investigations.

R. leguminosarum population biology

The first investigations into the genetic structure of populations of *R. leguminosarum* were carried out by Young and colleagues in the UK (Young 1985; Young, Demetriou & Apte 1987; Young & Wexler 1988; Harrison, Jones & Young 1989). Young (1985) obtained evidence that populations of the three biovars of *R. leguminosarum*, *viciae*, *trifolii*, and *phaseoli*, shared common chromosomal backgrounds. This finding was used as confirmatory evidence to support the taxonomic reorganization

of *Rhizobium* (Jordan 1984). Furthermore, these studies were the first to suggest that substantial levels of linkage disequilibrium, the non-random assortment of alleles at different genetic loci, exist in rhizobial populations. Although these studies contributed to a preliminary understanding of population genetic structure in *Rhizobium*, the analysis of a limited number (3) of genetic loci raises the question of whether the biovars might not share common alleles at the loci examined by chance, rather than as a result of close genetic relationship. Part of the work presented in this thesis is aimed at investigating the genetic diversity in populations of *R. leguminosarum* bvs. *viciae* and *trifolii* through the analysis of a greater number of loci.

Genetic diversity in populations of *E. coli* has generally been found to be three to five times greater than in eucaryotic species as defined by MLEE (Selander & Levin 1980; Selander, Caugant & Whittam 1987). However, the extent of genetic diversity in rhizobial populations is equivocal. For example, MLEE studies have shown that genetic diversity in *R. leguminosarum* is broadly comparable to that observed in *E. coli* populations (Young 1985; Young, Demetriou & Apte 1987; Piñero, Martínez & Selander 1988; Demezas *et al.* 1991). In contrast, methods such as plasmid profile and RFLP analyses have indicated a lower genetic diversity in soil rhizobial populations of *R. leguminosarum* bv. *viciae* recovered from pea (*Pisum sativum*; Mahler & Bezdicek 1978, 1980; Turco & Bezdicek 1987; Brockman & Bezdicek 1989; Hynes & O'Connell 1990; Laguerre, Mazurier & Amarger 1992). Evidence for the low diversity being due to the host plant selecting only certain genotypes of *R. leguminosarum* bv. *viciae* from the soil

population is equivocal. Although Hynes & O'Connell (1990) found that pea and faba bean (*Vicia faba*) nodulated with different members of the soil population, others have seen no evidence for a host effect by pea and lentil (*Lens culinaris*) grown in the same soil (Turco & Bezdicek 1987; Laguerre, Mazurier & Amarger 1992). The resolution of this controversy awaits a more detailed analysis of diversity within the soil rhizobial populations. I have addressed this issue in this thesis by analyzing diversity by MLEE in rhizobial populations recovered from a variety of hosts grown in soil at sites in Oregon, Washington, and the UK.

As statistical methods for the analysis of population structure such as bootstrapping and Monte Carlo simulations (Efron 1982), and procedures for determining both two-locus and multilocus linkage disequilibrium (Brown, Feldman & Nevo 1980; Brown & Feldman 1981; Hedrick & Thomson 1986) have become more widely used, studies have been published that delve more deeply into the underlying structure of populations of bacterial species. Istock and colleagues (Istock *et al.* 1992; Duncan, Ferguson, Kimura, Zhou & Istock, Evolution, In Press) investigated population structure in soil-borne populations of *Bacillus subtilis* and *B. licheniformis* recovered from a single microsite in the Sonoran desert. The authors concluded that recombination is frequent in these populations, at least on a local scale, although there was evidence of differences in the substructuring of populations of the two species. Souza *et al.* (1992) concluded from studies of linkage disequilibrium in local and diverse populations of *Phaseolus*-nodulating rhizobia from Mexico and the Western Hemisphere that migration in *Rhizobium* is

limited and recombination in local populations may be frequent. In contrast, investigations into the population structure of *E. coli* have indicated that migration is frequent and recombination is rare in this species. Whether the differences in population structure observed between these bacterial species reflects a fundamental dichotomy between soil-borne and commensal species is unknown. The work presented in this thesis will attempt to address this topic through analyses of population structure in local and geographically separated populations of *R. leguminosarum.*

Objectives

In consideration of the issues raised in this introduction, the following objectives were defined for the studies which make up this thesis.

(1) To determine the amount of genetic diversity in six populations of *Rhizobium leguminosarum* bvs. *viciae* and *trifolii*.

(2) To determine if there was any evidence of a host plant selection for certain genotypes from soil populations of *Rhizobium leguminosarum* bv. *viciae*.

(3) To examine the extent of multilocus structure in populations of *Rhizobium leguminosarum* bvs. *viciae* and *trifolii*, and to compare it with the findings of previous studies on soil-borne and human commensal species.

(4) To determine the genetic similarity between populations of *Rhizobium leguminosarum* bvs. *viciae* and *trifolii* recovered from geographically distinct locations.

Chapter II

Population Genetic Stucture and Ecology in Oregon Populations of *R. leguminosarum* Biovars *viciae* and *trifolii*

Introduction

Over the past 50 years, many studies have shown that soil populations of *Rhizobium* and *Bradyrhizobium* are composed of a large number of biotypically distinguishable strains (Bottomley 1992). Until recently, however, no attempts were made to analyze population diversity in a genotypically meaningful manner. In recent years, two strategies have been followed to examine population diversity. First, in an attempt to gain an overall picture of the genetic diversity within a rhizobial species, strains were chosen from culture collections with their origins on diverse hosts from geographically separated locations (Piñero, Martínez & Selander 1988; Eardly *et al.* 1990; Demezas *et al.* 1991). Secondly, to gain a picture of overall genetic diversity at a specific location, strains were recovered from nodules of a specific legume host planted at the site (Young 1985; Souza *et al.* 1992), or in soil recovered from the site (Harrison, Young & Jones 1987). To compare diversity at different locations, strains were recovered from the same host grown at various locations (Young, Demetriou & Apte 1987; Harrison, Jones & Young 1989). Unfortunately, the strategy of using a single legume species to trap soil rhizobia can lead to an underestimate of genetic diversity if the host legume selectively nodulates with only a subset of the total soil population, or if some strains are more competitive at nodulating than others within the soil population.

Although there are several reports in the literature showing that different clover (*Trifolium*) species can recover different subpopulations from the same soil population of *R. leguminosarum* bv. *trifolii* (Robinson 1969; Masterson & Sherwood 1974; Valdivia, Dughri & Bottomley

1988; Weaver *et al.* 1989; Leung, Yap, Dashti & Bottomley, Applied and Environmental Microbiology, In Press), the situation with *R. leguminosarum* bv. *viciae* is much less clear. Plasmid profile studies have indicated there to be limited diversity in soil populations of *R. leguminosarum* bv. *viciae* recovered from pea (Mahler & Bezdicek 1978, 1980; Turco & Bezdicek 1987; Brockman & Bezdicek 1989; Hynes & O'Connell 1990; Laguerre, Mazurier & Amarger 1992), whereas an MLEE study revealed significant diversity in a pea nodule population recovered from a UK site (Young, Demetriou & Apte 1987). Although there was no evidence for host preference being shown by pea and lentil grown in the same soil (Turco & Bezdicek 1987; Laguerre, Mazurier & Amarger 1992), Hynes & O'Connell (1990) obtained evidence that pea and faba bean (*Vicia faba*) nodulated with different members of the same population in a Canadian soil. Upon considering these issues I realized that all studies carried out to date had been conducted in arable soil with agriculturally important hosts of *R. leguminosarum* bv. *viciae* (peas, lentils, and faba beans). Furthermore, the effect of population density on the diversity in soil rhizobial populations has not been critically examined. Although Harrison, Jones & Young (1989) investigated diversity in bv. *trifolii* soil populations existing at a wide range of densities, it was not clear if extreme soil acidity or absence of the host played the dominant role in creating the low diversity measured in the soils with low density populations.

Preliminary studies identified an undisturbed open woodland site (site C) adjacent to pastoral land in which four of the most common vetch species of the Pacific Northwest were to be found (*V. hirsuta*, *V. villosa*, *V. sativa*, and *V. americana*). Vetch species (*Vicia* spp.) are

probably the primary hosts for bv. *viciae* in most natural ecosystems because they are the dominant legume species in the flora of grasslands and open woodlands in many parts of the world (Allen & Allen 1981; Kupicha 1981). The health and vigor of the legume population at the site, in combination with the lack of soil disturbance, had resulted in the development of a soil population of *R. leguminosarum* bv. *viciae* of substantial size ($>10^5$ per gram of soil). Furthermore, only a few *T. dubium* specimens were found on the site perimeter, and the bv. *trifolii* population was determined to be small ($<10^2$ per gram of soil). In contrast, on a neighboring site (site A) which had been sown to an improved grass-subclover pasture, the relative sizes of the two biovar populations were reversed ($>10^5$ per gram, bv. *trifolii*; $<10^2$ per gram, bv. *viciae*). These locations provided an opportunity to compare overall genetic diversities within soil populations of the two biovars existing at different densities, and to determine if pea would exert a different host preference for specific members of the populations than the native vetch species.

Materials and Methods

Study site descriptions

Site C was an open woodland area located midslope on the western side of Soap Creek Valley in Paul Dunn State Forest approximately five miles north of Corvallis, Oregon. The soil was a silty-clay loam of the Dixonville series (fine, mixed, mesic, Pachic Ultic Haploxeroll). The predominant herbaceous vegetation on site C were annual and perennial grasses with a legume component of *Vicia* spp. Site A, approximately one mile to the southeast of site C, was an improved subclover (*Trifolium subterraneum* L.)-orchard grass (*Dactylis glomerata* L.) pasture located on a north-facing toeslope in Soap Creek Valley. The soil was a silty clay loam of the Abiqua series (fine, mixed, mesic, Cumulic Ultic Haploxeroll).

The densities of the populations of bvs. *viciae* and *trifolii* on sites A and C were determined by the soil dilution/plant inoculation most probable number (MPN) procedure (Vincent 1970). *Vicia villosa* and *Trifolium hybridum* were used as specific hosts for bvs. *viciae* and *trifolii*, respectively. On site C, bvs. *viciae* and *trifolii* were present at respective densities of $>10^5$ and $<10^2$ g^{-1} oven dry soil. Population densities of the two biovars showed a reciprocal relationship on site A with bvs. *viciae* and *trifolii* present at $<10^2$ and $>10^5$ g^{-1} oven dry soil, respectively.

Individual populations of *R. leguminosarum*

Site C *R. leguminosarum* bv. *viciae* (CV population)

One hundred seventy-six isolates of *R. leguminosarum* bv. *viciae* were recovered from field-grown *V. hirsuta*, *V. villosa*, *V. sativa*, and *V. americana*. Additional isolates were recovered from *Pisum sativum* and *V. villosa* grown in soil transported from site C to the laboratory.

Site C *R. leguminosarum* bv. *trifolii* (CT population)

Seventy isolates of *R. leguminosarum* bv. *trifolii* were recovered from field-grown *T. dubium* and from subclover (*T. subterraneum*), crimson clover (*T. incarnatum*), white clover (*T. repens*), and alsike clover (*T. hybridum*) grown in soil transported from site C to the laboratory.

Site A *R. leguminosarum* bv. *viciae* (AV population)

One hundred twelve isolates of *R. leguminosarum* bv. *viciae* were recovered from root nodules of four host plants, *P. sativum*, *V. hirsuta*, *V. villosa*, and *V. sativa* grown under greenhouse conditions in site A soil.

Site A *R. leguminosarum* bv. *trifolii* (AT population)

One hundred ninety-eight isolates of *R. leguminosarum* bv. *trifolii* were recovered from various clover species grown under field or greenhouse conditions in site A soil.

Recovery of rhizobial isolates

Field-grown plants

Plants were brought from the field to the laboratory with their root systems intact and surrounded by soil. Excess soil was removed from the root system by gently breaking up the soil by hand under running water following which the root system was washed in a 0.1% (v/v) solution of Tween 20. Nodules were harvested as described below.

Pot-grown plants

Plastic pots were set up as follows for growing plants in the laboratory. A single layer of 2-cm diameter Bermuda rock was placed in the bottom of each pot to promote drainage. Approximately 1500 ml of dry, sterile, fine vermiculite was packed around a 20-cm length of 2-cm diameter PVC pipe with one end capped and holes drilled at approximately 3 cm intervals along its length. The vermiculite was moistened with 600 ml of sterile, distilled water. Surface soil (0-10 cm) was obtained from three locations on site C, combined, and broken up by hand. To avoid potential mineral deficiency problems during plant growth, the soil was amended with KH_2PO_4 (55 mg P kg^{-1}), $Na_2MoO_4 \cdot 2H_2O$ (1 mg Mo kg^{-1}), and K_2SO_4 (20 mg S kg^{-1}) in sufficient water (350 ml kg^{-1} oven dry soil) to raise the water potential of the soil to -30 kPa. Three replicate pots were prepared for each plant species, each containing a 1.5 kg portion of amended soil. The soil layer was gently packed to reduce future settling, and the weight of each pot was recorded.

Seeds of *V. villosa* subsp. *villosa* (Oregon State University Seed Laboratory) and *P. sativum* cv. 'Alaska' (Beal Seed Co., Portland, OR)

were surfaced sterilized separately. Seeds were placed in a sterile 250-ml Erlenmeyer flask and rinsed for 30 s in 100 ml of 95% (v/v) ethanol. The ethanol was decanted, and 100 ml of 25% (v/v) Chlorox bleach was added and swirled for 5-10 min. The seeds were then rinsed at least seven times (2 min each) by swirling with 100 ml portions of sterile distilled water. The seeds were allowed to imbibe in the final water rinse for 1-2 h at 4°C, spread evenly on 1% water agar plates, placed at 4°C overnight to allow further imbibition and to synchronize germination, and stored in the dark at room temperature until germination occured. For each plant species, approximately 20 seeds were sown 1-2 cm deep into each of the three replicate pots. The pots were covered with clear plastic wrap to prevent desiccation and placed in a growth room with a daylength of 14 h, temperature of 22 to 27°C, and lighting provided by 16 F48T12VHO daylight fluorescent lamps and six 25W incandescent bulbs. Upon germination, each pot was thinned to 8-10 seedlings and the plastic wrap was removed. Sterile distilled water was added periodically through the central watering tube to maintain the soil water content at approximately -30 kPa. Flowers were removed from pea plants as they formed to delay nodule senescence. After 6 wk of growth, the plants were removed from the pots, and adhering soil and vermiculite was removed from the root systems as described above.

Isolation of rhizobial strains from root nodules

For field-grown plants, 20 nodules were recovered from each plant species. For each plant species grown under greenhouse conditions, 20 nodules were recovered from each of the three replicate pots.

Nodules were surface sterilized by the same procedure used for surface sterilizing seeds except they were immersed for 10-15 min in a 0.1% (v/v) solution of Tween 20 prior to treatment with 95% ethanol. Rhizobial strains were recovered by crushing nodules onto yeast extract-mannitol (YEM) agar plates supplemented with 50 mg l^{-1} cycloheximide to prevent fungal contamination (Vincent 1970) followed by incubation at 27°C. After 2-3 d, cultures were streaked to purity. A single well-isolated colony was chosen from each plate and transferred to a YEM agar slant. Cultures were maintained on YEM agar slants stored at 4°C.

Multilocus enzyme electrophoresis (MLEE)

Preparation of cell extracts

Each isolate was grown at 28°C with shaking for 48-60 h in 50 ml of a defined glutamate-mannitol broth (pH 6.5) containing in grams per liter: mannitol, 10; L-(+)-glutamic acid, monosodium salt, monohydrate, 1; NaEDTA (27 µM), $FeCl_3$ (14 µM), $CaCl_2 \cdot 2H_2O$, 0.08; NaCl, 0.1; $MgSO_4 \cdot 7H_2O$, 0.2; KH_2PO_4, 0.5; 0.5 ml of the vitamin formulation of Chakrabarti, Lee & Gibson (1981); 10 ml of a trace elements solution (Evans 1974). Each culture was transferred to a 250-ml centrifuge bottle, pelleted at 16,300 x g for 15 min, resuspended in 1 ml of ice-cold Tris-EDTA buffer (5 mM Na_2EDTA, 10 mM Tris-HCl, pH 7.6), and transferred to a 1.7-ml capacity Eppendorf microcentrifuge tube. To each tube was added 30 µl of 100 mM dithiothreitol (Sigma, St. Louis, MO), 20 µl of freshly-prepared lysozyme solution (50 mg ml^{-1} in distilled water), and the suspensions were incubated for 1-2 h on ice with occasional mixing. Bacterial cells were broken by sonicating each sample twice for 10 s with a Branson Sonifier 200 equipped with a step-

down microtip and set at a power level of 6 on 50% duty cycle. Each sample was sonicated once and replaced on ice to cool before the second sonication cycle was begun. Cell debris was pelleted from each sonicated sample by centrifugation at 4°C (12,535 x g, 20 min), and the supernatant was aliquotted into two clean Eppendorf microcentrifuge tubes (1.7 ml capacity) and stored at -60°C.

Preparation of starch gel

Starch gels were prepared by suspending 17.1 g of hydrolyzed potato starch (Sigma, S-4501) in 150 ml of the appropriate gel buffer (Selander *et al.* 1986; Table A.1) in a 500-ml Erlenmeyer flask. The starch was dissolved by heating over a Bunsen burner with constant vigorous swirling to just past the boiling point. Air bubbles were removed by aspirating immediately until only large bubbles formed in the starch solution (about 20 s). Gels were cast by quickly pouring the molten starch into a preset plexiglass mold (13 x 10 x 0.6 cm) situated over a glass plate (13 x 10 cm). The gel was allowed to cool at room temperature for 45 min to 1 h, covered with clear plastic wrap, and stored at 4°C for at least 4 h prior to use.

Electrophoresis

A single-edge razor blade was used to cut a slit along the entire 13 cm length of the starch gel and 1.5 cm in from the edge. The 1.5 cm strip of the gel was gently pushed outward to create a slit about 2 mm in width. The gel was placed on the platform of a Bio-Phoresis horizontal electrophoresis unit (Bio-Rad, Richmond, CA) with the slit toward the cathode. The unit was pre-cooled to 2°C with circulating water from a

Lauda RMT-6 constant temperature circulating water bath (Brinkmann, Westbury, NY). Cell extracts were thawed partially, and a 15 µl aliquot of each extract was pipetted onto 5 x 10 mm strips of Whatman 3MM chromatography paper. The paper strips were placed in the slit of the gel and separated by 1 mm intervals. Paper strips loaded with extracts of rhizobial strains possessing specific electromorphic variants of enzymes were included as internal references. Strips loaded with an amaranth dye solution [1.2% (w/v) in 25% (v/v) ethanol] were placed at both extremes of the sample extracts to mark progression of the electrode buffer front, and the slit was closed to hold the paper strips in place. The gel was connected to the electrode buffer reservoirs with pieces of thick paper towel positioned with one end immersed in the electrode buffer and the other overlapping the gel by approximately 1 cm. Proteins in the cell extracts were allowed to migrate into the gel for 10 min under the voltage specified for the appropriate buffer system (Table A.1), after which the paper strips were removed, the gel covered with plastic wrap to prevent dessication, and electrophoresis continued at constant voltage for 1.5 to 2 h as appropriate for the buffer system in use. Following electrophoresis, the gel was removed from the gel mold, trimmed, and sliced horizontally into four 1.5 mm slices with a length of nylon fishing line. For enzymes with low activity, a 3 mm thick gel slice was routinely used. Each gel slice was placed in a plastic food container (15 x 10 x 2 cm) for assay of a particular enzyme. Electrode buffers were routinely used for three electrophoretic runs and then replaced with fresh buffer.

Enzyme assays

Specific enzyme assays were carried out essentially as described in Selander *et al.* (1986), with the exception that the quantities of many reagents were decreased by as much as 50% (Table A.2).

With the exception of β-galactosidase (BGA), which was assayed at 37°C, incubations were carried out at room temperature. The incubation times for the majority of the enzymes ranged from 10 to 30 min. Xanthine dehydrogenase (XDH) and nucleoside phosphorylase (NSP) required 2-3 h, and BGA was incubated overnight. Once the bands had developed sufficiently for scoring, the staining solutions were decanted, the gel slices rinsed in distilled water and fixed in a 1:4:5 (v/v) mixture of glacial acetic acid, methanol, and distilled water.

MLEE data analysis

Each isolate was characterized by its combination of allelic variants over the 13 enzymes assayed. Each distinct allelic variant profile was termed an electrophoretic type (ET; allelic profiles for the CV, AV, CT, and AT populations are presented in Tables A.4 through A.7, respectively). Genetic diversity (h) at an enzyme locus is calculated as $h = (1-\Sigma x_i^2)[n/(n-1)]$, where x_i is the frequency of the i^{th} allele at the locus, n is the number of ETs in the sample, and $n/(n-1)$ is a correction factor for bias in small samples (Nei 1978). Mean genetic diversity (H) is the arithmetic average of the h values over all enzyme loci examined. Coefficients of genetic differentiation [$G_{ST}=(H_T-H_S)/H_T$; Nei 1977] were used to apportion total diversity (H_T) over several populations into within-population (H_S) and between-population components. Relationships among ETs were revealed with a cluster

analysis program designed specifically for this purpose (ETCLUS; Dr. TS Whittam, Pennsylvania State University, State College, PA). The program was translated from FORTRAN to the C programming language with minor modifications by the thesis author prior to use. ETCLUS employs the UPGMA algorithm (Sneath & Sokal 1973) to cluster from a matrix of pairwise distances (calculated as the proportion of mismatched loci) between ETs. Relationships among ETs were displayed in the form of a dendrogram.

To determine the extent to which populations exhibit non-random combinations of alleles beween loci, the observed allelic mismatch frequency distribution was obtained by comparing each ET with every other ET once (for a total of $n(n\text{-}1)/2$ comparisons, where n is the number of ETs), and for each paired comparison, the number of dissimilar alleles (mismatches) was recorded. An equation for computing the variance of this distribution has been derived (Brown, Feldman & Nevo 1980; Brown & Feldman 1981):

$$V_o=\sum_j h_j-\sum_j h_j^2+2\sum_{j}\sum_{l>j}\sum_{i}\sum_{k}(2p_{ji}p_{lk}D_{ik,jl}+D_{ik,jl}^2).$$

If a population possesses random genetic structure (linkage equilibrium), all of the diallelic disequilibrium values ($D_{ik,jl}$) are equal to zero. In this case, the equation reduces to the difference between the first two terms which is the expected variance (V_e) of the distribution,

$$V_e=\sum_j h_j-\sum_j h_j^2,$$

and can be calculated directly from the single locus diversity values (h) of the observed data set. Thus, the inflation of the observed variance over the expected variance is related to the extent to which the population demonstrates non-random genetic structure. The ratio of V_o (the observed variance in the distribution of the number of mismatched loci over $n(n\text{-}1)/2$ paired ET comparisons) to V_e (the expected variance of such a distribution if all alleles were randomly associated) is a measure of linkage disequilibrium in a population. With increasing linkage disequilibrium the ratio V_o/V_e increases, and the inflation of V_o over V_e can be used to test the significance of linkage disequilibrium levels (Brown, Feldman & Nevo 1980). In practice, the above linkage disequilibrium statistics and allelic mismatch distributions were computed with the assistance of a program written by Dr. TS Whittam following translation by the thesis author from FORTRAN to C with minor modification.

Comparison of multilocus enzyme electrophoresis with DNA:DNA Hybridization

Jarvis, Dick & Greenwood (1980) reported on the relatedness of many strains of *R. leguminosarum* bv. *trifolii* to strain CC275e using DNA:DNA hybridization. Ten of the strains of *R. leguminosarum* bv. *trifolii* representing a range of hybridization values were used to determine if there was a correlation between the estimates of genetic relatedness based on MLEE with those obtained by DNA:DNA hybridization (Jarvis, Dick & Greenwood 1980; Table A.3). Strains CC275e, TA2, NZP549, NZP560, NZP550/2, NZP5117, and SU202 were obtained from Dr. BDW Jarvis (Massey University, Palmerston

North, New Zealand), and strains TA1, WU95, CC2480a, and UNZ29 were obtained from Dr. J Brockwell (CSIRO, Canberra City, ACT, Australia). Each of the 11 strains was analyzed by MLEE for allelic variation at 12 enzyme loci, and the number of allelic mismatches was recorded over the 12 loci between each isolate and reference strain CC275e (Jarvis, Dick & Greenwood 1980). Simple linear regression analysis was used to evaluate the correlation between estimates of genetic relatedness determined by the two methods.

Comparison of multilocus enzyme electrophoresis with REP (repetitive extragenic palindromic) and ERIC (enterobacterial repetitive intergeneric consensus) PCR

To further determine if MLEE data are good indicators of overall genetic relatedness among rhizobial strains, multiple isolates were chosen from three ETs and the profile of DNA fragments generated by PCR using REP (repetitive extragenic palindromic) and ERIC (enterobacterial repetitive intergeneric consensus) sequences as primers was determined as described by de Bruijn (1992). Oligonucleotide primers (REP1R-I: 3'-CGGICTACIGCIGCIIII-5'; REP2-I: 5'-ICGICTTATCIGGCCTAC-3'; ERIC1R: 3'-CACTTAGGGGTCCTC-GAATGTA-5'; ERIC2: 5'-AAGTAAGTGACTGGGGTGAGCG-3') were provided by Dr. FJ de Bruijn (Michigan State University, East Lansing, MI) and were synthesized with an Applied Biosystems 380B DNA synthesizer. A single colony of each isolate was suspended in a 25 µl volume containing 50 pmol of each of the appropriate pair of primers (REP or ERIC), 1.25 mM deoxynucleoside triphosphates, and 2 U of Ampli*Taq* DNA polymerase (Perkin-Elmer Cetus). Amplification

of the DNA was done using a DNA thermal cycler (Perkin-Elmer Cetus) set for the following series of cycles: 1 cycle at 95°C for 6 min; 30 cycles at 94°C for 1 min, at 40°C for 1 min, and at 65°C for 8 min; 1 cycle at 65°C for 16 min; and a final soak at 4°C. A 15 µl aliquot of the amplified DNA products was loaded onto a 1.5% agarose (low EEO; Boehringer Mannheim Corp., Indianapolis, IN) gel in TAE buffer (0.04 M Tris-acetate, 0.001 M EDTA), electrophoresed for 7 h at approximately 55 mA (70-80 V), stained with ethidium bromide, and photographed using Polaroid type 55 film. Similarities among the isolates were evaluated by comparing the banding patterns observed among the isolates combined over both REP- and ERIC-specific patterns, and were expressed as simple matching (S_m) coefficients. Banding patterns were coded by enumerating the total number of unique bands seen over all the isolates and, for each isolate, assigning each band position a 1 or a 0 to indicate the presence or absence of the band, respectively. For a pair of isolates, S_m was calculated as the sum of the number of bands present in both isolates and the number of bands absent from both isolates, divided by the total number of unique bands observed over all isolates examined. Similarity among the isolates was revealed in the form of a dendrogram constructed from the REP and ERIC PCR data using the NTSYS-pc analysis package (version 1.50; Exeter Software, Setauket, NY).

Results

Comparison of strain relatedness determined by multilocus enzyme electrophoresis, DNA:DNA hybridization, and REP (repetitive extragenic palindromic) and ERIC (enterobacterial repetitive intergeneric consensus) PCR

To determine the extent to which estimates of genetic relatedness revealed by MLEE agree with those obtained by DNA:DNA hybridization, 11 strains of *R. leguminosarum* bv. *trifolii* for which relative hybridization data were available (Jarvis, Dick & Greenwood 1980) were analyzed by MLEE. Simple linear regression of relative DNA:DNA hybridization values against the number of allelic mismatches between each strain and strain CC275e revealed a positive correlation (r^2=0.71, $P < 0.01$, Fig. 2.1).

The variability among isolates both within and between ETs was assessed by the PCR technique of de Bruijn (1992) using REP and ERIC sequence-specific primers. Figure 2.2 compares the relationships between 14 strains of *R. leguminosarum* determined by MLEE and REP and ERIC PCR. Whereas MLEE analysis, by definition, clustered strains of the same ET at a relative similarity of 1.0, REP and ERIC PCR revealed heterogeneity among strains within an ET, with no two isolates of the same ET being identical. However, the maximum divergence between isolates within any of the three groups was at 0.88 relative similarity, whereas the divergence between the two most closely related groups was at a relative similarity of 0.62. It is clear that strains of the same ET are more closely related to one another than to strains of other ETs. Moreover, this conclusion applies across biovars, because

strains 88FL1,7 and TA1 (bv. *trifolii*) cluster with strains of bv. *viciae*. Within a group, there was no trend for strains from the same site or of the same biovar being more closely related than to others in the group. For example, in the largest group containing three bv. *viciae* isolates from site C and three from site A, one of the site A strains (AVH14) was more closely related to the site C isolates than to others from site A.

Genetic diversity in the populations

For each of the four populations, genetic diversity values (*h*) at each of the 13 enzyme-encoding loci are presented (Table 2.1). Mean genetic diversity (H) is similar for each population, ranging from 0.45 (CT population) to 0.51 (AV population). Although genetic diversities at the majority of individual loci were generally similar across the four populations, diversity at three loci were greater within specific populations (PGM in CV, G6P in AV, and IDH in AT). Furthermore, if null alleles (those that show no detectable enzymatic activity) are excluded, only 25 of the 82 alleles are restricted to either site A or site C populations. In the majority of cases (23/25), these population-specific alleles were recovered in only one to three ETs. However, one AT population-specific HBD allele was identified in 12 ETs, and one PGM allele was identified exclusively in seven ETs recovered from the CV population.

Although genetic diversities at the majority of loci were similar, it is possible that the frequencies of the individual alleles at a given locus may be different among the populations. Such a result might be obtained if random genetic drift had operated on each population independently as a result of their geographic separation. To address this

possibility, coefficients of genetic differentiation (G_{ST}; Table 2.2) were used to evaluate the heterogeneity among allele frequencies at each locus between the populations. When each of the four populations was treated as a distinct subpopulation, the total G_{ST} was 0.042, indicating that only 4.2% of the total genetic diversity in the four populations was due to variation in the allele frequencies between the populations. There was no evidence for biological or geographical separation of the populations contributing more or less to the total variation in allele frequencies.

Linkage disequilibrium within the Oregon populations

Although G_{ST} coefficients indicate that genetic differentiation between the Oregon populations is low, they provide no information regarding the extent to which the populations are composed of genetically isolated lineages, i.e., demonstrate non-random combinations of alleles between loci. Values of V_o/V_e for each of the four Oregon populations of *R. leguminosarum* are presented in Table 2.3. V_o/V_e values ranged from 1.63 (CV population)to 2.03 (AV population),with the two bv. *trifolii* populations (CT and AT) expressing almost identical values (1.74 and 1.73, respectively). The extent of linkage disequilibrium in each of the populations was statistically significant at the 95% confidence level by the method of Brown, Feldman & Nevo (1980), indicating that the alleles at the different loci are associated non-randomly in multilocus combinations.

In an attempt to elucidate the subpopulations contributing to the chromosomal structure of the populations, linkage disequilibrium was systematically determined for combinations of subpopulations. The CV population was used as a base and the ETs were provisionally grouped

according to the results of a cluster analysis employing the unweighted pair-group method with averages (UPGMA) algorithm. Analyses of linkage disequilibrium were carried out both within and between the groups. As a result of the analyses, groups were defined as containing members that are in linkage equilibrium, but are in disequilibrium with the other groups. By combining the ETs in each individual population with the CV population, subjecting them to cluster analysis followed by disequilibrium analysis, a consensus population structure was developed which shows the overall relationships between major groups, and the occurrence of each group in the four populations (Fig. 2.3). Groups A and C were found in all four of the populations (Table 2.4), and made substantial contributions to the percentages of ETs in each of the populations. In contrast, groups D and F were not found in all populations and contributed minimally to the gene pool when they were present. Although groups B and E were also not found in all populations, their contributions to the populations were variable. For example, group B was well-represented in the AV population and group E in the CT and AT populations.

Linkage disequilibrium was examined between groups in each population to identify which were contributing significantly to the overall linkage disequilibrium in the individual populations (Table 2.5). In both bv. *viciae* populations (CV, AV) and one of the bv. *trifolii* populations (CT), disequilibrium was significant between each of the following comparisons (A+B+C+D, A+B+C, and A+B), demonstrating that disequilibrium is spread throughout the three populations in a similar fashion. In the AT population, however, exclusion of group E resulted in a nonsignificant level of disequilibrium indicating that this

group was the primary cause of linkage disequilibrium. The contribution of group E to the magnitude of linkage disequilibrium in the CV and CT populations could not be determined with any degree of confidence because the numbers of ETs within this group were small (Table 2.4).

Presence of identical ETs within more than one population

Although other workers have used the recovery of identical ETs from different locations as preliminary evidence for clonal structure (Selander & Levin 1980; Caugant, Levin & Selander 1984), the tremendous genetic diversity present in bacterial populations makes the recovery of identical ETs remote, even in moderately clonal populations. Nevertheless, I identified 15 ETs that were present in each of two populations. Eleven of these ET overlaps occurred between the AV and CV populations. Ten of the 11 were between ETs in group A, and the remaining one was between ETs in group C. Of the remaining four cases of ET overlap, one was observed in each of the following pairs of populations: AV/AT, AT/CV, AT/CT, and CV/CT. No ETs containing representatives of both the AV and CT populations were identified.

Distribution of the groups among the plant species

Three groups (A, B, C) accounted for 92 and 98% of the isolates recovered on sites A and C, respectively (Table 2.6). Chi-square analysis of the distribution of isolates from the various hosts (four and five hosts on sites A and C, respectively) over the three largest groups (A, B, C) revealed a highly significant degree of heterogeneity on both sites A ($\chi^2 = 21.94$, df = 6, $P < 0.005$) and C ($\chi^2 = 31.50$, df = 8, $P < 0.001$). However, if isolates from *V. villosa* are omitted from the

analysis, statistical significance is lost on both sites (site A: $\chi^2 = 8.90$, df = 4, $P > 0.05$; site C: $\chi^2 = 10.05$, df = 6, $P > 0.05$). These effects are caused primarily by group C isolates being overrepresented in nodules of *V. villosa* on site C relative to the other hosts. However, on site A the effect is caused by a greater representation by group A isolates in *V. villosa* nodules relative to the other hosts.

Despite only one ET overlap between the CV and CT populations, the majority of bv. *trifolii* strains recovered from site C soil clustered into the same groups as did the majority of bv. *viciae* strains from the same site (A, C; Table 2.6, Table 2.7). The only evidence for a host preference effect was shown by field-grown *T. dubium* which was nodulated primarily by isolates from group E. Although previous studies from this laboratory had shown a strong host preference by annual clover species for members of group E on site A (Leung, Yap, Dashti & Bottomley, Applied and Environmental Microbiology, In Press), neither of the two annual species used in this study (subclover and crimson clover) nodulated with any representatives from group E on site C. Because the annual species *T. dubium* was only found at one location on the site and the bv. *trifolii* population density is very low, it is possible that members of group E are not uniformly distributed on site C.

Table 2.1. Genetic diversity (*h*) at each of 13 enzyme-encoding loci in four *R. leguminosarum* populations in Oregon.

	Genetic diversity at a specific locus in population[a]			
	biovar *viciae*		biovar *trifolii*	
Locus	CV	AV	CT	AT
BGA	0.759 (6)	0.789 (6)	0.799 (5)	0.799 (6)
G6P	0.026 (2)	0.362 (4)	0.071 (2)	0.143 (3)
IDH	0.249 (3)	0.182 (4)	0.138 (2)	0.479 (5)
MDH	0.000 (1)	0.094 (3)	0.000 (1)	0.037 (2)
HBD	0.684 (6)	0.711 (5)	0.828 (6)	0.798 (10)
6PG	0.351 (5)	0.530 (6)	0.415 (3)	0.427 (3)
PEP	0.713 (6)	0.774 (8)	0.794 (6)	0.727 (5)
PGI	0.671 (6)	0.671 (5)	0.521 (4)	0.479 (5)
XDH	0.772 (6)	0.748 (6)	0.720 (5)	0.595 (6)
NSP	0.719 (6)	0.563 (5)	0.696 (4)	0.645 (4)
PGM	0.315 (3)	0.048 (2)	0.071 (2)	0.000 (1)
ADK	0.540 (4)	0.617 (4)	0.423 (2)	0.484 (2)
SOD	0.505 (2)	0.501 (3)	0.349 (2)	0.526 (3)
H[c]	0.485 (4.3)	0.507 (4.7)	0.448 (3.4)	0.472 (4.2)

a. Values in parentheses are number of alleles at locus.

b. C and A designate the sites from which the isolates were obtained.

c. H=mean genetic diversity in population.

Table 2.2. Coefficients of genetic differentiation (G_{ST}[a]) for Oregon populations of *R. leguminosarum* bvs. *viciae* and *trifolii*.

	Comparison between populations[b]				
Locus	All	Site C	Site A	bv. *viciae*	bv. *trifolii*
BGA	0.028	0.026	0.025	0.012	0.000
G6P	0.053	0.000	0.025	0.071	0.000
IDH	0.040	0.020	0.042	0.003	0.042
MDH	0.003	0.000	0.000	0.003	0.003
HBD	0.060	0.009	0.067	0.028	0.038
6PG	0.026	0.000	0.040	0.004	0.022
PEP	0.016	0.000	0.013	0.010	0.005
PGI	0.013	0.004	0.008	0.005	0.000
XDH	0.063	0.016	0.075	0.000	0.096
NSP	0.057	0.013	0.063	0.029	0.018
PGM	0.054	0.033	0.000	0.044	0.000
ADK	0.026	0.055	0.006	0.000	0.000
SOD	0.081	0.092	0.000	0.002	0.092
Total	0.042	0.021	0.035	0.012	0.029

a. G_{ST} is calculated as the ratio of between-population diversity to total diversity in all populations; see Materials and Methods.

b. All: four populations (CV, AV, CT, AT); Site C: CV vs. CT; Site A: AV vs. AT; bv. *viciae*: CV vs. AV; bv. *trifolii*: CT vs. AT.

Table 2.3. Linkage disequilibrium statistics for Oregon populations of *R. leguminosarum* bvs. *viciae* and *trifolii*.

		Mismatch Parameters				
Population	No. ETs	X[a]	V_o[b]	V_e[b]	V_o/V_e	Significance[c]
CV	78	6.30	3.86	2.38	1.63	$1.64<\mathbf{V_e}<3.11<\mathbf{V_o}$
CT	28	5.83	3.74	2.15	1.74	$1.03<\mathbf{V_e}<3.27<\mathbf{V_o}$
AV	42	6.59	5.05	2.49	2.03	$1.43<\mathbf{V_e}<3.54<\mathbf{V_o}$
AT	54	6.14	4.18	2.41	1.73	$1.52<\mathbf{V_e}<3.31<\mathbf{V_o}$

a. X is the mean number of mismatched loci (out of 13) over $n(n\text{-}1)/2$ paired ET comparisons.

b. Vo and Ve are the observed and expected variance, respectively, of the allelic mismatch frequency distribution for each population; see Materials and Methods for details.

c. 95% confidence limits calculated by method of Brown, Feldman & Nevo (1980).

Table 2.4. Distribution of isolates and ETs among major groups in four populations of *R. leguminosarum* bvs. *viciae* and *trifolii* from Oregon.

Group		% of isolates and ETs in the groups from the populations[a]			
		CV (176/78)	AV (112/42)	CT (70/28)	AT (198/54)
A	n[c]	54.5	48.2	25.7	28.3
	N	48.7	45.2	21.4	25.9
B[b]	n	5.7	26.8	0.0	1.0
	N	5.1	19.0	0.0	1.9
C	n	37.5	17.0	47.1	28.8
	N	43.6	14.3	67.9	55.6
D	n	0.6	7.1	1.4	0.0
	N	1.3	19.0	3.6	0.0
E	n	1.7	0.0	25.7	41.9
	N	1.3	0.0	7.1	16.7
F	n	0.0	0.9	0.0	0.0
	N	0.0	2.4	0.0	0.0

a. Values in parentheses give number of isolates/number of ETs in the population.

b. Includes groups B, G, and H; see Fig. 2.3.

c. n - percentage of isolates; N - percentage of ETs.

Table 2.5. Intergroup disequilibrium statistics in the CV, CT, AV, and AT populations of *R. leguminosarum*.

Comparison in population[a]	V_O/V_e	Significance[b]
CV:		
Total	1.63	$1.64<\mathbf{V_e=2.38}<3.11<\mathbf{V_O=3.86}$
A+B+C+D	1.62	$1.63<\mathbf{V_e=2.37}<3.11<\mathbf{V_O=3.84}$
A+B+C	1.64	$1.60<\mathbf{V_e=2.36}<3.12<\mathbf{V_O=3.88}$
A+B	1.49	$1.20<\mathbf{V_e=2.09}<2.97<\mathbf{V_O=3.10}$
AV:		
Total	2.03	$1.43<\mathbf{V_e=2.49}<3.54<\mathbf{V_O=5.05}$
A+B+C+D	1.66	$1.39<\mathbf{V_e=2.43}<3.48<\mathbf{V_O=4.04}$
A+B+C	1.71	$1.16<\mathbf{V_e=2.22}<3.27<\mathbf{V_O=3.79}$
A+B	1.94	$0.93<\mathbf{V_e=1.96}<2.98<\mathbf{V_O=3.80}$
CT:		
Total	1.74	$1.03<\mathbf{V_e=2.15}<3.27<\mathbf{V_O=3.74}$
A+D	1.52	$0.00<\mathbf{V_e=1.98}<\mathbf{V_O=3.01}<4.00$
C+E	1.47	$0.71<\mathbf{V_e=1.81}<\mathbf{V_O=2.66}<2.92$
AT:		
Total	1.73	$1.52<\mathbf{V_e=2.41}<3.31<\mathbf{V_O=4.18}$
A+B+C	1.40	$1.47<\mathbf{V_e=2.47}<\mathbf{V_O=3.47}<3.48$

a. Group B includes groups B, G, and H; see Fig. 2.3.

b. 95% confidence limits calculated by method of Brown, Feldman & Nevo (1980).

Table 2.6. Distribution of *R. leguminosarum* bv. *viciae* isolates from various hosts into major groups on sites A and C.

	No. of Isolates		% Total from Host Plant Species on Sites A or C									
			V. americana		*V. hirsuta*		*V. sativa*		*V. villosa*		*P. sativum*	
Group[a]	A	C	A	C	A	C	A	C	A	C	A	C
A	54	96		78.9	50.0	70.0	38.1	85.0	79.2	34.5	35.6	66.7
B	30	10		10.5	18.2	0.0	47.6	5.0	4.2	8.0	33.3	0.0
C	19	66		10.5	27.3	30.0	0.0	10.0	12.5	52.9	22.2	33.3
D	8	1		0.0	4.5	0.0	14.3	0.0	4.2	1.1	6.7	0.0
E		3		0.0		0.0		0.0		3.4		0.0
F	1				0.0		0.0		0.0		2.2	

a. Group B includes groups B, G, and H; see Fig. 2.3.

Table 2.7. Distribution of *R. leguminosarum* bv. *trifolii* isolates from various hosts into major groups on site C.

Group[a]	No. of Isolates	% Total from Host Plant Species				
		T. subterraneum	*T. dubium*	*T. repens*	*T. hybridum*	*T. incarnatum*
A	18	55.6	5.0	33.3	18.2	33.3
B	0	0.0	0.0	0.0	0.0	0.0
C	33	44.4	5.0	66.7	72.7	66.7
D	1	0.0	0.0	0.0	9.1	0.0
E	18	0.0	90.0	0.0	0.0	0.0
F	0	0.0	0.0	0.0	0.0	0.0

a. Group B includes groups B, G, and H; see Fig. 2.3.

Table 2.8. Number of observed and expected occurrences of ETs common to both the CV and AV populations of *R. leguminosarum* bv. *viciae*.

ET	Group	Representative isolate	Number of occurrences of ET	
			Observed	Expected[a]
2	A	APS2	3	0.0130
3	A	APS3	14	0.0096
7	A	APS12	4	0.0060
11	A	AVH3	19	0.0062
15	A	AVH8	7	0.0081
17	A	AVH15	4	0.0071
19	A	AVV3	19	0.0044
21	A	AVV11	5	0.0130
27	A	AVS4	4	0.0086
30	A	AVS15	7	0.0028
5	C	APS8	3	0.0170

a. Calculated by multiplying together the frequencies in the AVCV population of the alleles at each locus in the ET to get the expected frequency of isolation, and then multiplying by the number of ETs (n=109) in the population to get the expected number of occurrences.

Figure 2.1. Correlation between estimates of genetic relatedness of 10 *R. leguminosarum* bv. *trifolii* isolates to strain CC275e determined by MLEE and DNA:DNA hybridization. $r^2 = 0.71$.

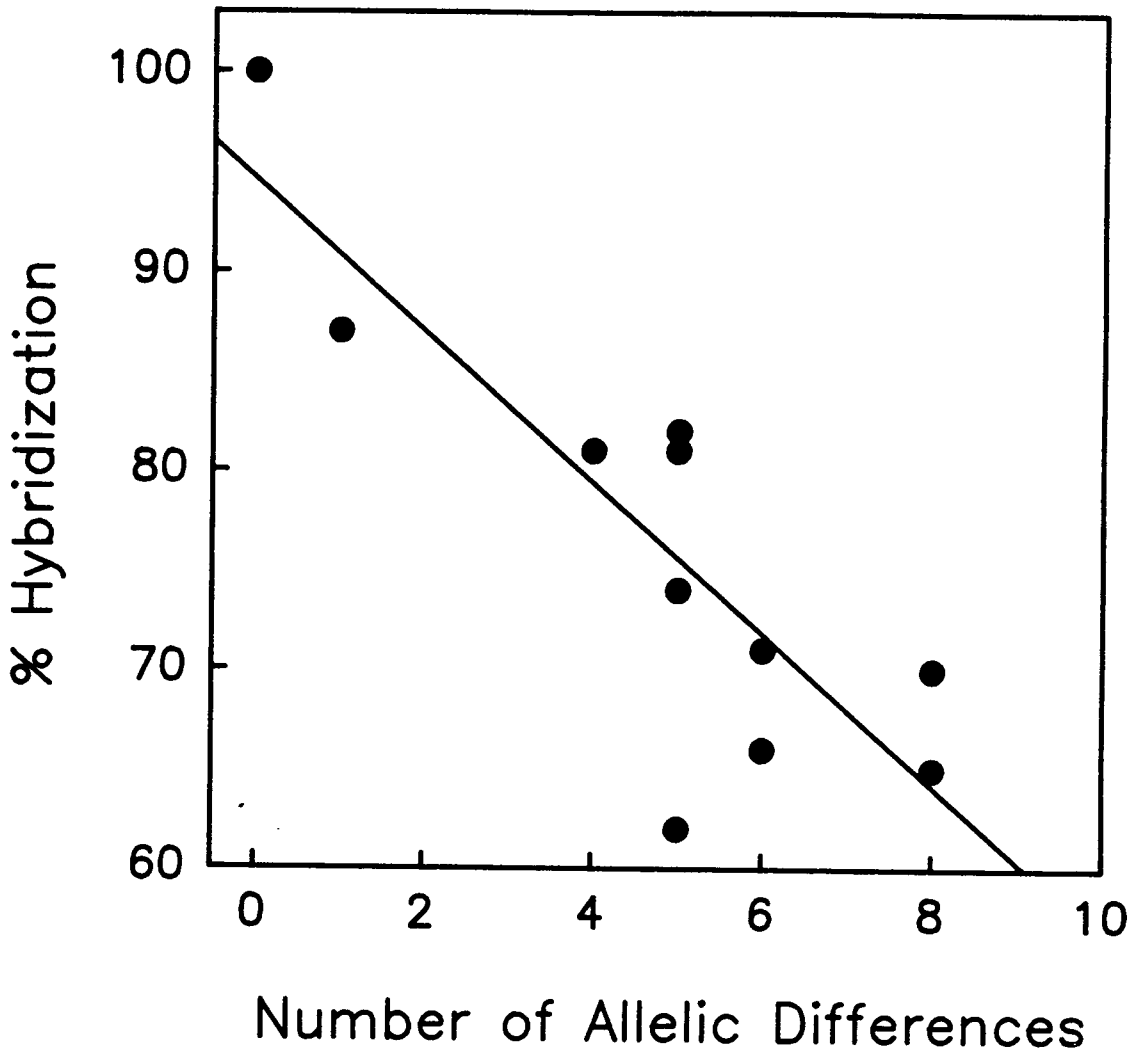
100
90
80
70
60
% Hybridization
0
2
4
6
8
10
Number of Allelic Differences

Figure 2.1

Figure 2.2. Comparison of estimates of genetic relationship between rhizobial strains determined by MLEE and PCR using REP and ERIC sequence-specific primers.

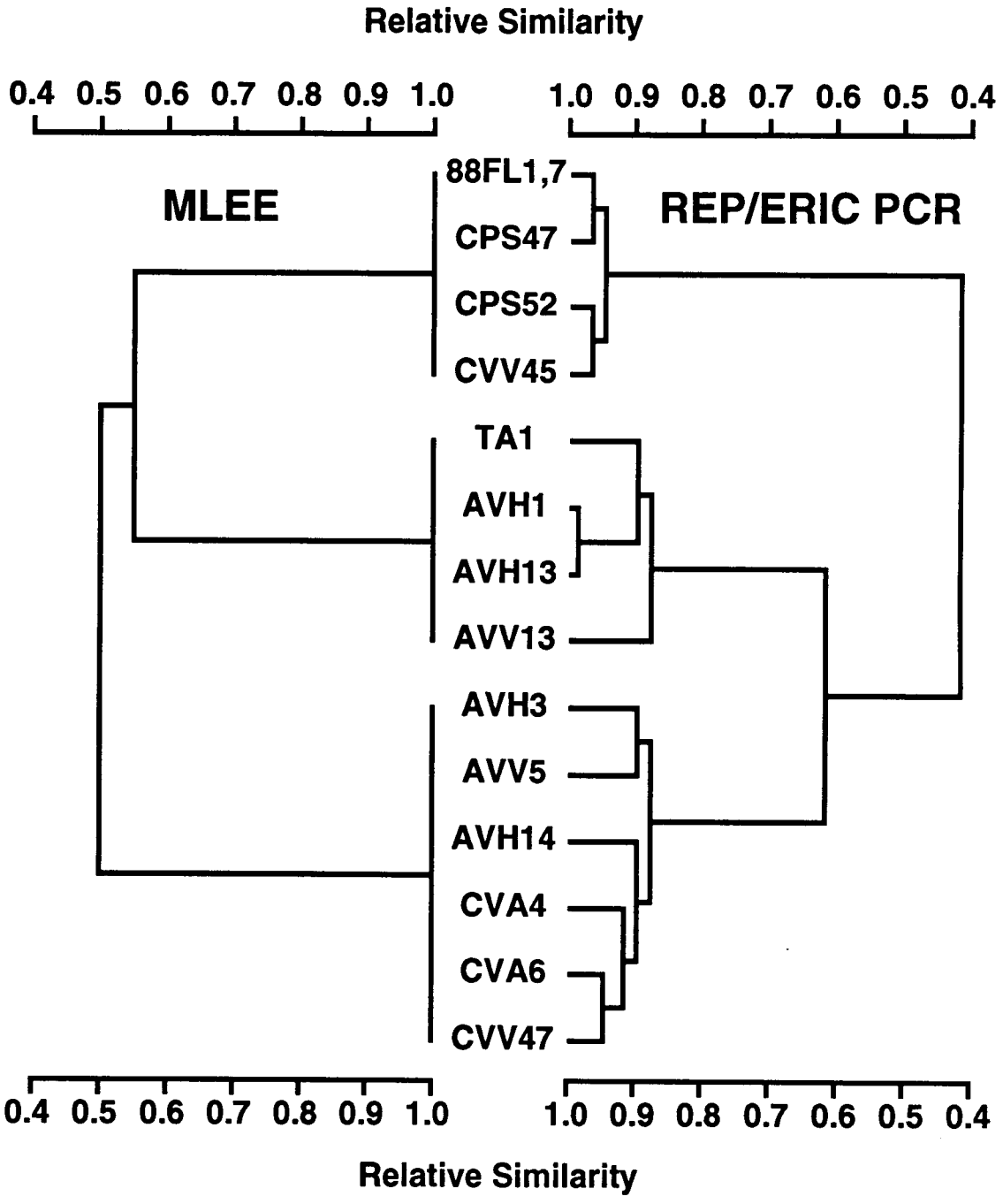
Relative Similarity
0.4 0.5 0.6 0.7 0.8 0.9 1.0
1.0 0.9 0.8 0.7 0.6 0.5 0.4
MLEE
REP/ERIC PCR
88FL1,7
CPS47
CPS52
CVV45
TA1
AVH1
AVH13
AVV13
AVH3
AVV5
AVH14
CVA4
CVA6
CVV47
0.4 0.5 0.6 0.7 0.8 0.9 1.0
1.0 0.9 0.8 0.7 0.6 0.5 0.4
Relative Similarity

Figure 2.2

Figure 2.3. Consensus dendrogram illustrating the overall relationships between major groups of ETs and their distribution across four populations of *R. leguminosarum* bvs. *viciae* and *trifolii* from Oregon.

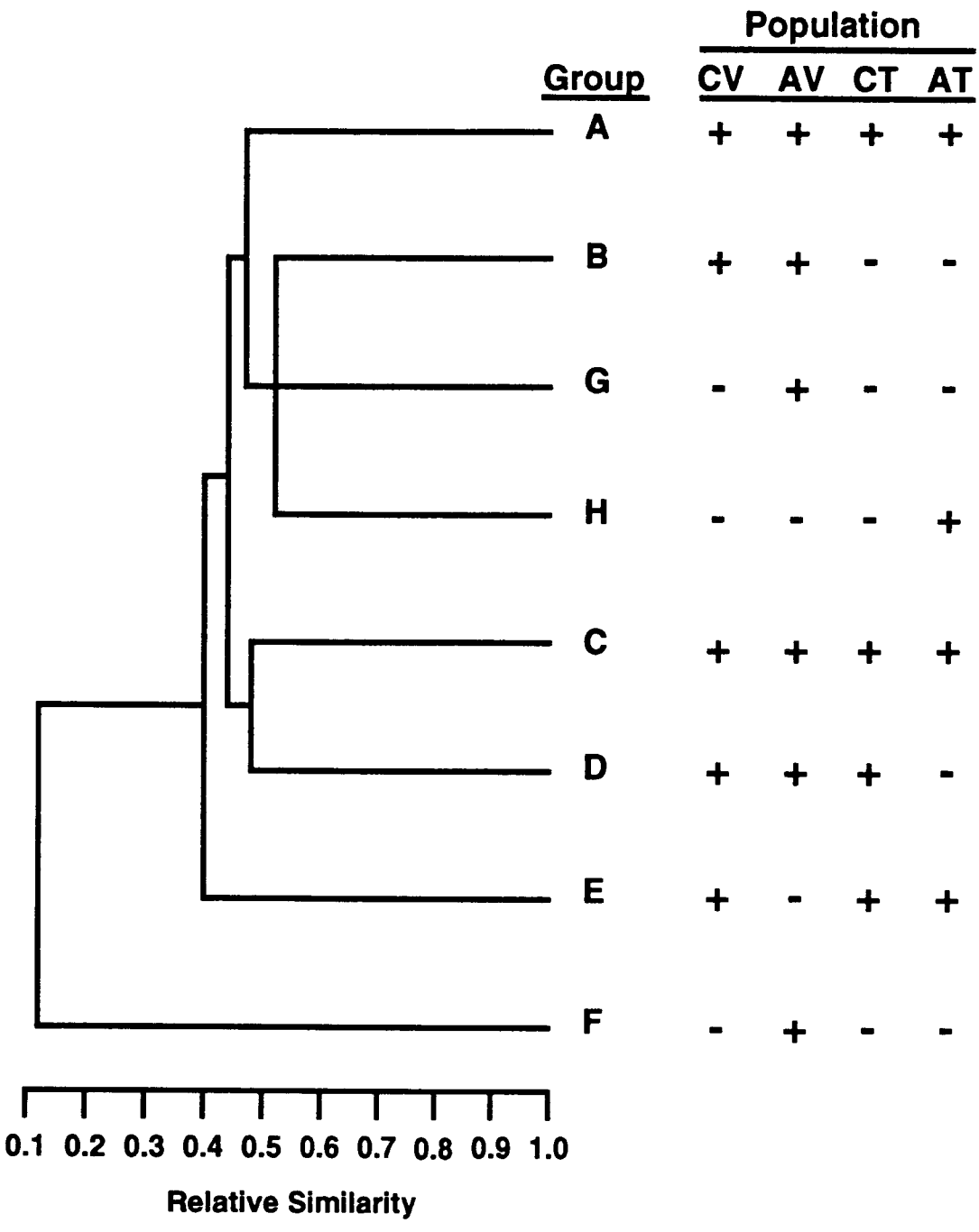
Population
Group
CV AV CT AT
A + + + +
B + + - -
G - + - -
H - - - +
C + + + +
D + + + -
E + - + +
F - + - -
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Relative Similarity

Figure 2.3

Discussion

The findings of this study have implications for *Rhizobium* biology and ecology, and, more generally, for the population biology of bacterial species. Consequently, the goal of this discussion will be to integrate in each of the following sections the significance of this work as it pertains to each of these fields. This study is the first to show a correlation between estimates of genetic relatedness established from data obtained by MLEE, REP and ERIC PCR, and DNA:DNA hybridization in *R. leguminosarum.* In this context, it is significant that REP and ERIC PCR grouped strains of the same ET into clusters that are well-defined relative to the relationships between groups. This result suggests that recombination in conjunction with mutation may lead to intra-ET variation, but is not sufficiently frequent between the ETs to break down multilocus structure. Nevertheless, I cannot exclude the possibility that intra-ET variation is due simply to chromosomal instability such as detected in *Phaseolus*-nodulating rhizobia (Flores *et al.* 1988).

Genetic diversity in natural populations of *R. leguminosarum* bvs. *viciae* and *trifolii*

It is noteworthy that the mean genetic diversity (H = 0.48) for the *R. leguminosarum* bvs. *viciae* and *trifolii* strains recovered from these populations is similar to values obtained from analyses of *R. leguminosarum* bv. *trifolii* strains from world-wide origins (H = 0.57; Demezas *et al.* 1991; Leung, Strain, de Bruijn & Bottomley, Applied and Environmental Microbiology, In Press). The H value is similar in

magnitude to mean genetic diversity values observed for several collections of *E. coli* strains recovered from local and international origins (Selander, Caugant & Whittam 1987). Obviously, there is as much genetic diversity within populations of *R. leguminosarum* bv. *viciae* and *trifolii* recovered from specific locations of between 2500 m^2 and 10,000 m^2 in area as contained within the species as a whole. A microscale assessment of genetic diversity as described recently by Istock *et al.* (1992) for *B. subtilis* recovered from 200 cm^3 of soil awaits further study.

The data presented in this thesis are the first to evaluate genetic diversity in bacterial populations existing in soil at different densities. Populations of commensal bacterial species, such as *E. coli*, are often sampled by suspending fecal swabs in liquid, followed by plating on solid media (Selander & Levin 1980; Caugant, Levin & Selander 1981, 1984). As a result, it is likely that the isolates recovered represent ETs that are present in the highest numbers. The potential exists to underestimate population diversity if other lineages are present in the environment at substantially lower densities. Legume species have the advantage of being able to recover rhizobial strains from soil that are present at low density even in the presence of higher density populations of the other biovar. However, as mentioned earlier, they can underestimate diversity if they select subpopulations preferentially. In this study, I saw no evidence for a specific chromosomal type dominating nodules of any host species regardless of the soil population density. However, because isolates of groups A or C dominated each of the host species (with the exception of *V. sativa* on site A), host preference may be operating at the group level. Because bv. *viciae*

strains of groups B, E, and F were recovered rarely (with the exception that group B isolates were prevalent on *V. sativa* on site A), I presume them to be poorly competitive or present in low numbers relative to groups A and C, or nonrandomly distributed. Interestingly, group E members are a nodule-dominant chromosomal type of bv. *viciae* in pea nodules at a site in Washington (see Chapter 3), in the AT population on subclover (Leung, Yap, Dashti & Bottomley, Applied and Environmental Microbiology, In Press), and in the CT population on *T. dubium* (Table 2.7). Further studies to evaluate the distribution and success of this group are required.

I have shown that the genetic diversity in both high and low density populations of bvs. *viciae* and *trifolii* are similar. I propose two different hypotheses to explain the low density seen in the CT and AV populations. First, because *Vicia* species are widespread throughout the hedgerows and non-pastoral areas in the Soap Creek valley, it is reasonable to assume that the AV population was probably of a density similar to the CV population before *Vicia* species went into a decline as a result of heavy grazing by bovines and hay harvests. The decline of the AV population must have been nonselective to account for the maintenance of its genetic diversity. In the case of the CT population, I propose that the low density is due to the inability of the host to establish itself in the plant community because of the predominance of grasses and small trees which shade the clover species and prevent their establishment. In this context, high diversity is probably due to continuous random immigration of bacteria onto the site without an opportunity being provided to proliferate because of the absence of the host. In the only other study where genetic diversity was investigated in

populations of *R. leguminosarum* bv. *trifolii* of differing densities, genotypic diversities were significantly lower at two sites where the rhizobial populations were present at only 60 cells g^{-1} of soil than at sites where densities were greater. In both of these cases, the soils were determined to be very acidic (pH 4.2-4.4). Because it is well known that neither *Trifolium* species nor *R. leguminosarum* bv. *trifolii* will thrive under these conditions (Bottomley 1992; Graham 1992), it is not clear what are the major factors contributing to this low diversity. More studies are required to gain an understanding of the roles of population collapse and immigration in determining the amount of genetic diversity in soil-borne bacterial species of low population density.

Relationships between site C isolates and rhizobial strains of other origins

When bv. *viciae* and bv. *trifolii* strains from diverse origins were analyzed by MLEE they were found to associate with groups A and C established in the Oregon populations. For example, seven of nine USDA strains (including the ATCC type strains) recovered from widely separated geographic locations (five states in the USA, and Yugoslavia) over a period spanning 44 years (1912 to 1955) clustered within group C. Furthermore, isolates representing (a) ATCC type strains of bv. *trifolii*; (b) one of the dominant ETs (MFF) recovered from pea nodules at a UK site (Young 1985); (c) one of the nodule-dominant serotypes on peas in the Palouse of eastern Washington (Brockman & Bezdicek 1989); and (d) two well-studied strains of bv. *trifolii* (TA1, ANU843) also fell into group C. In this connection, Eardly *et al.* (1992) found that the ATCC type strain of *R. leguminosarum* bv. *phaseoli* was similar

to the ET-MFF from the UK collection, and Laguerre, Mazurier & Amarger (1992) showed that the dominant chromosomal type from their studies of *R. leguminosarum* corresponded with the MFF chromosomal type from Young's UK collection. Several well-studied strains of bv. *trifolii* of Australian origin (WA67, WU95, NA30, CC275e, and UNZ29) clustered into group A. Two widely-studied *R. leguminosarum* strains, RBL5280 (bv. *viciae*) and USDA2124 (bv. *trifolii*) fell outside of any group defined in the Oregon rhizobial populations. Obviously, the chromosomal types of *R. leguminosarum* represented in groups A and C have become distributed globally throughout the biovars, yet, this in itself in no way guarantees them nodulating success at a specific site.

A comparison of multilocus population genetic structure in *Rhizobium* and other bacterial species

The recovery of identical ETs from different populations has been used as circumstantial evidence to support the hypothesis that populations of *E. coli* exhibit a clonal structure (Selander & Levin 1980; Caugant, Levin & Selander 1984; Selander, Caugant & Whittam 1987). The recovery of identical ETs is very unlikely given the enormous number of multilocus genotypes that could theoretically exist in bacterial populations if the alleles were randomly assorted. However, Maynard Smith *et al.* (1993) cautioned against using the simple fulfillment of this criterion as evidence for clonal structure without first proving that recovery of multiple instances of an ET is not due to random chance alone. In a study of population structure in clinical isolates of *N. gonorrhoeae*, Maynard Smith *et al.* (1993) noted that the most prevalent ET was recovered 35 times, and consisted of the most frequently

observed allele at each locus (i.e., represents the modal ET). However, calculation of the expected frequency of this ET in the population revealed that it was expected to occur 32.2 times. The frequency of recovery of other common ETs from the population also fit well with expected frequencies under the assumption of random assortment of alleles. In contrast, if the 11 ETs recovered from both the AV and CV populations are considered (Table 2.8), the lowest observed number of occurrences for any of the ETs in the AV/CV population is 3, whereas the highest expected number of occurrences is 1.74×10^{-2}. Even in the case of the two ETs recovered 19 times, the expected number of occurrences of these ETs in the population is only 6.2×10^{-3} and 4.4×10^{-3}, respectively. This discrepancy in the observed and expected recovery frequencies of these ETs suggests that the rhizobial populations I have studied possess a substantial amount of clonal structure, and that ET overlap between populations is a result of migration of the genotypes between sites. Alternatively, Maynard Smith *et al.* (1993) proposed a population structure intermediate between clonal and freely-recombining that they term an epidemic structure. In this case, the members of the population undergo relatively frequent recombination; however, the clonal expansion of certain ETs which have a selective advantage under certain conditions produces a superficially clonal structure in the population. The fact that all of the ET overlaps between the AV and CV populations are clustered into the widely- distributed groups A and C (Table 2.8) raises the possibility that *R. leguminosarum* populations exhibit an epidemic structure. It is noteworthy that the identification of ETs or groups of ETs that are potentially capable of clonal expansion under appropriate conditions may allow for a more directed approach in

selecting rhizobial strains for testing as plant inoculants, and for asking questions about the microbial ecology of soil-borne bacterial species.

Studies on *Phaseolus*-nodulating rhizobia and *B. subtilis* have revealed that local populations of these species may show varying degrees of multilocus structure. Souza *et al.* (1992) demonstrated in a local population of *Phaseolus*-nodulating rhizobia that 50% of the subpopulations recovered from the nodules of individual plants showed significant levels of linkage disequilibrium, whereas the other half were in linkage equilibrium. Similarly, recent studies into the population structure in *B. subtilis* indicated that the population was composed primarily of two groups which were undergoing intra-group recombination frequently and inter-group recombination to a lesser degree (Istock *et al.* 1992; Duncan, Ferguson, Kimura, Zhou & Istock, Evolution, In Press). Further study of a related species (*B. licheniformis*) recovered from the same microsite showed no evidence for inter-group recombination even though the population in this species was also composed of two groups (Duncan, Ferguson, Kimura, Zhou & Istock, Evolution, In Press). Maynard Smith *et al.* (1993) showed that even among bacterial species of the same genus (*Neisseria gonorrhoeae* and *N. meningitidis*), evidence for clonal structure and recombination may vary considerably. I now provide evidence that linkage disequilibrium is structured differently in the bv. *viciae* and bv. *trifolii* populations in two Oregon soils. Whether this difference is an artifact of incomplete sampling of the bv. *viciae* populations, or reflects some fundamental difference in the history of the bv. *viciae* and bv. *trifolii* populations on these sites awaits further study. Furthermore, although I have presented data that certain rhizobial strains from geographically

distant sites (Washington and the UK) cluster with isolates recovered in Oregon, it is unclear whether the population structure seen in *R. leguminosarum* on my field sites can be extended to rhizobial populations elsewhere. These issues will be explored more fully in the next chapter.

Chapter III

Hierarchical Analysis of Linkage Disequilibrium in Populations of *R. leguminosarum* Biovars *viciae* and *trifolii*

Introduction

Over the past several years, multilocus enzyme electrophoresis (MLEE) has been used to investigate the genetic structure and diversity of bacterial populations (Selander *et al.* 1986). The ability of MLEE to detect specific alleles at gene loci makes it possible to assess the non-random association of alleles at different loci, a phenomenon known as linkage disequilibrium (Lewontin & Kojima 1960). The extent of linkage disequilibrium within a population provides insight into the processes which influence its genetic structure. For example, in the absence of selective forces favoring certain multilocus allele combinations, a relatively high frequency of genetic recombination relative to growth rate will result in the decay of linkage disequilibrium and the population will move toward a state of linkage equilibrium (Whittam 1992). Conversely, in situations where either reproduction via binary fission occurs at a greater rate than recombination, or where recruitment of novel genotypes into a local population occurs at a measurable rate, or where periodic selection of genotypes occurs (Atwood, Schneider & Ryan 1951; Levin 1981), linkage disequilibrium will be maintained.

It is generally accepted that natural populations of *Escherichia coli* demonstrate a clonal population structure due to extensive global migration and to greater rates of asexual reproduction than of recombination (Selander & Levin 1980; Whittam, Ochman & Selander 1983b; Ochman & Selander 1984). Over the past 10 years, researchers analyzed MLEE-derived data sets from several bacterial species and concluded that a clonal population structure is the norm (Maynard Smith

et al. 1993). However, a combination of new evidence and reexamination of old data sets has challenged the clonal paradigm (Maynard Smith *et al.* 1993). Istock *et al.* (1992) observed no evidence of linkage disequilibrium within each of two MLEE-defined clusters of strains in a soil population of *Bacillus subtilis* recovered from a single microsite. Only modest levels of linkage disequilibrium were measured between the two groups. Subsequent studies of a *B. licheniformis* population recovered from the same microsite (Duncan, Ferguson, Kimura, Zhou & Istock, Evolution, In Press) revealed significant levels of linkage disequilibrium within one of two MLEE-defined groups and no evidence of disequilibrium within the other. The authors concluded that recombination was occurring frequently within groups but not between the groups of each *Bacillus* species. In contrast to the microscale studies of Istock and colleagues, Souza *et al.* (1992) characterized 276 strains of *Phaseolus*-nodulating rhizobia recovered from five populations at three sites in the state of Morelos, Mexico. Through an analysis which detects non-random associations of alleles at several enzyme loci, the authors concluded that only modest amounts of linkage disequilibrium existed within local soil populations and speculated that restricted migration between the populations contributed significantly to the development and maintenance of linkage disequilibrium among populations. Of course, this explanation is contrary to that given to explain linkage disequilibrium in local populations of *E. coli* into which extensive immigration is known to occur (Caugant, Levin & Selander 1981, 1984; Whittam, Ochman & Selander 1983b).

Although these findings might be used to support the hypothesis that the forces which drive the population biology of soil-borne and animal commensalistic bacterial species are fundamentally different, there are limitations to the study of Souza *et al.* (1992). *Phaseolus*-nodulating bacteria are known to be composed of several genetically distinct lineages (Martínez-Romero *et al.* 1991; Segovia, Young & Martínez-Romero 1993). Furthermore, it was not clear if the authors had gained a complete picture of the genetic diversity at specific locations since they did not consider the possibility of a particular host species only recovering a subset of each soil population.

I analyzed several populations of *R. leguminosarum* bvs. *viciae* and *trifolii* recovered from different locations. In total, 682 isolates representing 234 unique electrophoretic types (ETs) were examined. These data sets provided an opportunity to address the following issues relevant to the recent developments described above: i) to determine the extent of genetic diversity and linkage disequilibrium within six populations of *R. leguminosarum* bvs. *viciae* and *trifolii* recovered from four geographically distinct sites; ii) to determine if geographical separation contributes to the genetic distinctiveness of the populations.

Materials and Methods

Descriptions of Oregon sites

Site A was an improved subclover (*Trifolium subterraneum* L.)-orchard grass (*Dactylis glomerata* L.) pasture located on a north-facing toeslope in Soap Creek Valley approximately five miles north of Corvallis, Benton County, Oregon. The soil was a silty clay loam of the Abiqua series (fine, mixed, mesic, Cumulic Ultic Haploxeroll). Site C, approximately one mile to the northwest of site A, was an open woodland area located midslope on the western side of Soap Creek Valley in Paul Dunn State Forest. The soil was a silty-clay loam of the Dixonville series (fine, mixed, mesic, Pachic Ultic Haploxeroll). Several *Vicia* spp. were found on site C along with a few specimens of *Trifolium dubium* scattered on the perimeter of this site along a fence row.

Individual populations of *R. leguminosarum*

Site C *R. leguminosarum* bv. *viciae* (CV population)

One hundred seventy-six isolates of *R. leguminosarum* bv. *viciae* were recovered from field-grown *V. hirsuta*, *V. villosa*, *V. sativa*, and *V. americana*. Additional isolates were recovered from *Pisum sativum* and *V. villosa* grown in soil transported from site C to the laboratory.

Site C *R. leguminosarum* bv. *trifolii* (CT population)

Seventy isolates of *R. leguminosarum* bv. trifolii were recovered from field-grown *T. dubium* and from subclover (*T. subterraneum*), crimson clover (*T. incarnatum*), white clover (*T. repens*), and alsike clover (*T. hybridum*) grown in soil transported from site C to the laboratory.

Site A *R. leguminosarum* bv. *viciae* (AV population)

One hundred twelve isolates of *R. leguminosarum* bv. *viciae* were recovered from root nodules of four host plants, *P. sativum*, *V. hirsuta*, *V. villosa*, and *V. sativa* grown under greenhouse conditions in site A soil.

Site A *R. leguminosarum* bv. *trifolii* (AT population)

One hundred ninety-eight isolates of *R. leguminosarum* bv. *trifolii* were recovered from various clover species grown under field or greenhouse conditions in site A soil.

Washington State (WA population)

Seventy-six isolates of *R. leguminosarum* bv. *viciae* recovered from root nodules of peas (*Pisum sativum* cv. Latah) grown on two 48 m^2 field plots separated by 110 m in the Palouse region of eastern Washington (Brockman & Bezdicek 1989) were provided by Dr. DF Bezdicek (Dept. of Agronomy and Soils, Washington State University, Pullman, WA). Two of the strains, C4202 and M344, were used by Mahler & Bezdicek (1978) for the production of antisera WA-01 and WA-02, respectively. Of the remaining 74 isolates, equal numbers came from bottomland and south slope topographic positions, and 22%, 45%, 24%, and 9% of the strains serotyped as WA-01, WA-02, WA-03, and no serological reaction, respectively (Brockman & Bezdicek 1989).

Norfolk, England (UK population)

Fifty strains of *R. leguminosarum* bv. *viciae* recovered from root nodules of field-grown pea (*P. sativum*) were kindly provided by Dr. JPW Young (Dept. of Biology, University of York, York, UK). An attempt was made to obtain a cross-spectrum of chromosomal types from this population by obtaining representative isolates of the 10 most common electrophoretic types (ETs) from a larger collection (Young 1985; Young, Demetriou & Apte 1987): FMP, MFF, MFK, MFM, MSK, MSL, MSM, MSS, SSM, and SSQ.

Multilocus Enzyme Electrophoresis (MLEE)

MLEE was used to evaluate allelic variation in the rhizobial strains from Oregon, Washington, and the UK as described in chapter 2. The present analysis determined allelic variation at the same enzyme-encoding loci as described therein. Allelic profiles of the CV, AV, CT, AT, WA, and UK populations are presented in Tables A.4 through A.9, respectively. Genetic diversity (h) at each enzyme locus was calculated as $h = (1-\Sigma x_i^2)[n/(n-1)]$, where x_i is the frequency of the i^{th} allele at the locus, and n is the number of ETs in the sample. Mean genetic diversity (H) is the arithmetic average of the h values over all enzyme loci examined. Genetic distance between each pair of ETs was estimated as the proportion of loci at which the two ETs differed with respect to allelic variants. Relationships between the ETs were displayed in the form of a dendrogram constructed using a clustering program designed specifically for this purpose (Dr. TS Whittam, Pennsylvania State University, State College, PA). The program groups ETs from a matrix

of pairwise distances using the unweighted pair-group method with averages (UPGMA; Sneath & Sokal 1973).

Statistical analyses

If a bacterial species or a specific subpopulation of it is composed of individuals that recombine frequently relative to their rate of cell division, then the species should be composed of individuals possessing all combinations of alleles. In this case, the species or the subpopulation is said to possess random genetic structure. Conversely, if a species is composed of subpopulations which do not recombine frequently relative to the rate of cell division, or if recruitment of novel genotypes or extinction of local genotypes is occurring, then genetic structure develops and can be detected as non-random association of alleles at different loci.

Various statistical procedures have been developed to determine the extent of non-random association of alleles within populations of a species. Although all of the following tests have been used in studies of prokaryotes, each has its own strengths and weaknesses.

Multilocus linkage disequilibrium

A useful method for detecting multilocus genetic structure involves calculation of the allelic mismatch frequency distribution for the observed data, and comparing this distribution to one in which linkage disequilibrium between loci is equal to zero (i.e., there is random association of alleles among loci). An allelic mismatch frequency distribution is calculated by comparing each unique ET to every other unique ET once for a total of $n(n\text{-}1)/2$ pairwise comparisons, where n is

the number of ETs in the population. For each paired ET comparison, the number of mismatched loci are recorded.

The variance of frequency distributions of allelic mismatches has been formulated (Brown, Feldman & Nevo 1980; Brown & Feldman 1981) and is given by:

$$V_o = \sum_j h_j - \sum_j h_j^2 + 2\sum_{j}\sum_{l>j}\sum_{i}\sum_{k}(2p_{ji}p_{lk}D_{ik,jl} + D_{ik,jl}^2),$$

where h_j is the genetic diversity at the j^{th} locus, and p_{ji} and p_{lk} are the frequencies of the i^{th} allele at the j^{th} locus and the k^{th} allele at the l^{th} locus, respectively. $D_{ik,jl}$ is the coefficient of linkage disequilibrium for the specific alleles at the corresponding loci and is equal to the difference between the observed and expected ($= p_{ji}p_{lk}$) frequencies of the diallelic genotype in the population. Summation is carried out over the range between one and the number of loci for the j and l indices (constraining l to be greater than j ensures that each possible two-locus pair is considered only once), and between one and the number of alleles at the respective locus for i and k. If there is no linkage disequilibrium among loci, all of the $D_{ik,jl}$ values are zero, and the variance reduces to $V_o = \Sigma h_j - \Sigma h_j^2$. Thus, the inflation of the observed variance (V_o) over the expected variance (V_e) due to the third term, which includes the linkage disequilibrium coefficients ($D_{ik,jl}$), gives an indication of the extent of linkage disequilibrium among loci.

Three methods can be used to determine if the inflation of V_o over V_e is statistically significant. (i) Brown, Feldman & Nevo (1980) calculated 95% confidence limits around V_e, assuming the null hypothesis that no linkage disequilibrium exists. There is uncertainty in

this approach since it has not been shown conclusively that the sampling distribution of V_o is normal (Brown, Feldman & Nevo 1980). (ii) A method based on the χ^2 distribution (Sokal & Rohlf 1981) can also be employed to construct asymmetrical confidence limits around V_o. This procedure is also based on the assumption that the sampling distribution of V_o is approximately normal, and requires that similar caution be used. (iii) The nominal significance of the difference between V_o and V_e can be calculated using a Monte Carlo simulation (Souza *et al.* 1992), which makes no assumption about the sampling distribution of V_o. For this procedure, all of the alleles (equal in number to the number of ETs in the sample) at each locus are randomly shuffled three times to eliminate the effect of linkage disequilibrium, and the variance of the allelic mismatch distribution based on the randomized data is calculated. This process is repeated 10,000 times to simulate the sampling distribution of V_o. Randomization of the alleles is carried out without replacement, resulting in each allele being present in each randomized data set at the same frequency as in the observed data set. Statistical significance is calculated as the probability of obtaining a variance in the distribution of allelic mismatches equal to or greater than the observed V_o.

Bootstrap procedure

A bootstrap procedure was used to establish confidence limits around the ratio V_o/V_e for each population. By so doing, it is possible to determine which populations demonstrate levels of linkage disequilibrium significantly different from the other populations. In the only published example that I am aware of where this procedure was used on a bacterial species, the bootstrap procedure was carried out by

treating each collection of rhizobial isolates (not ETs) as a statistical population (Souza *et al.* 1992). For each population, 1000 new samples were randomly selected with replacement, each having the same number of isolates as the original population. By sampling with replacement, each new random sample may, by chance, contain multiple instances of some isolates, whereas others are not represented. V_o/V_e was calculated for each of the 1000 random samples. The 1000 values of V_o/V_e were sorted numerically, and the upper and lower 95% confidence limits were obtained by choosing an upper and lower value of V_o/V_e which together encompassed 95% of the 1000 values. There is controversy, however, among the experts in this field as to whether or not it is more appropriate to analyze the isolates or the ETs within a data set (Lenski 1993; TS Whittam, pers. comm.). Because of the importance of this analysis to our findings, I chose to analyze the data both in terms of isolates and ETs.

Results and Discussion

Analysis of rhizobial populations by multilocus enzyme electrophoresis

The six populations (CV, CT, AV, AT, WA, UK) were analyzed separately and also in hierarchical combinations designed to represent increasing geographical scales. Thus, the AV and CV populations were combined to produce the AVCV population, as were the AT and CT populations (ATCT population). AVCV was combined with the CT and AT populations to form the Oregon (OR) population; OR and WA were combined to produce the Pacific Northwest (PNW) population; and the PNW and UK populations joined to form the International (INT) population. Table 3.1 summarizes the results of MLEE analysis of each of the 11 populations. With the exception of the WA population, all of the populations show similar levels of genetic diversity ranging from 0.47 (AT and UK populations) to 0.52 (INT population). The WA population showed a somewhat lower mean genetic diversity than the other populations (H = 0.38). The mean number of isolates per electrophoretic type (ET) ranged from 1.6 (UK population) to 3.7 (AT population) indicating that the majority of the ETs in each population represent few isolates. The mean genetic diversity values for our six individual populations are similar to those observed from MLEE analyses of strains of *R. leguminosarum* bv. *trifolii* of diverse origins (0.57; Demezas *et al.* 1991; Leung, Strain, de Bruijn & Bottomley, Applied and Environmental Microbiology, In Press) and in studies of *E. coli* populations (0.34 - 0.54; Selander, Caugant & Whittam 1987). In brief, the genetic diversity within a single-site population of *R.*

leguminosarum can be as high as in the species as a whole. The values are lower than reported from analyses of *Phaseolus*-nodulating rhizobial strains [H=0.61 to 0.69; Piñero, Martínez & Selander 1988; excluding two ETs reclassified as *R. tropici* (Martínez-Romero *et al.* 1991) and any more divergent ETs; Segovia *et al.* 1991].

Analysis of allele frequencies from various rhizobial populations

The frequencies of the alleles at each of the 13 enzyme loci in the CV, CT, AV, AT, WA, and UK populations are summarized in Table 3.2. At many loci, one or two alleles were consistently identified at high frequency in all populations (i.e., G6P, MDH, 6PG, PGI, PGM, ADK, SOD). However, at each locus, there are alleles which are found exclusively in an individual population. For example, alleles were found at high frequency in the WA population that were seen rarely in other populations (i.e., BGA, IDH, HBD, PEP, XDH, NSP). The significance of novel alleles in the WA population will be discussed in more detail later in this paper.

Further evidence of the similarity in allele frequencies between the populations is provided by calculating G_{ST} coefficients (Nei 1977) at each locus (Table 3.2). G_{ST} is defined as the proportion of the total genetic diversity over all populations that is accounted for by variation between populations. Under the neutral mutation-random drift theory (Kimura 1968, 1983), the majority of molecular mutations are selectively neutral or nearly neutral and the fate of mutations in a population is determined largely by random genetic drift rather than by selective processes. Random genetic drift should, over time, result in different alleles being represented in genetically isolated populations. Moreover,

even if the same set of alleles were present at a specific locus in each population, geographic separation coupled with random drift should lead to different frequencies of the alleles in different populations. Interpopulation variation should increase with a resulting increase in G_{ST} values. Values of G_{ST} in this study ranged from 0.01 (MDH) to 0.23 (IDH) with a mean G_{ST} per locus of 0.10, indicating that only 10% of the variability in allele frequencies is attributable to differences between populations. This observation can be explained if there is frequent mixing of the population gene pools thereby preventing substantial divergence in allele frequencies.

Similar data were obtained from studies of *E. coli* and were interpreted to mean that migration of organisms between populations is frequent enough to prevent significant divergence in allele frequencies between populations (Caugant *et al.* 1983; Whittam, Ochman & Selander 1983a; Selander, Caugant & Whittam 1987). For example, in a comparison between three pathogenic *E. coli* UTI populations and a combined fecal population from Sweden and North America (Selander & Levin 1980), only 3.8% of the total variation in allele frequencies was due to variation between the four populations (Caugant *et al.* 1983). Furthermore, interpopulation variation accounted for only 1% of the total variation in a comparison between the fecal population from Sweden and a fecal population from North America (Selander & Levin 1980). A comparison of natural *E. coli* populations isolated from humans in Sweden, Iowa, and Tonga revealed that only 2% of the total genetic diversity was due to geographic variation (Whittam, Ochman & Selander 1983a).

Although the low average value of G_{ST} for the *R. leguminosarum* populations in this study indicates that migration is sufficiently frequent to prevent substantial differentiation of allele freqencies between the populations, migration does not appear to be as frequent as in *E. coli*. Although the transfer of *E. coli* strains around the world can be explained in terms of the global movement of humans, other routes of migration might be more significant for other bacterial species. The major migratory conduits of *R. leguminosarum*, for example, may be more dependent on slower processes such as movement on soil particles through the atmosphere and via water, the transfer on plant materials, or by migratory animals of either a feral or domestic nature. In this context it is interesting to speculate why Souza *et al.* (1992) favored limited migration to explain their linkage disequilibrium data on *Phaseolus*-nodulating rhizobia when so many migratory pathways can be proposed.

Linkage disequilibrium within rhizobial populations

Although the analysis of allele frequencies suggests that the six rhizobial populations are genetically related, it does not exclude the possibility that unique multilocus combinations of alleles exist within each population and the frequencies of the alleles themselves are similar due to chance alone. If this were the case and the six populations were genetically distinct (although not necessarily of different soil populations), analysis of the hierarchically combined populations should reveal a greater degree of linkage disequilibrium than seen in the individual populations (Maynard Smith *et al.* 1993). If the individual populations shared a substantial proportion of similar genotypes,

however, linkage disequilibrium can be expected to decrease or remain unchanged when populations are combined.

Table 3.3 summarizes the results of the analyses of linkage disequilibrium for each of the populations considered separately and in hierarchical combinations. V_o and V_e are the observed and expected variances in the frequency distribution of the number of allelic mismatches between pairs of ETs, respectively. The ratio of these values, V_o/V_e, provides a measure of multilocus disequilibrium in which V_e scales for the effects of single-locus genetic diversities (Brown, Feldman & Nevo 1980). The inflation of V_o over V_e, which is indicative of linkage disequilibrium, was significant for all of the individual and combined populations regardless of the method used to test for statistical significance. The only exception was the UK population which did not show significant disequilibrium by the method of Brown, Feldman & Nevo (1980) but did by the other two procedures. With the exception of the UK population, the highest values of V_o/V_e were observed in the individual populations, ranging from 1.63 (CV population) to 2.57 (WA population). Values of V_o/V_e decreased or remained essentially unchanged as geographical scale increased in the hierarchical populations. In contrast to our findings, Souza *et al.* (1992) reported an increase in disequilibrium as geographic scale was expanded, as well as an increase in mean genetic diversity (H). Mean genetic diversity is very similar in nearly all of the populations studied here (Table 3.1). It is possible that the increase in disequilibrium with increasing scale observed by Souza *et al.* (1992) is an artifact of the limited number of isolates analyzed at the larger scale (i.e., Western Hemisphere). Alternatively, the conclusions of Souza *et al.* (1992) may

be a result of the Western Hemisphere population containing several different phyletic lineages relative to the local population analyzed. Piñero, Martínez & Selander (1988) suggested that *Phaseolus*-nodulating rhizobia may represent a polyphyletic assemblage of lineages. This proposal has been strengthened by the reassignment of some strains analyzed in Piñero, Martínez & Selander (1988) to new species, *R. tropici* (Martínez-Romero *et al.* 1991) and *R. etli* (Segovia, Young & Martínez-Romero 1993).

The results of computer simulations employing the bootstrap procedure to establish 95% confidence intervals around V_o/V_e for each individual and hierarchically combined population are presented in Fig. 3.1. Two values for V_o/V_e were determined for each population by analyzing the data on the basis of either ETs or isolates (see section on bootstrap procedures in Materials and Methods). In general, larger values of V_o/V_e were obtained when the data were analyzed by isolates than by ETs. In addition, the ET-derived V_o/V_e values were often found at the lower limit of the randomly generated distribution of V_o/V_e values. This effect is due primarily to the majority of random values of V_o being larger than the value observed in the original data; V_e was more centrally located in the distribution of random V_e values (data not shown). Indeed, the values of V_o/V_e for two populations (CT and UK) were located outside of the lower 95% confidence limit (Fig. 3.1). Isolate-derived V_o/V_e values were less skewed, and all of the V_o/V_e values fell within the 95% limits. When analyzed by ETs, the V_o/V_e value for the WA population was significantly different from the values generated for the CV, AT, and UK populations. None of the other values were significantly different from one another. In the case of

isolate-derived V_o/V_e values, more significant differences were identified. Not only was the WA population V_o/V_e value significantly different from any other population, but the values for the CT and AT populations were also significantly greater than values for the UK and CV populations. With the exception of the combined ATCT population, V_o/V_e values (either ET- or isolate-derived) for the various combined populations were either not significantly different from, or were smaller than individual populations indicating that geographical separation is not an important contributor to linkage disequilibrium in these populations of *R. leguminosarum* from temperate zones. This conclusion contrasts markedly with that of Souza *et al.* (1992) but is consistent with conclusions drawn about *E. coli* populations (Whittam, Ochman & Selander 1983a).

Interestingly, the three populations (AT, WA, ATCT) which gave significantly greater values of V_o/V_e when analyzed on the basis of isolates had the highest isolate:ET ratios of any of the populations (Table 3.1; Fig. 3.1). A high isolate:ET ratio can indicate that the host plant is nodulating with only a subset of the genotypes present in the soil population, or that some genotypes are being excluded from nodules by more competitive members of the same population. In consideration of this issue, both host plant genotype and superior competitiveness by bacterial strains have been observed to influence the outcome of nodulation of *Phaseolus* beans (Hilali *et al.* 1989; Vargas & Graham 1989; Josephson *et al.* 1991; Wolff *et al.* 1991; George & Robert 1992). It would be interesting to know if the conclusions of Souza *et al.* (1992) would withstand analysis of their data on the basis of ETs, and if genetic

diversity within local populations was underestimated because of host choice.

Comparison of the WA and UK populations

Since the WA (V_o/V_e = 2.57) and UK (V_o/V_e = 1.44) populations represent the extremes of linkage disequilibrium that I have observed in individual populations, I sought to examine their population structures in greater detail. Figure 3.2 presents the results of cluster analyses carried out on each of the populations. A total of five groups were defined by cluster analysis. Groups A and C were common to both populations, whereas two groups (H, I) and one group (E) were unique to the UK and WA populations, respectively. Analyses of linkage disequilibrium were conducted both intra- and inter-group in the UK population (Table 3.4). Of the various comparisons available, only that between H and I showed significant disequilibrium. In the case of the WA population, no disequilibrium was found within any of the three groups but was detected between each of them (Table 3.4). Groups A and C were confirmed to be the same in both populations because there was no significant linkage disequilibrium even when the respective groups from WA and UK were combined (Table 3.5). In addition to inter-group population structure creating linkage disequilibrium, further analysis of the data revealed the importance of intra-group structure in creating linkage disequilibrium. At first glance, a discrepancy seems to exist in that there is disequilibrium between C and A in the WA population, but not in the UK population. However, a more detailed examination of lineage C in the WA population and A in the UK population showed a substantial difference in the genetic substructuring within those lineages.

Although the mean number of allelic mismatches between groups A and C in each population is identical (6.7), the respective values within groups A and C are 1.0 and 2.7 for the WA population, and 5.6 and 4.0 for the UK population. The lower intra- to inter-group ratios in the WA population indicate the role that intra-group diversity plays in creating linkage disequilibrium within the WA population. The effect on linkage disequilibrium is more easily seen by plotting the relative frequencies of the number of allelic mismatches between all possible pairwise comparisons of ETs in each population (Fig. 3.3). The WA population shows a distinctly bimodal distribution of allelic mismatches (Fig. 3.3; panel A). The two modes at 3 and 7 mismatches represent pairwise comparisons between ETs within and between groups, respectively. The separation between the two is a reflection of the low diversity within groups relative to between groups, and of the distinctness of the groups in the population. In contrast, the UK population (Fig. 3.3; panel B) displays only a single mode at 7 mismatches indicating that within-group diversity is closer to that between groups than in the WA population.

There is the possibility that the difference in linkage disequilibrium between the two populations is an artifact. It is possible that strains from group E, and from groups H and I are present in the UK and WA, respectively, and were simply not recovered. This possibility could be examined by sampling the soil populations using other trap hosts such as *Vicia* or *Lathyrus* spp. Group E in the WA population may represent a species distinct from *R. leguminosarum*. Segovia, Young & Martínez-Romero (1993) proposed that *Phaseolus*-nodulating rhizobia from the Americas can belong to three species, *R. leguminosarum* bv. *phaseoli*, *R. tropici*, and *R. etli*. Furthermore, the

authors identified a rhizobial strain nodulating pea that possessed a chromosomal background similar to *R. etli*. Since group E is not found in the UK population, it is possible that group E in the WA population represents *R. etli* or some other lineage not found in the UK. In support of this possibility, alleles were identified at four loci (IDH, HBD, XDH, NSP) in group E which were not observed in any other group in the WA or UK populations. Nevertheless, I should temper this speculation with the fact that in the WA population group E diverges from group A at a moderate distance of only 0.48 (Fig. 3.2). Analysis of 16s rRNA sequences of representative isolates from group E and authentic *R. etli* strains will be required to conclusively settle this issue.

Although recent publications have shown that population genetic structure of soil-borne bacterial species might be different from that of *E. coli*, our analyses of a different species do not support such a theory. Nevertheless, as the study of the population biology of procaryotes is unquestionably in its infancy, far more research effort will be needed to clarify the situation.

Table 3.1. Summary of MLEE analysis of individual and hierarchical combinations of populations of *R. leguminosarum* bvs. *viciae* and *trifolii*.

Population[a]	No. of Isolates	No. of ETs	Mean genetic diversity (variance)
CV	176	78	0.49 (0.076)
CT	70	28	0.45 (0.085
AV	112	42	0.51 (0.072)
AT	198	54	0.47 (0.074)
WA	76	22	0.38 (0.082)
UK	50	32	0.47 (0.076)
AVCV	288	109	0.51 (0.071)
ATCT	268	81	0.48 (0.078)
OR	556	187	0.51 (0.075)
PNW	632	208	0.51 (0.076)
INT	682	234	0.52 (0.075)

a. See Materials and Methods for origin of each population.

Table 3.2. Allele frequencies at each of 13 enzyme-encoding loci in six populations of *R. leguminosarum* bvs. *viciae* and *trifolii*.

		Population						
Locus	Allele	AV	AT	CV	CT	WA	UK	G_{ST}[a]
BGA	1		0.02			0.50		0.07
	2	0.17	0.32	0.39	0.25	0.05	0.28	
	3	0.36	0.17	0.26	0.14	0.27	0.22	
	4	0.12		0.09			0.03	
	6	0.21	0.19	0.14	0.11	0.18	0.22	
	7	0.12	0.11	0.09	0.18		0.25	
	8	0.02						
	9		0.20	0.04	0.32			
G6P	1	0.79	0.93	0.99	0.96	1.00	0.63	0.12
	2	0.17	0.04				0.25	
	10	0.02	0.04		0.04		0.13	
	11			0.01				
	0	0.02						
IDH	1		0.17	0.01	0.07	0.36		0.23
	2	0.02	0.06					
	3		0.02					
	4	0.91	0.70	0.86	0.93	0.23	0.94	
	5	0.05	0.06	0.13		0.41	0.06	
	10	0.02						
MDH	1	0.95	0.98	1.00	1.00	1.00	1.00	0.01
	2		0.02					
	10	0.02						
	0	0.02						
HBD	1			0.01		0.05		0.10
	2	0.02	0.02	0.21	0.25	0.55	0.34	
	3		0.04	0.01	0.14	0.27		
	4	0.36	0.35	0.45	0.25	0.05	0.34	
	6	0.29		0.04				
	7	0.31	0.04	0.28	0.21	0.09	0.16	
	8		0.22					
	11		0.19		0.07			

Table 3.2 (continued)

	15		0.04					
	16		0.06				0.03	
	17		0.02					
	21						0.09	
	0	0.02	0.04		0.07		0.03	
6PG	2	0.67	0.72	0.80	0.75	0.96	0.84	0.04
	4		0.24		0.07			
	5	0.17		0.14	0.18	0.05	0.13	
	6	0.02		0.04				
	9	0.10		0.01				
	12	0.02						
	15	0.02	0.04	0.01			0.03	
PEP	1	0.14		0.03			0.16	0.10
	2	0.41	0.41	0.45	0.39	0.09	0.47	
	3			0.04	0.07			
	4	0.05	0.24	0.10	0.14	0.23	0.22	
	5	0.02		0.13	0.18	0.68	0.03	
	6	0.19	0.11	0.26	0.11		0.06	
	7	0.02						
	8	0.02						
	9	0.14	0.22		0.11		0.06	
	10		0.02					
PGI	1	0.07		0.05				0.09
	2	0.52	0.70	0.47	0.64	1.00	0.59	
	3	0.12	0.06	0.01			0.31	
	4	0.21	0.17	0.31	0.29		0.06	
	5	0.07	0.06	0.04			0.03	
	6		0.02	0.12	0.04			
	11				0.04			
XDH	1	0.02	0.02	0.08	0.04	0.09	0.25	0.08
	2	0.24	0.61	0.23	0.21	0.18	0.19	
	3	0.33	0.11	0.35	0.18	0.32	0.38	
	4	0.31	0.11	0.22	0.46	0.05	0.16	
	5	0.07		0.10	0.11			
	6	0.02	0.13	0.03				

Table 3.2 (continued)

	8		0.02				0.03	
	9					0.36		
NSP	1	0.02						0.16
	2	0.62	0.32	0.37	0.21	0.09	0.38	
	3	0.10		0.09				
	4	0.24	0.50	0.35	0.39		0.56	
	6	0.02	0.07	0.03		0.18		
	8		0.11	0.15	0.36	0.73	0.06	
	9			0.01				
	0				0.04			
PGM	1	0.02		0.09	0.04	0.05	0.03	0.03
	2	0.98	1.00	0.82	0.96	0.96	0.97	
	3			0.09				
ADK	1			0.03				0.03
	2	0.43	0.39	0.53	0.29	0.55	0.31	
	4	0.45	0.61	0.44	0.71	0.46	0.69	
	6	0.10		0.01				
	8	0.02						
SOD	2	0.36	0.46	0.47	0.79	0.59	0.22	0.09
	3						0.09	
	4	0.62	0.52	0.53	0.21	0.41	0.69	
	6	0.02						
	8		0.02					

a. See Materials and Methods in Chapter 2 for calculation of G_{ST}.

Table 3.3. Linkage disequilibrium analysis of individual and hierarchical combinations of populations of *R. leguminosarum* bvs. *viciae* and *trifolii*.

		Mismatch Parameters				Statistical Significance[a]		
Population	No. ETs	X[b]	V_O[c]	V_e[c]	V_O/V_e	A	B	C
CV	78	6.30	3.86	2.38	1.63	$V_e<3.67<V_O<4.06$	$1.64<V_e<3.11<V_O$	P<0.0001
CT	28	5.83	3.74	2.15	1.74	$V_e<3.26<V_O<4.34$	$1.03<V_e<3.27<V_O$	P<0.0001
AV	42	6.59	5.05	2.49	2.03	$V_e<4.60<V_O<5.56$	$1.43<V_e<3.54<V_O$	P<0.0001
AT	54	6.14	4.18	2.41	1.73	$V_e<3.89<V_O<4.50$	$1.52<V_e<3.31<V_O$	P<0.0001
AVCV	109	6.64	3.92	2.47	1.58	$V_e<3.78<V_O<4.07$	$1.82<V_e<3.13<V_O$	P<0.0001
ATCT	81	6.23	3.60	2.31	1.56	$V_e<3.43<V_O<3.78$	$1.61<V_e<3.01<V_O$	P<0.0001
OR	187	6.59	3.36	2.39	1.41	$V_e<3.29<V_O<3.43$	$1.91<V_e<2.88<V_O$	P<0.0001
WA	22	4.91	5.26	2.05	2.57	$V_e<4.42<V_O<6.37$	$0.87<V_e<3.23<V_O$	P<0.0001
PNW	208	6.68	3.37	2.36	1.43	$V_e<3.31<V_O<3.43$	$1.91<V_e<2.81<V_O$	P<0.0001
UK	32	6.14	3.41	2.38	1.44	$V_e<3.02<V_O<3.88$	$1.22<V_e<V_O<3.53$	P=0.0001
INT	234	6.70	3.25	2.39	1.36	$V_e<3.20<V_O<3.31$	$1.96<V_e<2.82<V_O$	P<0.0001

a. Significance of the inflation of V_O over V_e. A: asymmetric χ^2 method (Sokal and Rohlf 1981); B: method of Brown, Feldman & Nevo (1980); C: probability of V_O exceeding V_e by observed amount by chance alone.

b. X is the mean number of mismatched loci (out of 13) over $n(n-1)/2$ paired ET comparisons.

c. V_O and V_e are the observed and expected variance, respectively, of the allelic mismatch frequency distribution for each population; see Materials and Methods for details.

Table 3.4. Linkage disequilibrium within and between groups in the WA and UK populations of *R. leguminosarum* bv. *viciae*.

Comparison[a]	V_O/V_e	Significance[b]
UK Population:		
Total Population	1.44	$1.22<\mathbf{V_e=2.38}<\mathbf{V_O=3.41}<3.53$
Group A	1.29	$0.73<\mathbf{V_e=2.28}<\mathbf{V_O=2.95}<3.84$
Group H	0.64	$0.14<\mathbf{V_O=0.90}<\mathbf{V_e=1.41}<2.67$
Group I	1.51	$0.00<\mathbf{V_e=1.62}<\mathbf{V_O=2.44}<3.54$
Groups A + C	1.38	$0.84<\mathbf{V_e=2.35}<\mathbf{V_O=3.24}<3.87$
Groups H + I	1.94	$0.59<\mathbf{V_e=2.09}<3.59<\mathbf{V_O=4.04}$
WA Population:		
Total Population	2.57	$0.87<\mathbf{V_e=2.05}<3.23<\mathbf{V_O=5.26}$
Group C	0.63	$0.28<\mathbf{V_O=0.75}<\mathbf{V_e=1.20}<2.13$
Group E	0.87	$0.06<\mathbf{V_O=1.23}<\mathbf{V_e=1.41}<2.76$
Groups A + C	2.83	$0.51<\mathbf{V_e=1.80}<3.08<\mathbf{V_O=5.07}$
Groups C + E	2.57	$0.81<\mathbf{V_e=2.05}<3.29<\mathbf{V_O=5.27}$

a. See Fig. 2.3 for inter-group relationships; group I, not shown in that figure, clusters adjacent to group C at a relative similarity of approximately 0.55.

b. Significance of V_O inflation over V_e determined by method of Brown, Feldman & Nevo (1980).

Table 3.5. Linkage disequilibrium within and between groups A and C in the WA and UK populations of *R. leguminosarum* bv. *viciae*.

Comparison	V_O/V_e	Significance[a]
Group A (UK)	1.29	$0.73<\mathbf{V_e=2.28<V_O=2.95}<3.84$
Group A (WA + UK)	1.29	$0.73<\mathbf{V_e=2.28<V_O=2.95}<3.84$
Group C (WA)	0.63	$0.28<\mathbf{V_O=0.75<V_e=1.20}<2.13$
Group C (WA + UK)	0.90	$0.39<\mathbf{V_O=1.24<V_e=1.38}<2.37$

a. Significance of V_O inflation over V_e determined by method of Brown, Feldman & Nevo (1980).

Figure 3.1. Results of bootstrap analyses showing 95% confidence limits for the sampling distributions of V_o/V_e in individual and hierarchical combinations of populations of *R. leguminosarum* bvs. *viciae* and *trifolii* calculated based on (A) ETs and (B) isolates.

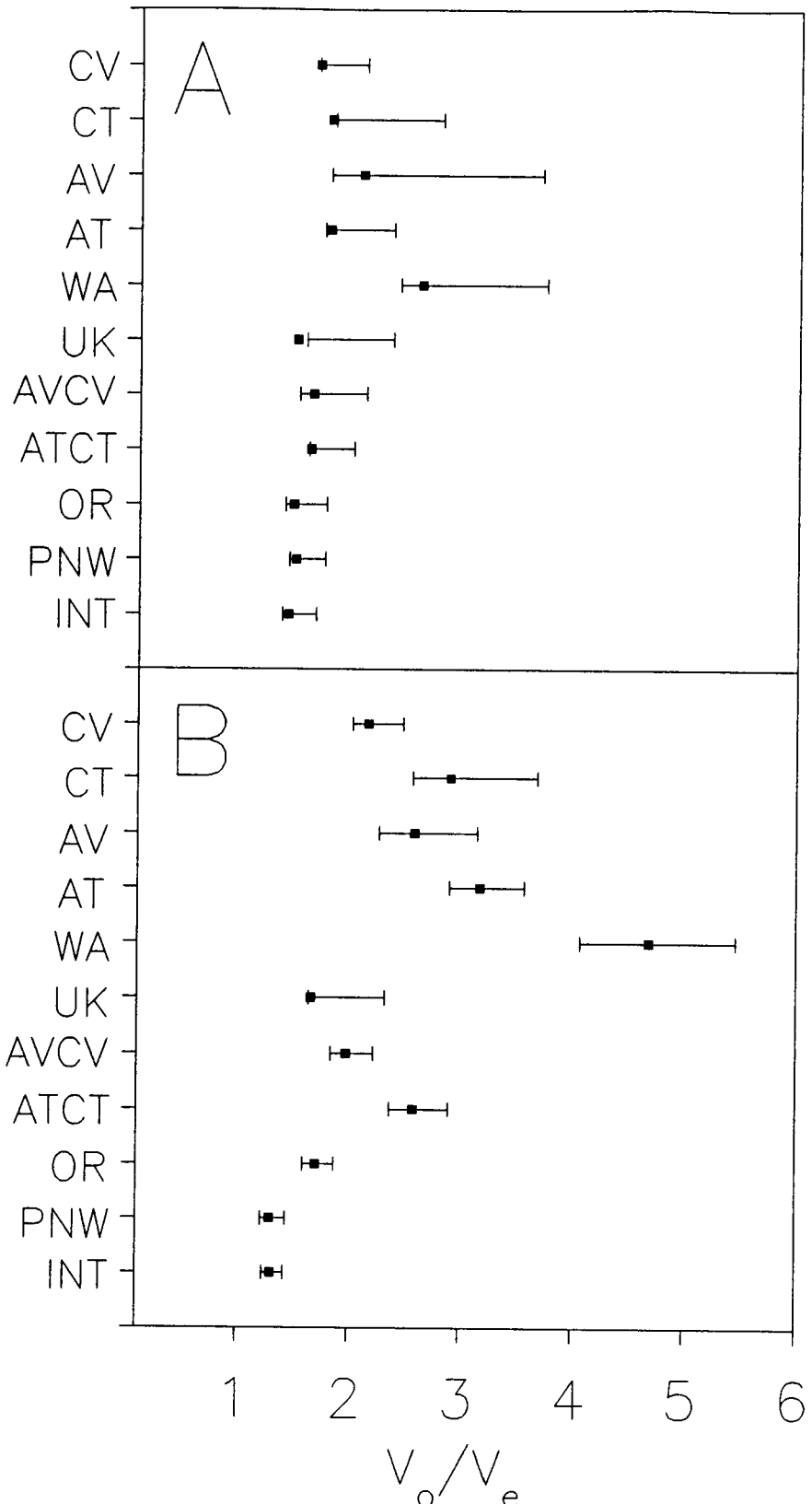
A
CV
CT
AV
AT
WA
UK
AVCV
ATCT
OR
PNW
INT
B
CV
CT
AV
AT
WA
UK
AVCV
ATCT
OR
PNW
INT
1
2
3
4
5
6
V_e/V_o

Figure 3.1

Figure 3.2. Comparison of the major groups of ETs defined in the WA and UK populations of *R. leguminosarum* bv. *viciae*.

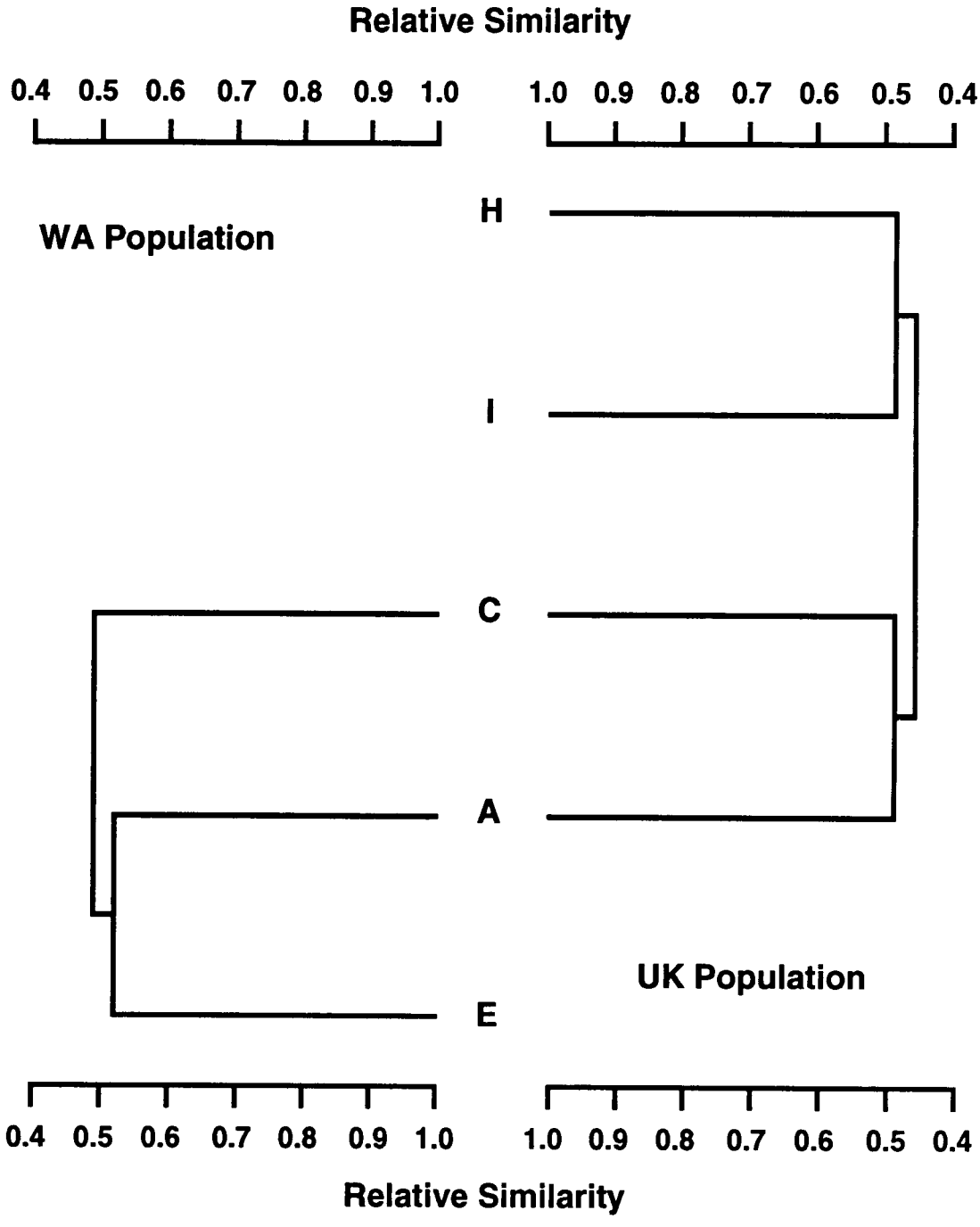
Relative Similarity
0.4 0.5 0.6 0.7 0.8 0.9 1.0
1.0 0.9 0.8 0.7 0.6 0.5 0.4
WA Population
H
I
C
A
E
UK Population
0.4 0.5 0.6 0.7 0.8 0.9 1.0
1.0 0.9 0.8 0.7 0.6 0.5 0.4
Relative Similarity

Figure 3.2

Figure 3.3. Allelic mismatch frequency distributions based on $n(n\text{-}1)/2$ paired ET comparisons for the (A) WA (n=22) and (B) UK (n=32) populations of *R. leguminosarum* bv. *viciae*.

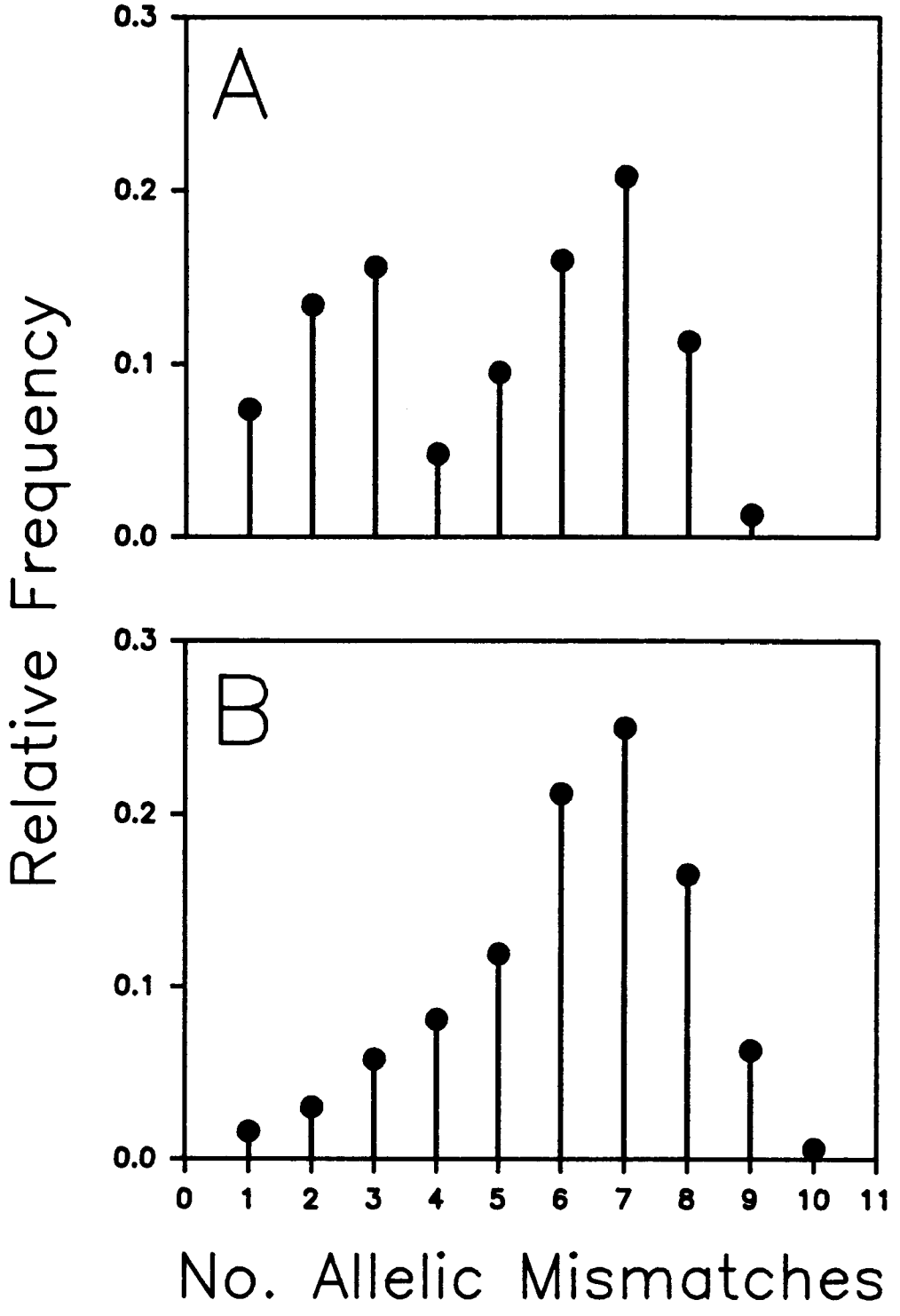
A
B
0.3
0.2
0.1
0.0
Relative Frequency
0 1 2 3 4 5 6 7 8 9 10 11
No. Allelic Mismatches

Figure 3.3

Chapter IV

Summary to the Thesis

In this thesis, I have spoken of genetic recombination, migration of organisms, and mutation as forces influencing the structure of populations of *Rhizobium leguminosarum* bvs. *viciae* and *trifolii*. In this chapter, I will attempt to unify my findings into an overall theme addressing population structure and dynamics in this species.

Although the evidence for migration between the populations studied herein is indirect, it is nonetheless compelling. I have presented data that indicates that genetic diversity in local populations of *R. leguminosarum* bvs. *viciae* and *trifolii* is very close to that seen for the species as a whole. Furthermore, the majority of alleles detected at each locus are similar across each of the six populations I studied, and differentiation in allele frequencies is low between the populations even though certain of them are separated from one another by hundreds or thousands of miles. If migration were not occurring frequently between the populations, it is unlikely that the populations would have maintained genetic similarities over time in the face of stochastic processes such as mutation and random genetic drift. Alternatively, it can be hypothesized that founder effects (Avery & Hill 1979) whereby a single population gave rise at some point to the other populations may have generated the genetic similarity between the populations. However, the complexity of the *Rhizobium*-legume symbiosis suggests that it is not an evolutionarily recent event (Long 1992; Sprent & Raven 1992), and stochastic processes would most likely have operated within the separate populations as described previously to diverge the populations genetically.

Although I have defined groups which are found in only one or a few of the populations, this observation is not inconsistent with the idea

that frequent migration is occurring between the populations. Several hypotheses may be proposed to explain the absence of groups from certain populations. Simply because organisms representing a genotype or range of genotypes are introduced onto a site by migration by no means indicates that the immigrants will become established in the population. The success or failure of the immigrants to establish themselves could be a reflection of the extent of local adaptation to the conditions on a specific site, or could be due to random chance. In this context, the wide distribution of ETs in groups A and C could be a manifestation of the migration of successful genotypes throughout the world. Alternatively, the differences in group representation between populations might be due to periodic selection (Atwood, Schneider & Ryan 1951; Levin 1981), or to the random extinction of ETs in the populations. In these cases, the implication is that the rate of migration is sufficiently low to allow these differences to accumulate, but high enough that the populations do not permanently diverge. In the absence of direct measurements on rates of migration, this possibility cannot be excluded.

Although I cannot discount the existence of genetic recombination between ETs within a population, the existence of significant levels of linkage disequilibrium in each of the populations of *R. leguminosarum* indicates that recombination cannot be so frequent as to break down multilocus structure. However, the large amounts of genetic diversity within certain of the groups, most notably A and C, may be a result of random mutation and genetic recombination which would be effective within but not between sites. Thus, the observation that there are common groups of ETs in geographically separated populations, but that

ET overlap is a relatively rare occurrence, may be a consequence of local diversification. In this regard, perhaps ecological studies on bacterial populations should be focussed at the group level rather than on individual ETs.

The field of bacterial population genetics and ecology is still quite young. Consequently, the underlying principles governing population structure in bacterial species are just now beginning to be resolved, and there are a number of issues that have arisen in recent years that must be addressed in detail. As further studies on populations of bacteria both in the soil and in other environments are undertaken, a more firm foundation will be established from which to formulate ecologically and evolutionarily meaningful hypotheses.

Bibliography

Allen ON, Allen EK (1981) The Leguminosae. The University of Wisconsin Press, Madison, Wisconsin.

Atwood KC, Schneider LK, Ryan FJ (1951) Periodic selection in *Escherichia coli*. Proceedings of the National Academy of Sciences of the USA, 37, 146-155.

Avery PJ, Hill WG (1979) Distribution of linkage disequilibrium with selection and finite population size. Genetics Research, 33, 29-48.

Bottomley PJ (1992) Ecology of *Bradyrhizobium* and *Rhizobium*. In: Stacey G, Evans HJ, Burris RH (eds.) Biological Nitrogen Fixation, pp. 293-348. Chapman and Hall, New York.

Brockman FJ, Bezdicek DF (1989) Diversity within serogroups of *Rhizobium leguminosarum* bv. *viceae* in the Palouse region of eastern Washington as indicated by plasmid profiles, intrinsic antibiotic resistance and topography. Applied and Environmental Microbiology, 55, 109-115.

Brown AHD, Feldman MW (1981) Population structure of multilocus associations. Proceedings of the National Academy of Sciences of the USA, 78, 5913-5916.

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics, 96, 523-536.

Caugant DA, Levin BR, Lidin-Janson G, Whittam TS, Svanborg Edén C, Selander RK (1983) Genetic diversity and relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. Progress in Allergy, 33, 203-227.

Caugant DA, Levin BR, Selander RK (1984) Distribution of multilocus genotypes of *Escherichia coli* within and between host families. Journal of Hygiene, 92, 377-384.

Caugant DA, Levin BR, Selander RK (1981) Genetic diversity and temporal variation in the *E. coli* population of a human host. Genetics, 98, 467-490.

Chakrabarty S, Lee MS, Gibson AH (1981) Diversity in the nutritional requirements of strains of various *Rhizobium* species. Soil Biology and Biochemistry, 13, 349-354.

de Bruijn FJ (1992) Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* and other soil bacteria. Applied and Environmental Microbiology, 58, 2180-2187.

Demezas DH, Reardon TB, Watson JM, Gibson AH (1991) Genetic diversity among *Rhizobium leguminosarum* bv. *trifolii* strains revealed by allozyme and restriction fragment length polymorphism analyses. Applied and Environmental Microbiology, 57, 3489-3495.

Eardly BD, Materon LA, Smith NH, Johnson DA, Rumbaugh MD, Selander RK (1990) Genetic structure of natural populations of the nitrogen-fixing bacterium *Rhizobium meliloti.* Applied and Environmental Microbiology, 56, 187-194.

Efron B (1982) The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Apllied Mathematics, Philadelphia.

Evans HJ (1974) Symbiotic nitrogen fixation in legume nodules. In: Moore TC (ed.) Research Experiences in Plant Physiology, pp. 417-426. Springer-Verlag, Berlin.

Evans HJ, Burris RH (1992) Highlights in biological nitrogen fixation during the last 50 years. In: Stacey G, Evans HJ, Burris RH (eds.) Biological Nitrogen Fixation, pp. 1-42. Chapman and Hall, New York.

Flores M, González V, Pardo MA, Leija A, Martínez E, Romero D, Piñero D, Dávila G, Palacios R (1988) Genomic instability in *Rhizobium phaseoli*. Journal of Bacteriology, 170, 1191-1196.

George MLC, Robert FM (1992) Competition among *Rhizobium leguminosarum* bv. *phaseoli* strains for nodulation of common bean. Canadian Journal of Microbiology, 38, 157-160.

Gilmour MN, Whittam TS, Kilian M, Selander RK (1987) Genetic relationships among the oral streptococci. Journal of Bacteriology, 169, 5247-5257.

Graham PH (1992) Stress tolerance in *Rhizobium* and *Bradyrhizobium*, and nodulation under adverse soil conditions. Canadian Journal of Microbiology, 38, 475-484.

Harrison SP, Jones DG, Young JPW (1989) *Rhizobium* population genetics: genetic variation within and between populations from diverse locations. Journal of General Microbiology, 135, 1061-1069.

Harrison SP, Young JPW, Jones DG (1987) *Rhizobium* population genetics: effect of clover variety and inoculum dilution on the genetic diversity sampled from natural populations. Plant and Soil, 103, 147-150.

Hedrick PW, Thomson G (1986) A two-locus neutrality test: applications to humans, *E. coli* and lodgepole pine. Genetics, 112, 135-156.

Hilali A, Aurag J, Molina JAE, Schmidt EL (1989) Competitiveness and persistence of strains of *Rhizobium phaseoli* introduced into a Moroccan sandy soil. Biology and Fertility of Soils, 7, 213-218.

Hynes MF, O'Connell MP (1990) Host plant effect on competition among strains of *Rhizobium leguminosarum*. Canadian Journal of Microbiology, 36, 864-869.

Istock CA, Duncan KE, Ferguson N, Zhou X (1992) Sexuality in a natural population of bacteria - *Bacillus subtilis* challenges the clonal paradigm. Molecular Ecology, 1, 95-103.

Jarvis BDW, Dick AG, Greenwood RM (1980) Deoxyribonucleic acid homology among strains of *Rhizobium trifolii* and related species. International Journal of Systematic Bacteriology, 30, 42-52.

Jordan DC (1984) Rhizobiaceae. In: Krieg NR (ed.) Bergey's Manual of Systematic Bacteriology, vol. 1, pp. 234-256. Williams & Wilkins, Ancaster, Ontario.

Josephson KL, Bourque DP, Bliss FA, Pepper IL (1991) Competitiveness of KIM 5 and VIKING 1 bean rhizobia: strain by cultivar interactions. Soil Biology and Biochemistry, 23, 249-253.

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, UK.

Kimura M (1968) Evolutionary rate at the molecular level. Nature, 217, 624-626.

Kupicha FK (1981) Vicieae. In: Polhill RM, Raven PH (eds.) Advances in Legume Systematics, vol.2, Part 1, pp. 377-381. Royal Botanic Gardens, England.

Laguerre G, Mazurier SI, Amarger N (1992) Plasmid profiles and restriction fragment length polymorphism of *Rhizobium leguminosarum* bv. *viciae* in field populations. FEMS Microbiology Ecology, 107, 17-26.

Lenski RE (1993) Assessing the genetic structure of microbial populations. Proceedings of the National Academy of Sciences of the USA, 90, 4334-4336.

Levin BR (1981) Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. Genetics, 99, 1-23.

Lewontin RC, Kojima K (1960) The evolutionary dynamics of complex polymorphisms. Evolution, 14, 450-472.

Long SR (1992) Genetic analysis of *Rhizobium* nodulation. In: Stacey G, Evans HJ, Burris RH (eds.) Biological Nitrogen Fixation, pp. 560-597. Chapman and Hall, New York.

Mahler RL, Bezdicek DF (1980) Serogroup distribution of *Rhizobium leguminosarum* in peas in the Palouse of eastern Washington. Soil Science Society of America Journal, 44, 292-295.

Mahler RL, Bezdicek DF (1978) Diversity of *Rhizobium leguminosarum* in the Palouse of eastern Washington. Applied and Environmental Microbiology, 36, 780-782.

Martínez-Romero E, Segovia L, Mercante FM, Franco AA, Graham P, Pardo MA (1991) *Rhizobium tropici*, a novel species nodulating *Phaseolus vulgaris* L. beans and *Leucaena* sp. trees. International Journal of Systematic Bacteriology, 41, 417-426.

Masterson CL, Sherwood MT (1974) Selection of *Rhizobium trifolii* strains by white and subterranean clovers. Irish Journal of Agricultural Research, 13, 91-99.

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? Proceedings of the National Academy of Sciences of the USA, 90, 4384-4388.

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York, New York.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small sample of individuals. Genetics, 89, 583-590.

Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Annals of Human Genetics, 41, 225-233.

Ochman H, Selander RK (1984) Evidence for clonal population structure in *Escherichia coli*. Proceedings of the National Academy of Sciences of the USA, 81, 198-201.

Ochman H, Whittam TS, Caugant DA, Selander RK (1983) Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. Journal of General Microbiology, 129, 2715-2726.

Pasteur N, Pasteur G, Bonhomme F, Catalan J, Britton-Davidian J (1988) Practical Isozyme Genetics. Ellis Horwood Limited, England.

Piñero D, Martínez E, Selander RK (1988) Genetic diversity and relationships among isolates of *Rhizobium leguminosarum* biovar *phaseoli*. Applied and Environmental Microbiology, 54, 2825-2832.

Robinson AC (1969) Host selection for effective *Rhizobium trifolii* by red clover and subterranean clover in the field. Australian Journal of Agricultural Research, 20, 1053-1060.

Segovia L, Piñero D, Palacios R, Martínez-Romero E (1991) Genetic structure of a soil population of nonsymbiotic *Rhizobium leguminosarum*. Applied and Environmental Microbiology, 57, 426-433.

Segovia L, Young JPW, Martínez-Romero E (1993) Reclassification of American *Rhizobium leguminosarum* biovar *phaseoli* type I strains as *Rhizobium etli sp. nov.* International Journal of Systematic Bacteriology, 43, 374-377.

Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittam TS (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Applied and Environmental Microbiology, 51, 873-884.

Selander RK, Caugant DA, Whittam TS (1987) Genetic structure and variation in natural populations of *Escherichia coli*. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE (eds.) *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, vol. 2, pp. 1625-1648. American Society for Microbiology, Washington, D.C.

Selander RK, Levin BR (1980) Genetic diversity and structure in *Escherichia coli* populations. Science, 210, 545-547.

Sneath PHA, Sokal RR (1973) Numerical Taxonomy. W.H. Freeman and Co., San Francisco.

Sokal RR, Rohlf FJ (1981) Biometry. W.H. Freeman and Co., San Francisco.

Souza V, Nguyen TT, Hudson RR, Piñero D, Lenski RE (1992) Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? Proceedings of the National Academy of Sciences of the USA, 89, 8389-8393.

Sprent JI, Raven JA (1992) Evolution of nitrogen-fixing symbioses. In: Stacey G, Evans HJ, Burris RH (eds.) Biological Nitrogen Fixation, pp. 461-496. Chapman and Hall, New York.

Turco RF, Bezdicek DF (1987) Diversity within two serogroups of *Rhizobium leguminosarum* native to soils in the Palouse of eastern Washington. Annals of Applied Biology, 111, 103-114.

Valdivia B, Dughri MH, Bottomley PJ (1988) Antigenic and symbiotic characterization of indigenous *Rhizobium leguminosarum* bv. *trifolii* recovered from root nodules of *Trifolium pratense* L. sown into subterranean clover pasture soils. Soil Biology and Biochemistry, 20, 267-274.

Vargas AAT, Graham PH (1989) Cultivar and pH effects on competition for nodule sites between isolates of *Rhizobium* in beans. Plant and Soil, 117, 195-200.

Vincent JM (1970) A Manual for the Practical Study of Root-Nodule Bacteria. IBP handbook no. 15. Blackwell Scientific Publications, Oxford, England.

Weaver RW, Sen D, Coll JJ, Dixon CR, Smith GR (1989) Specificity of arrowleaf clover for rhizobia and its establishment in soil from crimson clover pastures. Soil Science Society of America Journal, 53, 731-734.

Whittam TS (1992) Sex in the soil. Current Biology, 2, 676-678.

Whittam TS, Ochman H, Selander RK (1983a) Geographic components of linkage disequilibrium in natural populations of *Escherichia coli*. Molecular Biology and Evolution, 1, 67-83.

Whittam TS, Ochman H, Selander RK (1983b) Multilocus genetic structure in natural populations of *Escherichia coli*. Proceedings of the National Academy of Sciences of the USA, 80, 1751-1755.

Wolff AB, Streit W, Kipe-Holt JA, Vargas H, Werner D (1991) Competitiveness of *Rhizobium leguminosarum* bv. *phaseoli* strains in relation to environmental stress and plant defense mechanisms. Biology and Fertility of Soils, 12, 170-176.

Young JPW (1985) *Rhizobium* population genetics: enzyme polymorphism in isolates from peas, clover, beans and lucerne grown at the same site. Journal of General Microbiology, 131, 2399-2408.

Young JPW, Demetriou L, Apte RG (1987) *Rhizobium* population genetics: enzyme polymorphism in *Rhizobium leguminosarum* from plants and soil in a pea crop. Applied and Environmental Microbiology, 53, 397-402.

Young JPW, Wexler M (1988) Sym plasmid and chromosomal genotypes are correlated in field populations of *Rhizobium leguminosarum*. Journal of General Microbiology, 134, 2731-2739.

Appendix

Table A.1. Buffer systems used for multilocus enzyme electrophoresis of rhizobial enzymes.

Electrode Buffer	Gel Buffer	Voltage	Enzyme
Tris (0.69 M)-citrate (0.16 M) (pH 8.0) (To prepare a stock of Tris-citrate soln.: 208.0 g of Tris, 90.8 g of anhyd. citric acid, 1.00 liter of deionized water, adjust pH to 8.25 with HCl. To prepare electrode buffer: mix 300 ml of the stock soln. with 450 ml of deionized water.)	Tris-citrate (pH 8.0) (30X dilution of the electrode buffer	130 V	IDH BGA MDH PGI SOD PEP XDH NSP G6PD
Tris (0.22 M)-citrate (0.09 M) (pH 6.3) (27.0 g of Tris, 16.5 g of anhyd. citric acid, 1.00 liter of deionized water, pH adjusted with NaOH)	Tris (8 mM)-citrate (3 mM) (pH 6.7) (1.0 g of Tris, 0.6 g of anhyd. citric acid, 1.00 liter of deionized water, pH adjusted with NaOH)	150 V	6PG HBD
Borate (0.30 M) (pH 8.2) (18.5 g of boric acid, 2.4 g of NaOH, 1.00 liter of deionized water)	Tris (10 mM)-HCl (pH 8.5) (1.2 g of Tris, 1.00 liter of deionized water, pH adjusted with HCl)	250 V	ADK PGM

Table A.1 (continued)

Enzyme	Abbreviation	Enzyme	Abbreviation
1. Isocitrate dehydrogenase	IDH	8. Nucleoside phosphorylase	NSP
2. β-Galactosidase	BGA	9. 6-Phosphogluconate dehydrogenase	6PG
3. Glucose-6-phosphate dehydrogenase	G6PD	10. β-Hydroxybutyrate dehydrogenase	HBD
4. Malate dehydrogenase	MDH	11. Superoxide dismutase	SOD
5. Phosphoglucose isomerase	PGI	12. Adenylate kinase	ADK
6. Peptidase	PEP	13. Phosphoglucomutase	PGM
7. Xanthine dehydrogenase	XDH		

Table A.2. Composition of enzyme assay solutions.

Enzyme	Substrate, coupling enzymes and coenzyme	Assay buffer and salt supplements	Dye
IDH	0.1 M isocitrate[a] (trisodium salt) (1 ml) NADP[b] (1 ml)	Tris-HCl[c] (pH 8.0) (10 ml) $MgCl_2$[d] (2 ml) deionized water (40 ml)	MTT[e] (1 ml) PMS[f] (0.5 ml)
BGA	6-bromo-2-naphthyl-β-galactopyranoside (10 mg in 5 ml methanol)	Phosphate-citrate[g] (pH 5.0) (8.5 ml) deionized water (30 ml)	Tetrazotized *o*-dianisidine[g] (30 mg)
G6PD	Glucose-6-phosphate (disodium salt hydrate) (50 mg) NADP (1 ml)	Tris-HCl (pH 8.0) (10 ml) $MgCl_2$ (1 ml) deionized water (40 ml)	MTT (1 ml) PMS (0.5 ml)
MDH	2.0 M malate[h] (sodium salt) (3 ml) NAD[i] (1 ml)	Tris-HCl (pH 8.0) (8 ml) deionized water (32 ml)	MTT (1 ml) PMS (0.5 ml)
PGI[j]	Fructose-6-phosphate (10 mg) Glucose-6-phosphate dehydrogenase (3 units) NADP (1 ml)	Tris-HCl (pH 8.0) (5 ml) $MgCl_2$ (0.3 ml) deionized water (20 ml)	MTT (1 ml) PMS (0.5 ml)

Table A.2 (continued)

PEP[j]	Dipeptide Gly-Leu (20 mg) Peroxidase (5 mg) Snake Venom (5 mg)	Tris-HCl (pH 8.0) (5 ml) $MnCl_2$[k] (0.5 ml) deionized water (20 ml)	*o*-Dianisidine dihydrochloride (10 mg)
XDH	Hypoxanthine (100 mg) NAD (1 ml)	Tris-HCl (pH 8.0) (10 ml) deionized water (40 ml)	MTT (1 ml) PMS (0.5 ml)
NSP[j]	Inosine (20 mg) Xanthine Oxidase (2 units)	Sodium phosphate[l] (pH 7.0) (1 ml) deionized water (24 ml)	MTT (1 ml) PMS (0.5 ml)
6PG	6-Phosphogluconic acid (barium salt) (10 mg) NADP (1 ml)	Tris-HCl (pH 8.0) (4 ml) $MgCl_2$ (10 ml) deionized water (16 ml)	MTT (1 ml) PMS (0.5 ml)
HBD	DL-β-hydroxybutyric acid (disodium salt) (100 mg) NAD (1 ml)	Tris-HCl (pH 8.0) (10 ml) deionized water (40 ml)	MTT (1 ml) PMS (0.5 ml)
SOD[j]	Assayed together with NSP	Sodium phosphate (pH 7.0) (1 ml) deionized water (24 ml)	MTT (1 ml) PMS (0.5 ml)

Table A.2 (continued)

ADK[j]	Glucose (100 mg) ADP (25 mg) Hexokinase (1 mg) Glucose-6-phosphate dehydrogenase (15 units) NADP (0.1 ml)	Tris-HCl (pH 8.0) (5 ml) $MgCl_2$ (1 ml) deionized water (20 ml)	MTT (1 ml) PMS (0.5 ml)
PGM	Glucose-1,6-diphosphate and Glucose-1-phosphate[m] (5 mg) Glucose-6-phosphate dehydrogenase (15 units) NADP (50 µl)	Tris-HCl (pH 8.0) (1 ml) $MgCl_2$ (5 ml) deionized water (29 ml)	MTT (1 ml) PMS (0.5 ml)

a. Isocitrate solution (trisodium salt) (0.1 M): dissolve 2.94 g of DL-isocitric acid·H_2O (trisodium salt) in 100 ml of deionized water.

b. NADP (disodium salt) solution (13 mM): dissolve 1.0 g NADP (disodium salt) in 100 ml of deionized water; store solution in small aliquots (1 ml) at -20°C.

c. Tris-HCl buffer solution (1.0 M): dissolve 121.0 g of Trizma Base in 1.00 liter of deionized water; adjust pH to 8.0 with HCl.

d. $MgCl_2$ solution (0.1 M): dissolve 2.03 g of $MgCl_2 \cdot 6H_2O$ in 100 ml of deionized water.

Table A.2 (continued)

e. MTT solution (24 mM): suspend 1.00 g of 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyltetrazolium bromide in 100 ml of deionized water; store solution in small aliquots (1 ml) at -20°C.

f. PMS solution (33 mM): dissolve 1.00 g of phenazine methosulfate in 100 ml of deionized water; store solution in small aliquots (0.5 ml) at -20°C.

g. Phosphate-citrate buffer solution (pH 5.0): dissolve 1.46 g of Na_2HPO_4 (anhyd.) (0.1 M) and 0.93 g of citric acid (anhyd.) (0.05 M) in 100 ml of deionized water; adjust pH to 5.0 with HCl or NaOH if necessary. To stain for BGA, incubate gel slice in the substrate solution at 37°C for approximately 30 min, stain with a solution containing 30 mg of dye in 30 ml of deionized water (adjust the dye solution to pH 7.8 with a saturated $NaHCO_3$ solution), then continue to incubate at 37°C overnight.

h. Malate solution (2.0 M): dissolve 26.8 g of DL-malic acid and 16.0 g of NaOH in 100 ml of deionized water in an ice bath to prevent overheating.

i. NAD solution (15 mM): dissolve 1.0 g of NAD (free acid) in 100 ml of deionized water; store solution in small aliquots (1 ml) at -20°C.

j. Agarose overlay: suspend 0.5 g of agarose in a solution mixture of 1.0 M Tris-HCl buffer (pH 8.0, 5 ml) and 20 ml of deionized water, steam for 20 min, cool in a 55°C water bath, mix with the stain solution, and pour evenly over gel slice.

k. $MnCl_2$ solution (0.25 M): dissolve 4.90 g of $MnCl_2 \cdot H_2O$ in 100 ml of deionized water.

Table A.2 (continued)

l. Sodium phosphate buffer solution (0.2 M, pH 7.0): dissolve 1.38 g of $NaH_2PO_4 \cdot H_2O$ and 2.68 g of $Na_2HPO_4 \cdot 7H_2O$ in 100 ml deionized water.

m. The glucose-1-phosphate preparation contains sufficient glucose-1,6-diphosphate as a contaminant for the staining reaction.

Table A.3. Allelic profiles of *R. leguminosarum* bv. *trifolii* strains previously characterized by DNA:DNA hybridization.

	Allele at enzyme locus													
Strain	BGA	G6PD	IDH	MDH	HBD	6PG	PEP	PGI	NSP	PGM	ADK	SOD	ΔEM[a]	% Hybrid.[b]
CC275e	3	1	2	1	16	2	5	2	4	2	4	4	0	100
CC2480a	4	1	4	1	2	2	2	2	5	2	4	2	6	71
TA1	2	1	4	1	4	5	2	2	3	2	2	2	8	65
TA2	2	1	4	1	4	5	2	2	3	2	2	2	8	70
NZP549	7	1	2	1	17	2	6	3	4	2	6	4	5	82
NZP560	3	1	2	1	16	2	5	4	4	2	4	4	1	87
NZP550/2	2	1	6	1	2	2	5	2	2	2	4	2	5	62
NZP5117	3	1	2	1	2	2	5	3	3	2	6	2	5	81
SU202	3	1	2	1	17	5	5	2	5	2	6	4	4	81
UNZ29	6	1	4	1	7	2	6	2	2	2	4	2	6	66
WU95	3	1	2	1	2	2	5	3	5	2	6	2	5	74

a. Number of loci (out of 12) at which the strain differs from CC275e.
b. Percent DNA hybridization between strain and CC275e (Jarvis *et al.*, 1980).

Table A.4. Allelic profiles of *R. leguminosarum* bv. *viciae* isolates in the CV population.

			Allele at Locus												
Group	ET	n	Bga	G6p	Idh	Mdh	Hbd	6pg	Pep	Pgi	Xdh	Nsp	Pgm	Adk	Sod
A	110	1	2	1	5	1	7	2	6	2	3	4	2	2	4
	100	3	2	1	5	1	7	2	6	2	4	4	2	2	4
	141	1	2	1	5	1	7	2	5	2	3	4	2	2	4
	163	1	2	1	5	1	4	2	4	6	4	4	2	2	4
	222	1	2	1	4	1	4	5	2	4	3	4	1	2	4
	174	1	2	1	4	1	4	5	2	4	3	4	2	2	4
	130	1	2	1	4	1	4	5	2	4	2	3	2	2	4
	220	2	7	1	4	1	4	2	2	4	3	4	2	2	4
	217	1	7	1	4	1	4	2	2	4	4	4	2	2	4
	221	1	7	1	4	1	4	2	2	4	2	4	2	2	4
	89	1	7	1	4	1	7	2	2	4	3	4	3	2	4
	21	2	6	1	4	1	7	2	2	2	3	2	2	2	4
	17	3	6	1	4	1	7	2	2	4	3	2	2	2	4
	19	18	6	1	4	1	7	2	2	4	2	2	2	2	4
	30	4	4	1	4	1	7	2	2	4	2	2	2	2	4
	99	1	6	1	4	1	7	2	2	4	1	2	2	2	4

Table A.4 (continued)

162	1	6	1	4	1	7	2	3	4	2	2	2	2	4
3	2	6	1	4	1	7	2	2	2	4	2	2	2	4
7	3	4	1	4	1	7	2	2	2	4	2	2	2	4
218	1	3	1	5	1	7	2	2	2	4	2	2	2	4
2	1	3	1	4	1	7	2	6	2	3	2	2	4	4
15	6	3	1	4	1	7	2	6	2	2	2	2	4	4
27	3	2	1	4	1	7	2	6	2	2	2	2	2	4
94	1	2	1	4	1	7	2	6	2	3	2	2	2	4
133	2	2	1	4	1	7	2	6	2	1	2	2	2	4
97	1	3	1	4	1	7	2	6	4	1	2	2	2	4
95	2	3	1	4	1	7	2	6	4	2	2	2	2	4
104	1	2	1	4	1	7	2	6	4	1	2	2	2	4
142	1	9	1	4	1	7	2	5	2	1	2	2	2	4
102	4	3	1	4	1	4	2	6	4	2	4	2	2	2
11	14	3	1	4	1	4	2	6	4	2	4	2	2	4
216	1	3	1	4	1	4	2	6	4	3	4	2	2	2
143	1	6	1	4	1	4	2	5	4	2	2	2	2	4
129	1	6	1	4	1	4	2	6	4	2	2	2	2	4

Table A.4 (continued)

	214	4	6	1	4	1	4	2	6	4	3	2	2	2	4
	107	1	4	1	4	1	4	2	6	4	3	2	2	1	4
	92	2	6	1	4	1	4	2	6	4	4	2	2	1	4
	87	1	7	1	4	1	1	2	6	2	4	2	2	4	4
B	103	1	7	1	4	1	4	6	1	5	5	2	2	4	4
	86	2	6	1	4	1	6	6	1	5	5	2	2	4	4
	161	2	7	1	4	1	6	6	2	5	5	2	2	4	4
	159	5	3	1	4	1	6	5	2	4	5	2	2	4	4
C	101	4	2	1	4	1	2	2	5	2	4	8	2	4	2
	85	1	2	1	4	1	2	2	5	2	1	8	2	4	2
	108	2	2	1	4	1	2	2	5	2	3	8	2	4	2
	213	2	2	1	5	1	2	2	4	2	3	8	2	4	2
	144	1	2	1	4	1	2	2	4	2	3	8	2	4	2
	212	2	2	1	5	1	2	2	4	2	3	4	2	4	2
	115	1	2	1	5	1	2	5	5	2	2	8	2	4	2
	109	1	2	1	5	1	2	2	5	2	3	9	2	6	2
	106	1	4	1	4	1	2	2	2	1	4	8	2	4	2
	90	2	4	1	4	1	2	2	2	1	4	8	1	4	2

Table A.4 (continued)

113	2	4	1	4	1	2	2	2	1	4	2	2	2	2
137	1	3	1	4	1	2	2	2	1	3	8	1	4	2
105	1	2	1	4	1	2	2	2	2	5	8	1	4	2
98	2	3	1	4	1	2	2	3	2	5	8	1	4	2
128	1	9	1	4	1	2	2	4	2	4	8	1	4	2
140	1	3	1	4	1	2	15	4	2	2	6	2	4	2
136	1	2	1	4	1	4	2	2	6	3	3	3	4	2
132	2	3	1	4	1	4	2	2	6	3	3	3	4	2
138	1	4	1	4	1	4	2	2	6	3	3	3	4	2
114	1	3	1	4	1	4	2	2	6	5	4	3	4	2
88	1	3	1	4	1	4	2	2	6	5	4	2	4	2
139	1	3	1	4	1	4	2	2	6	3	4	3	4	2
96	1	3	1	4	1	4	2	2	6	4	4	3	4	2
74	20	3	1	4	1	4	2	2	6	4	4	2	4	2
135	1	2	1	4	1	4	5	4	2	2	3	2	4	2
127	3	2	1	4	1	4	5	5	2	2	3	2	4	2
219	1	2	1	4	1	4	9	4	2	3	4	2	4	2
112	1	2	1	4	1	4	5	2	2	2	4	2	2	2

Table A.4 (continued)

	91	1	2	1	4	1	4	5	2	2	4	4	2	2	2
	134	1	2	1	4	1	4	5	2	2	2	3	2	2	2
	5	2	2	1	4	1	4	2	6	2	3	4	2	2	2
	93	1	2	1	4	1	4	2	6	2	4	4	2	2	2
	111	1	2	1	4	1	4	2	2	2	3	4	2	2	2
	215	1	3	1	4	1	4	5	2	3	3	4	2	2	2
D	160	1	9	11	5	1	4	2	3	2	6	4	2	4	2
E	131	3	3	1	1	1	3	2	5	2	6	6	1	2	4

Table A.5. Allelic profiles of *R. leguminosarum* bv. *viciae* isolates in the AV population.

			Allele at Locus												
Group	ET	n	Bga	G6p	Idh	Mdh	Hbd	6pg	Pep	Pgi	Xdh	Nsp	Pgm	Adk	Sod
A	1	11	6	1	4	1	7	2	2	2	2	2	2	4	4
	2	2	3	1	4	1	7	2	6	2	3	2	2	4	4
	3	12	6	1	4	1	7	2	2	2	4	2	2	2	4
	4	5	6	1	4	1	7	2	2	2	3	2	2	4	4
	6	1	4	1	4	1	7	15	2	2	4	2	2	2	4
	7	1	4	1	4	1	7	2	2	2	4	2	2	2	4
	8	1	3	1	4	1	4	2	6	2	2	2	2	2	4
	11	5	3	1	4	1	4	2	6	4	2	4	2	2	4
	15	1	3	1	4	1	7	2	6	2	2	2	2	4	4
	16	1	2	1	5	1	4	2	2	5	4	4	2	2	4
	17	1	6	1	4	1	7	2	2	4	3	2	2	2	4
	18	1	6	1	4	1	6	2	2	2	2	2	2	4	4
	19	1	6	1	4	1	7	2	2	4	2	2	2	2	4
	20	1	6	1	4	1	4	2	2	2	1	2	2	2	4
	21	3	6	1	4	1	7	2	2	2	3	2	2	2	4
	24	2	2	1	4	1	7	2	9	2	2	4	2	4	4

Table A.5 (continued)

	27	1	2	1	4	1	7	2	6	2	2	2	2	2	4
	30	3	4	1	4	1	7	2	2	4	2	2	2	2	4
	125	1	3	1	4	1	4	2	6	4	2	3	2	2	2
B	23	8	3	1	4	1	6	5	1	4	4	2	2	4	4
	10	4	3	1	4	1	6	5	1	4	5	2	2	4	4
	26	2	3	1	4	1	6	5	1	5	6	2	2	4	4
	31	1	6	1	4	1	6	6	1	5	4	2	2	4	4
C	14	1	4	1	4	1	2	2	2	1	4	2	1	4	2
	119	5	2	1	4	1	4	5	2	2	3	3	2	2	2
	9	7	2	1	4	1	4	5	2	2	3	4	2	2	2
	13	1	2	1	4	0	4	5	2	2	3	2	2	2	2
	116	4	3	1	4	1	4	5	2	2	3	3	2	2	2
	5	1	2	1	4	1	4	2	6	2	3	4	2	2	2
D	28	1	3	2	4	1	4	2	4	2	4	2	2	4	2
	22	1	3	2	4	1	4	2	4	2	3	4	2	4	2
	126	1	7	2	4	1	6	2	5	2	4	2	2	4	2
	29	1	7	2	2	1	4	2	9	4	3	2	2	4	2
	117	1	7	2	4	1	6	2	1	3	5	3	2	6	2

Table A.5 (continued)

	12	1	7	2	4	1	6	2	1	3	4	2	2	6	2
	25	1	8	0	4	1	6	2	7	3	4	2	2	6	2
	122	1	7	2	5	1	6	2	6	2	3	6	2	6	2
F	123	1	4	10	10	10	0	12	8	4	4	1	2	8	6
G	121	2	3	1	4	1	6	9	9	3	3	4	2	4	4
	120	8	3	1	4	1	4	9	9	3	3	4	2	4	4
	124	4	3	1	4	1	4	9	9	1	4	4	2	4	4
	118	1	3	1	4	1	6	9	9	1	5	4	2	4	4

Table A.6. Allelic profiles of *R. leguminosarum* bv. *trifolii* isolates in the CT population.

			Allele at Locus												
Group	ET	n	Bga	G6p	Idh	Mdh	Hbd	6pg	Pep	Pgi	Xdh	Nsp	Pgm	Adk	Sod
A	153	1	2	1	4	1	4	4	4	4	4	4	2	2	4
	152	1	7	1	4	1	0	4	4	4	4	0	2	2	2
	234	1	9	1	4	1	4	5	9	4	3	4	2	2	4
	225	4	9	1	4	1	4	5	2	4	3	4	2	2	4
	223	10	7	1	4	1	0	5	2	4	3	4	2	2	2
	148	1	7	1	4	1	11	5	2	4	2	4	2	2	2
C	147	1	6	1	4	1	2	2	2	2	4	8	1	4	2
	149	1	2	10	4	1	7	2	4	2	2	4	2	4	2
	150	1	7	1	4	1	7	2	2	2	2	2	2	4	2
	151	1	2	1	4	1	3	2	6	2	4	8	2	4	2
	154	1	2	1	4	1	3	2	5	2	4	4	2	4	2
	155	1	3	1	4	1	4	2	4	6	4	4	2	4	2
	156	1	2	1	4	1	3	2	5	2	4	8	2	4	2
	157	1	7	1	4	1	7	2	2	2	3	2	2	4	2
	101	1	2	1	4	1	2	2	5	2	4	8	2	4	2
	224	4	9	1	4	1	4	2	2	4	3	2	2	4	2

Table A.6 (continued)

	226	4	2	1	4	1	7	2	2	2	2	4	2	4	2
	227	1	9	1	4	1	7	2	6	2	2	2	2	4	4
	81	1	9	1	4	1	4	2	2	4	2	2	2	4	2
	228	1	6	1	4	1	3	2	2	2	5	4	2	4	2
	229	3	9	1	4	1	2	2	9	2	4	8	2	4	2
	230	1	9	1	4	1	2	2	3	11	4	8	2	4	2
	231	1	9	1	4	1	7	2	6	2	1	2	2	4	4
	232	7	6	1	4	1	2	2	2	2	4	8	2	4	2
	233	1	3	1	4	1	2	2	9	2	4	8	2	4	2
D	158	1	9	1	4	1	4	5	3	2	4	4	2	4	4
E	145	2	3	1	1	1	2	2	5	2	5	8	2	2	2
	146	16	3	1	1	1	11	2	5	2	5	8	2	2	2

Table A.7. Allelic profiles of *R. leguminosarum* bv. *trifolii* isolates in the AT population.

			Allele at Locus												
Group	ET	n	Bga	G6p	Idh	Mdh	Hbd	6pg	Pep	Pgi	Xdh	Nsp	Pgm	Adk	Sod
A	77	2	3	1	2	1	16	2	4	2	4	2	2	4	4
	42	5	3	1	2	1	16	4	4	2	4	2	2	4	4
	43	31	3	1	2	1	4	2	4	2	4	2	2	4	4
	46	1	3	1	4	1	4	2	2	2	4	2	2	4	4
	39	1	3	1	4	1	3	2	2	2	3	2	2	4	4
	52	1	2	1	4	1	8	2	6	2	2	2	2	2	4
	55	1	2	1	5	1	3	2	6	2	2	2	2	2	4
	16	1	2	1	5	1	4	2	2	5	4	4	2	2	4
	58	2	3	1	3	1	11	2	6	4	2	4	2	2	4
	53	6	3	1	4	1	11	2	6	4	2	4	2	2	4
	82	1	7	1	4	1	0	4	10	4	2	4	2	2	2
	83	1	7	1	4	1	4	2	6	4	2	4	2	2	4
	59	2	7	1	4	1	0	4	2	4	3	4	2	2	2
	72	1	2	1	4	1	4	4	2	4	2	4	2	2	4
C	84	1	1	1	4	1	2	2	4	2	2	8	2	4	2
	74	1	3	1	4	1	4	2	2	6	4	4	2	4	2

Table A.7 (continued)

66	1	7	1	4	2	8	2	9	2	2	2	2	4	2
65	3	7	1	4	1	8	2	9	2	2	2	2	4	2
64	1	6	1	4	1	7	2	9	2	2	2	2	4	2
67	1	7	2	4	1	8	2	6	2	2	2	2	4	2
81	1	9	1	4	1	4	2	2	4	2	2	2	4	2
71	1	2	1	4	1	4	2	2	4	2	2	2	4	2
57	1	2	1	4	1	11	2	2	5	2	4	2	4	2
54	1	2	1	5	1	11	2	2	5	2	4	2	4	2
76	1	2	10	4	1	8	4	2	2	2	4	2	4	2
75	3	2	10	4	1	8	2	2	2	2	4	2	4	2
44	1	2	1	4	1	7	2	2	2	2	2	2	4	2
38	1	2	1	4	1	8	2	2	2	2	2	2	4	2
37	1	2	1	4	1	8	2	9	2	2	4	2	4	2
32	1	2	1	4	1	8	2	2	2	2	4	2	4	2
61	7	2	1	4	1	4	2	2	2	2	4	2	4	2
49	1	9	1	4	1	8	2	9	2	2	4	2	4	2
47	12	9	1	4	1	8	2	9	2	2	4	2	4	4
45	1	3	1	4	1	8	2	2	2	2	4	2	4	4

Table A.7 (continued)

	73	1	2	1	4	1	4	4	9	2	2	4	2	4	4
	70	2	2	1	4	1	4	4	2	2	2	4	2	4	4
	79	1	9	1	4	1	4	2	2	2	2	4	2	4	4
	78	1	9	1	4	1	4	4	2	2	2	4	2	4	4
	80	1	9	1	4	1	4	2	2	2	3	4	2	4	4
	51	2	9	1	4	1	4	4	9	2	3	4	2	2	4
	48	3	9	1	4	1	4	4	9	4	3	4	2	2	4
	50	1	9	1	4	1	4	4	9	2	2	2	2	2	4
	60	3	9	1	4	1	4	4	9	2	3	4	2	4	8
	68	1	2	1	4	1	4	4	2	2	2	4	2	2	2
E	33	23	6	1	1	1	11	15	4	2	6	8	2	2	2
	34	18	6	1	1	1	11	15	4	2	6	8	2	2	4
	35	13	6	1	1	1	11	2	4	2	6	8	2	2	4
	36	13	6	1	1	1	15	2	4	2	2	6	2	2	4
	40	1	6	1	1	1	11	2	4	3	6	6	2	2	2
	41	2	6	1	1	1	11	2	4	3	6	6	2	4	2
	56	1	6	1	1	1	17	2	4	2	6	8	2	2	4
	62	2	6	1	1	1	15	2	4	2	1	6	2	2	4

Table A.7 (continued)

	63	10	6	1	1	1	11	2	4	2	6	8	2	2	2
H	69	2	9	2	4	1	16	2	9	3	8	2	2	4	4

Table A.8. Allelic profiles of *R. leguminosarum* bv. *viciae* isolates in the WA population.

			Allele at Locus												
Group	ET	n	Bga	G6p	Idh	Mdh	Hbd	6pg	Pep	Pgi	Xdh	Nsp	Pgm	Adk	Sod
A	21	1	6	1	4	1	7	2	2	2	3	2	2	2	4
	189	1	6	1	4	1	7	2	2	2	2	2	2	2	4
C	194	3	1	1	4	1	2	2	5	2	3	8	2	4	2
	193	34	1	1	4	1	2	2	4	2	3	8	2	4	2
	210	2	1	1	4	1	2	2	5	2	4	8	2	4	2
	204	1	1	1	5	1	2	2	4	2	3	8	2	4	2
	199	3	1	1	5	1	2	2	4	2	2	8	2	4	2
	206	4	1	1	5	1	2	2	4	2	3	8	2	2	2
	207	1	1	1	5	1	3	2	4	2	3	8	2	4	2
	208	2	1	1	5	1	2	2	5	2	1	8	2	2	2
	205	1	1	1	5	1	2	2	5	2	2	8	2	2	2
	211	1	1	1	5	1	2	2	5	2	2	8	2	4	2
	209	1	1	1	5	1	2	2	5	2	1	8	2	4	2
	196	1	2	1	5	1	2	2	5	2	3	8	2	4	2
E	203	1	6	1	1	1	2	2	5	2	9	8	2	2	4
	202	1	6	1	1	1	4	2	5	2	9	8	2	2	4

Table A.8 (continued)

197	5	3	1	1	1	3	2	5	2	9	8	2	2	4
195	2	3	1	1	1	3	2	5	2	9	8	2	4	4
198	8	3	1	1	1	3	2	5	2	9	6	2	2	4
192	1	3	1	1	1	3	2	5	2	9	6	1	2	4
201	1	3	1	1	1	1	2	5	2	9	6	2	2	4
200	1	3	1	1	1	3	5	5	2	9	6	2	2	2

Table A.9. Allelic profiles of *R. leguminosarum* bv. *viciae* isolates in the UK population.

			Allele at Locus												
Group	ET	n	Bga	G6p	Idh	Mdh	Hbd	6pg	Pep	Pgi	Xdh	Nsp	Pgm	Adk	Sod
A	21	1	6	1	4	1	7	2	2	2	3	2	2	2	4
	164	1	4	1	4	1	7	2	2	2	3	2	2	2	4
	184	1	6	1	4	1	7	2	4	2	3	2	2	2	4
	189	1	6	1	4	1	7	2	2	2	2	2	2	2	4
	170	1	3	1	4	1	4	2	2	2	3	2	2	4	4
	169	1	3	1	4	1	4	2	6	2	3	2	2	4	4
	46	1	3	1	4	1	4	2	2	2	4	0	2	4	4
	178	1	7	1	4	1	4	2	4	2	3	2	2	4	4
	165	1	3	1	4	1	4	2	5	2	4	2	1	4	4
	191	1	2	1	4	1	4	5	4	2	3	4	2	2	2
	190	3	2	1	4	1	7	5	4	2	2	4	2	2	2
	5	1	2	1	4	1	4	2	6	2	3	4	2	2	2
	174	1	2	1	4	1	4	5	2	4	3	4	2	2	4
	168	1	2	1	4	1	4	2	2	4	3	4	2	4	4
	177	1	2	1	4	1	4	2	2	2	4	4	2	2	4
	16	1	2	1	5	1	4	2	2	5	4	0	2	2	4

Table A.9 (continued)

C	188	1	2	1	5	1	2	2	2	3	2	8	2	4	2
	186	5	2	1	4	1	2	2	4	2	4	8	2	4	2
H	176	1	7	2	4	1	2	2	9	3	1	4	2	4	4
	172	1	3	2	4	1	2	2	9	3	1	4	2	4	4
	182	2	6	2	4	1	2	2	4	3	1	4	2	4	4
	167	1	3	2	4	1	16	2	2	3	1	4	2	4	4
	180	2	7	2	4	1	0	2	1	3	2	4	2	4	4
	171	6	3	1	4	1	2	2	1	3	2	4	2	4	4
	185	1	6	2	4	1	2	2	1	3	3	4	2	4	4
	181	1	6	2	4	1	2	5	1	3	3	4	2	4	4
	173	1	6	2	4	1	2	2	1	3	2	2	2	4	4
I	187	3	7	10	4	1	21	2	2	2	1	4	2	4	3
	175	1	7	10	4	1	21	2	2	2	1	2	2	4	3
	183	2	7	10	4	1	21	2	4	2	1	4	2	4	3
	166	2	7	10	4	1	2	2	2	2	8	4	2	4	2
	179	2	7	1	4	1	2	15	2	2	1	2	2	4	2