

Reticulated and epidemic population genetic structure of *Rhizobium etli* biovar *phaseoli* in a traditionally managed locality in Mexico

C. SILVA, L. E. EGUIARTE and V. SOUZA

Departamento de Ecología Evolutiva, Instituto de Ecología, Apartado Postal 70–275, Universidad Nacional Autónoma de México, Mexico D.F., 04510, Mexico

Abstract

We conducted a multilocus enzyme electrophoresis (MLEE) study to assess the genetic structure of the nitrogen-fixing bacteria *Rhizobium etli* bv. *phaseoli*. We analysed the genetic variation at 10 enzyme-encoding chromosomal loci of 482 isolates from root nodules of *Phaseolus vulgaris* and *P. coccineus* bean plants. The isolates were obtained from six traditionally managed agricultural plots in two localities in the State of Puebla, in Central Mexico. The total mean genetic diversity (H_E) for the six plots was 0.531. Among the 482 isolates collected, 126 distinctive multilocus genotypes (electrophoretic types [ETs]) were obtained, and approximately half of the isolates are represented by five widespread ETs. A significant degree of genetic differentiation among the six plots ($G_{ST} = 0.072$) and between the two localities ($G_{ST} = 0.022$) was detected. The main part of the observed variability (70%) was found among the isolates within the plants. The cluster analysis revealed two deeply diverging lineages, separated at a genetic distance of 0.7. When a multilocus linkage disequilibrium analysis was performed at different hierarchical levels, we found significant linkage disequilibrium, but when the analysis was performed for the genotypes within the two diverging lineages, we found evidence of recombination. We propose for *R. etli* bv. *phaseoli* a reticulated and epidemic genetic structure, in which few genotypes increase in frequency to produce numerically dominant clones, and genetic exchange occurs mainly among genotypes within each lineage.

Keywords: allozyme electrophoresis, genetic structure, linkage disequilibrium, migration, *Rhizobium etli*, selection

Received 5 February 1998; revision received 28 July 1998; accepted 25 September 1998

Introduction

Most studies on the interaction between *Rhizobium* and plants have a strong emphasis on controlled environments and monocultural systems, where influences of other plants are few, due to agricultural management practices (Brockwell *et al.* 1995). Although some studies have been carried out in the sites of origin and diversification of the symbionts (Eardly *et al.* 1990; Demezas *et al.* 1991; Segovia *et al.* 1991; Souza *et al.* 1992, 1994), none of them had studied the genetic structure of rhizobia in a traditionally managed agricultural system. This is an interesting question, as the traditionally managed plots are complex plant

communities, where interactions among different plant species, at the aerial and root level, are the rule (Gliessman 1986, 1990; Alcorn 1990; Souza *et al.*, in press).

In Mexico, the annual bean *Phaseolus vulgaris* and the related perennial species *P. coccineus* are normally nodulated by strains of the nitrogen-fixing bacteria *Rhizobium etli* bv. *phaseoli* (Segovia *et al.* 1993; Souza *et al.* 1994). *P. vulgaris* and *P. coccineus* bean landraces are traditionally cultivated in an agricultural system known as milpa (Altieri & Merrick 1987; Alcorn 1990). In this system, beans are intercropped with corn and squash, as well as with diverse plant species that are locally used for medicinal and nutritional purposes (Altieri & Merrick 1987; Alcorn 1990; Souza *et al.*, in press). The milpa system is an ancient method of bean culture, and is probably the original ecosystem in which the beans were

Correspondence: C. Silva. Fax: +52-5-622-89-95; E-mail: csilva@miranda.ecologia.unam.mx

domesticated by the inhabitants of Mesoamerica (Bush 1986; Altieri & Merrick 1987).

Several studies on the genetic structure of rhizobia associated with beans have been carried out using collections of isolates from different geographical origins (Piñero *et al.* 1988; Geniaux *et al.* 1993; Laguerre *et al.* 1993). Piñero *et al.* (1988) analysed *Phaseolus* isolates from various geographical sources, mainly in Mexico, and they described a high genetic diversity and a strong linkage disequilibrium, suggesting a clonal structure (i.e. lack of genetic exchange). A number of those Mexican isolates originally classified as *R. leguminosarum* bv. *phaseoli* were reclassified as *R. etli* bv. *phaseoli*, and also significant multilocus linkage disequilibrium was detected when only *R. etli* bv. *phaseoli* electrophoretic types (ETs) were considered (Souza *et al.* 1992). But as pointed out by Maynard Smith *et al.* (1993), linkage disequilibrium may arise in samples in which strains or populations are geographically isolated. Furthermore, linkage disequilibrium may be minimal in some local populations of bacterial species such as *Bacillus subtilis* (Istock *et al.* 1992), *Bradyrhizobium* sp. (Bottomley *et al.* 1994), *Burkholderia cepacia* (Wise *et al.* 1995) and *Pseudomonas* (Haubold & Rainey 1996). In *R. etli*, some studies at a local scale have found lower levels of linkage disequilibrium than the one reported by Piñero *et al.* (1988) (Souza *et al.* 1992, 1994; Gordon *et al.* 1995), suggesting that recombination occurs locally, and that the previous evidences of clonality were the product of geographical isolation. The reproductive isolation among genetically distant isolates can also be a source of linkage disequilibrium (Maynard Smith *et al.* 1993). Evidence of this has been observed in *R. meliloti* from wild *Medicago* in the Mediterranean basin and from *Medicago* cultivated worldwide (Eardly *et al.* 1990). This population structure is described by Maynard Smith *et al.* (1993) as reticulated, 'in which recombination does not occur between isolates from the major branches, but frequent recombination occurs between isolates within each major branch'. An epidemic structure could be another source of linkage disequilibrium in bacteria (Maynard Smith *et al.* 1993). In these populations, recombination occurs but, occasionally, a highly successful genotype arises and increases in frequency to produce an epidemic clone. Maynard Smith *et al.* (1993) reported an epidemic genetic structure for clinical isolates of *Neisseria meningitidis* collected worldwide.

In the present study we analysed the genetic structure of native *R. etli* bv. *phaseoli* populations from nodules in beans grown in a traditionally managed locality in Puebla, Mexico. This study was designed with a sampling scheme at different hierarchical levels, and the multilocus enzyme electrophoretic (MLEE) data were examined with the purpose of elucidating the importance of the different evolutionary forces in shaping the population genetic structure of *R. etli* bv. *phaseoli* in a complex agricultural system.

Materials and methods

Description of the site and sampling procedure

San Miguel Acuexcomac is a small village located in the state of Puebla, in Central Mexico (98°05'W, 18°50'N), at 2100 m above sea level, \approx 36 km from Puebla city. We studied the nitrogen-fixing bacteria *Rhizobium etli* bv. *phaseoli* associated with two species of bean landraces: *Phaseolus vulgaris* (common bean, 'mantequilla' variety) and *P. coccineus* (red runner or ayocote bean) in six traditionally managed agricultural plots. In order to study the geographical differentiation among local populations, the sampled plots were distributed in two localities, one called 'town', which includes the cultivated plots that are near the village houses, and the other called 'field', which includes plots that are \approx 2 km from the village (Fig. 1). The soil conditions are different between these localities. Town soils have a sandy clay loam texture and a higher content of total nitrogen (139.0 vs. 117.2 p.p.m.), nitrates (1.9 vs. 0.9 p.p.m.), organic matter (2.9 vs. 1.9%) and pH (8.4 vs. 8.0) than field soils. In contrast, field soils have a clay texture and a higher content of magnesium (85.0 vs. 17.5 meq/100 g) and calcium (431.7 vs. 185.0 meq/100 g) than town soils (Souza *et al.*, in press, unpublished data).

In each locality, three milpa plots were sampled: plots A, B and C are town plots, and D, E, and F are field plots (Fig. 1). The mean distance among the plots of each locality was \approx 100 m, and the mean distance between the town and field localities was \approx 2 km. In each plot, 10 plants were randomly chosen (including *P. coccineus* plants when present) at a mean distance of \approx 5 m, and 15 nodules were sampled per plant. The nodules were washed with 15% sodium hypochlorite, rinsed twice with sterile water, squashed on to Petri dishes with peptone yeast extract medium (PY) and incubated for 2 days at 30 °C (Souza *et al.* 1994). A single

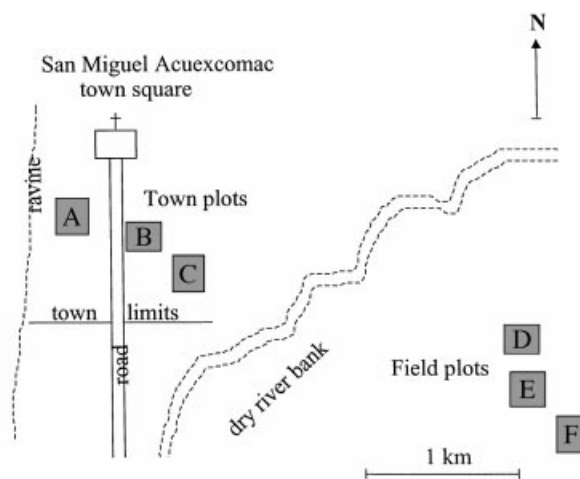


Fig. 1 Schematic representation of the six plots in both field and town localities. The letters refer to plot designations.

colony was restreaked in a new PY Petri dish and incubated for 2 days. In order to obtain pure *R. etli* bv. *phaseoli* isolates, a single colony was restreaked on to PY medium with added nalidixic acid (20 µg/mL) and incubated for 2 days. As described by Segovia *et al.* (1993), all the *R. etli* isolates are resistant to nalidixic acid. Each isolate so obtained was kept at -80 °C in glycerol peptone liquid medium (UL) (Souza *et al.* 1994). We eliminated from the analysis the plants in which less than five isolates could be recovered, to ensure a statistically more adequate sample size. The sampling scheme and the number of recovered isolates are shown in Table 1.

MLEE

Frozen isolates were streaked on to PY plates and incubated for 2 days at 30 °C and then grown in 50 mL of PY for 1 day at 30 °C. The cultures were then centrifuged at 6000 r.p.m. for 5 min, the supernatant was eliminated and the pellets were resuspended in 1 mL of buffer Tris-HCl pH 8. To break bacterial cell walls, 0.1 mL of lysozyme (7.5 mg/mL) was added; the resulting suspension was frozen at -80 °C twice for 15 min and centrifuged at 12 000 r.p.m. for 3 min. The supernatant, which contains the protein lysate, was distributed into three 0.5 mL plastic tubes and stored at -80 °C.

Electrophoresis was performed on acetate cellulose membranes following the procedures recommended by Hebert & Beaton (1993). Six enzymes were assessed: isocitrate dehydrogenase (IDH, EC 1.1.1.42), peptidase (PEP, EC 3.4.13), phosphoglucomutase (PGM, EC 5.4.2.2), glucose-6-phosphate dehydrogenase (G6PDH, EC 1.1.1.49), xanthine dehydrogenase (XDH, EC 1.1.1.204) and malate dehydrogenase (MDH, EC 1.1.1.37). For all of them, the buffer system used was Tris glycine (TG) pH 8.5 (Hebert & Beaton 1993). The enzymes IDH, PEP and PGM showed

one band of activity, G6PDH and XDH two bands, and MDH three bands, yielding a total of 10 analysed loci.

Genetic diversity

Distinctive mobility variants of each enzyme, numbered in order of decreasing anodal mobility, were considered alleles at the corresponding locus (Selander *et al.* 1986). In the case of enzymes with more than one band, each one was equated with a locus. The absence of enzyme activity was scored as a null allele and was treated as an ordinary allele. The allele profiles (ETs) for the 10 loci analysed were equated with multilocus genotypes (Selander *et al.* 1986).

Based on allele frequencies for ETs, genetic diversity for an enzyme locus was calculated as $h = (1 - \sum x_i^2)[n/(n-1)]$, where x_i is the frequency of the i th allele and n is the number of ETs (Selander *et al.* 1986). The mean genetic diversity per locus (H_E) is the arithmetic mean of h values for all loci, and represents the proportion of loci at which two randomly chosen genotypes can be expected to differ. To compute the H_E values we used the program ETDIV version 2.2 (Whittam 1990). To compare the allelic frequencies in a given locus, we used the Workman & Niswander (1970) chi-square test of heterogeneity.

Genetic differentiation

To estimate the relative genetic differentiation at different levels, we used various modified indices related to Nei's $G_{ST} = (H_T - H_S)/H_T$ (Souza *et al.* 1994), where H_T is the expected diversity in an equivalent random mating total population and H_S is the average diversity of the subpopulations. In the case of plot populations, and for total *P. vulgaris* and *P. coccineus* populations, H_S is the average genetic diversity of the plants. In the case of town, field and total populations, H_S is the average genetic diversity

Plot	No. of plants	No. of <i>P. v.</i> *	No. of <i>P. c.</i> †	No. of isolates	No. of <i>P. v.</i> ‡	No. of <i>P. c.</i> §
A	10	5	5	94	43	51
B	7	7	0	69	69	0
C	9	8	1	96	85	11
Town	26	20	6	259	197	62
D	7	7	0	77	77	0
E	7	3	4	72	30	42
F	8	8	0	74	74	0
Field	22	18	4	223	181	42
Total	48	38	10	482	378	104

Table 1 Sampling scheme and number of recovered isolates of *Rhizobium etli* associated with *Phaseolus vulgaris* and *P. coccineus* plants recovered from six plots in two localities (town and field) in San Miguel, Puebla, Mexico

*Number of *P. vulgaris* plants.

†Number of *P. coccineus* plants.

‡Number of *R. etli* isolates recovered from *P. vulgaris*.

§Number of *R. etli* isolates recovered from *P. coccineus*.

of the plots. To compute the G_{ST} values we used the program *ETDIV* version 2.2 (Whittam 1990). These indices range from 0, if there is no genetic differentiation at a given level, to 1 if there is maximal genetic differentiation (Nei 1987). To test if the G_{ST} indices were significantly different from 0, we performed a chi-square test of independence as $\chi^2 = nG_{ST}(a - 1)$, where n is the number of individuals and a is the total number of alleles. Degrees of freedom are $(k - 1)(a - 1)$, where k is the number of subdivisions. Degrees of freedom and χ^2 values were summed across loci and significance was examined at $P < 0.05$ (Workman & Niswander 1970; Hagen & Hamrick 1996b).

The amount of gene flow among populations was calculated by substituting Nei's G_{ST} for F_{ST} in Wright's island model of gene flow (Wright 1951), that describes the effective number of migrants per generation as $Nm = (1 - G_{ST}) / 2G_{ST}\alpha$, in which $\alpha = (n/(n - 1))^2$, where n is the number of populations (Crow & Aoki 1984).

Also, we calculated the average Nei's genetic identities (I) for all pairs of plots based on the allele frequencies for the ETs (Nei 1987).

We performed a variance component estimation using both isolates and ETs. In order to obtain estimates of the proportion of the variability explained at each sampling level we performed a three-level nested ANOVA (Weir & Cockerham 1984; Weir 1996) with the levels: plots, plants within plots, and isolates within plants. These analyses consider the isolates within plants as the error of the ANOVA.

Linkage disequilibrium

To determine the extent to which populations exhibit non-random associations of alleles between loci, we used a multilocus index based on the distribution of allelic mismatches between pairs of isolates over all loci. An allelic mismatch frequency distribution is calculated by comparing each isolate with every other isolate for a total of $n(n - 1)/2$ pairwise comparisons, where n is the number of isolates in the population. An equation of the variance of the distribution has been derived by Brown *et al.* (1980) and the procedures have been described in detail elsewhere (Souza *et al.* 1992; Leung *et al.* 1994). The ratio of the variance in mismatches observed in a population (V_O) to the expected variance in a corresponding population at linkage equilibrium (random association of alleles due to mixis, V_E), provides a measure of multilocus linkage disequilibrium. If there is no linkage disequilibrium $V_O/V_E = 1$. The significance of the difference between V_O and V_E was calculated using a Monte Carlo procedure with 1000 iterations, performed with the LDV program (Souza *et al.* 1992). If the ratio V_O/V_E is significantly higher than 1 (alleles are associated in a nonrandom way), we suspect a lack of recombination (clonal structure) or selection to a certain combination of alleles.

Cluster analysis

The genetic distance between each pair of different multilocus genotypes (ETs) was estimated as the proportion of loci with different alleles (mismatches). Clustering from a matrix of pairwise mismatched distances was performed by the UPGMA method (unweighted pair-group method with arithmetic mean) (Sokal & Michener 1958). The analysis was carried out using the program *ETCLUS* version 2.1 (Whittam 1990). For this analysis we included four reference strains, one of *R. meliloti* (strain 1021, kindly provided by Frans de Bruijn, Center for Microbial Ecology, Michigan State University) and three *R. etli* bv. *phaseoli* (strains CFN42, TAL182 and CIAT895, kindly provided by David Romero, Centro de Investigación sobre Fijación de Nitrógeno, Universidad Nacional Autónoma de México).

Results

Genetic diversity

Levels of genetic variation of *Rhizobium etli* bv. *phaseoli* in San Miguel were high (Table 2) and displayed a high degree of genotype dominance, meaning that among the ETs identified, some of them were represented by many isolates. Among the 482 isolates collected from six plots and 48 plants (Table 1), only 126 distinctive multilocus genotypes (ETs) were found (Table 2). The total mean genetic diversity (H_E) for the six plots was 0.531. All the loci examined were polymorphic, and the mean number of alleles per locus was 4.2.

The plots showed a genetic diversity H_E ranging from 0.327 to 0.549 (Table 2; plots F and E, respectively). We found that the mean genetic diversity for the town plots ($H_E = 0.523$) was higher than for the field plots ($H_E = 0.481$). In Table 2, it can be observed that the mean number of alleles and genetic diversity were higher for the plots in which *Phaseolus coccineus* plants were sampled, than for the plots with only *P. vulgaris*. These results suggest that the *P. coccineus* plants nodulate with a wider range of *R. etli* bv. *phaseoli* genotypes. This is also reflected in the *P. coccineus* ET/isolates ratio (0.48), which is twice the *P. vulgaris* ET/isolates ratio (0.24).

Of the 126 ETs found from the 485 isolates, 85 were represented by only one isolate, comprising 67.5% of the ETs and 17.6% of the isolates. Of the 41 ETs represented by more than one isolate, ETs 7, 9, 11, 13 and 27 (with 26, 42, 70, 34 and 59 isolates, respectively) were shared by all the plots, comprising approximately half (48%) of the total number of isolates. A chi-square test showed that the frequencies of these five widespread ETs varied significantly among the six plots ($\chi^2_{20} = 78.2$, $P < 0.005$), and for all the pairwise combinations between town and field plots (AD $\chi^2_4 = 19$, AE $\chi^2_4 = 22.2$, AF $\chi^2_4 = 17.2$, BD $\chi^2_4 = 23.2$, BE $\chi^2_4 = 28$, BF $\chi^2_4 = 21.7$, CD $\chi^2_4 = 19.3$, CE $\chi^2_4 = 24.6$, CF $\chi^2_4 = 19.7$;

Data set	No. of isolates	No. of ETs	Mean no. of alleles	H_E †	G_{ST} ‡	Nm §
A	94	46	3.3	0.521 (0.069)	0.068 (0.020) ^a	5.6
B	69	18	2.7	0.346 (0.068)	0.040 (0.014) ^a	8.8
C	96	33	3.4	0.473 (0.065)	0.135 (0.033) ^a	2.5
D	77	27	2.6	0.336 (0.083)	0.015 (0.005) ^a	24.1
E	72	33	3.5	0.549 (0.054)	0.066 (0.015) ^a	5.2
F	74	30	2.9	0.327 (0.077)	0.090 (0.035) ^a	3.9
Town	259	75	3.7	0.523 (0.066)	0.062*(0.015) ^b	3.4
Field	223	69	3.9	0.481 (0.060)	0.046*(0.019) ^b	4.6
<i>P. v.</i> **	378	91	3.7	0.501 (0.066)	0.045 (0.010) ^a	10.1
<i>P. c.</i>	104	50	4.0	0.548 (0.067)	0.173 (0.039) ^a	1.9
Total	482	126	4.2	0.531 (0.066)	0.072*(0.017) ^b	4.5

Table 2 Genetic diversity and genetic differentiation estimates at different hierarchical levels for the 10 loci surveyed in *Rhizobium etli* associated with *Phaseolus vulgaris* and *P. coccineus* plants in two localities (town and field) in San Miguel, Mexico

ET, electrophoretic type.

†Mean genetic diversity, values in parenthesis represent \pm standard error (SE).

‡Genetic differentiation, values in parenthesis represent \pm SE.

§Effective number of migrants per generation.

***P. vulgaris*.

††*P. coccineus*.

* G_{ST} values significantly different from 0 at $P < 0.05$.

^aGenetic differentiation at the plant level.

^bGenetic differentiation at the plot level.

$P < 0.005$). In contrast, the differences in the frequencies of these widespread ETs among the town plots and among the field plots were not significant. This suggests that the differences are due to the local characteristics of the plots, but are averaged (and lost) when considering the localities.

Genetic differentiation

The genetic differentiation indices (G_{ST}) are shown in Table 2. The genetic differentiation of the six plots in relation to the total genetic diversity found in San Miguel was significantly different from 0 ($G_{ST} = 0.072$; Table 2), indicating that 7.2% of the variability in allele frequencies is attributable to differences among plots. The differences between the two localities accounted for a smaller, but a significant, fraction of the observed variability ($G_{ST} = 0.022 \pm 0.007$ standard error (SE), $P < 0.001$; not shown in Table 2). We also found a significant degree of differentiation among town ($G_{ST} = 0.062$) and field ($G_{ST} = 0.046$) plots (Table 2).

At the plant level, none of the mean genetic differentiation values were significant; G_{ST} ranged from 0.015 to 0.135 (plots D and C, respectively; Table 2), indicating considerable variation. When the plants of the two bean species were analysed separately, we found that the bacteria associated with *P. coccineus* plants showed a higher, but nonsignificant, degree of differentiation than the *P. vulgaris* plants (Table 2).

From the G_{ST} related values, the migration parameter Nm was estimated. All the Nm values at the plot level were relatively high and similar (Table 2). We also found that the number of effective migrants per generation

between town and field localities was high ($Nm = 5.1$; not shown in Table 2).

In contrast, the migration parameter at the plant level varied considerably (Table 2). The lowest value was found for the plants from plot C ($Nm = 2.5$), while the highest value was found for the plants from plot D ($Nm = 24.1$). The migration parameter for the total *P. vulgaris* sample was more than fivefold higher than the parameter for the total *P. coccineus* sample (Table 2), suggesting a lower degree of migration among *P. coccineus* plants compared with *P. vulgaris* plants.

The pairwise genetic identities (I) and percentage of shared ETs are shown in Table 3. Despite heterogeneity in allele frequencies among the six plots, the pairwise genetic identity values were very high (mean $I = 0.914$), with the exception of most comparisons with plot A. This supports our earlier analysis indicating a low degree of differentiation among plots. Moreover, because the identity values were not higher between pairs of plots of the same locality than the pairs among localities (Table 3), we found no evidence of isolation by distance between the two localities. These findings are reinforced by the results of the percentage of shared ETs (Table 3).

The three-level nested ANOVA showed that, in the case of isolates, 70.1% of the variability was due to differences among isolates within plants, while less variability can be explained by differences at higher levels: 13.1% is due to differences among plants and a slightly higher proportion (16.8%) to differences among plots. Similar results were obtained when the analysis was performed using only ETs.

Plot	A	B	C	D	E	F	Average†
A		0.811	0.918	0.804	0.891	0.791	0.843
B	10.9		0.955	0.984	0.911	0.978	0.928
C	13.9	21.6		0.934	0.954	0.936	0.939
D	8.2	20.0	13.3		0.922	0.990	0.927
E	7.6	13.7	10.6	13.3		0.929	0.921
F	7.9	16.7	15.9	19.3	12.7		0.925
Average	9.7	16.6	15.1	14.8	11.6	14.5	

ET, electrophoretic type.

*Above the diagonal are the genetic identities and below the diagonal are the percentages of shared ETs.

†Average over all the comparisons for each plot.

The genetic component of the variance at the isolates level was $\approx 88.3\%$, the variance due to differences among plants was 5.3% and, finally, the variance due to differences among plots was 6.4% . These results indicate that the main proportion of the variability is found within the plants level, rather than at higher local geographical scales.

Linkage disequilibrium at hierarchical levels

Table 4 summarizes the results of the linkage disequilibrium analysis. In general, the V_O/V_E ratios were lower when only the ETs were analysed, while the mean number

Table 4 Linkage disequilibrium parameters at different hierarchical levels at the 10 loci surveyed for *Rhizobium etli* associated with *Phaseolus vulgaris* and *P. coccineus* in two localities (town and field) in San Miguel, Mexico

Data set	Isolates		ETs		
	mism*	V_O/V_E †	mism	V_O/V_E	P‡
A	4.7	3.97	5.2	2.25	<0.001
B	1.9	3.13	3.5	2.71	<0.001
C	3.2	4.58	4.7	3.23	<0.001
D	2.3	1.76	3.4	1.46	<0.001
E	4.1	4.23	5.6	3.25	<0.001
F	2.5	1.71	3.3	1.76	<0.001
Town	4.2	4.60	5.2	2.44	<0.001
Field	3.1	3.16	4.8	2.51	<0.001
<i>P. v.</i> §	3.1	3.36	5.0	2.30	<0.001
<i>P. c.</i> ¶	5.0	4.10	5.5	2.91	<0.001
Total	3.8	3.93	5.4	2.27	<0.001

ET, electrophoretic type.

*Mean number of mismatches.

†Observed variance/expected variance of the mismatch distribution.

‡Probability of rejecting by chance alone the null hypothesis that $V_O = V_E$ for both isolates and ETs.

§*P. vulgaris*.

¶*P. coccineus*.

Table 3 Nei's genetic identities (I) and the percentage of shared ETs between pairs of plots*

of mismatches increased. The V_O/V_E ratio for the 482 isolates was 3.93, while it dropped to 2.27 if performed for the 126 ETs; both values are high and significantly different from 1 ($P < 0.001$; Table 4), indicating nonrandom association of alleles.

The V_O/V_E ratio for isolates within plots ranged from 1.71 to 4.58 (Table 4; plots F and C, respectively), and when only ETs were taken into account, the value ranged from 1.46 to 3.23 (Table 4; plots D and C, respectively).

The isolates' V_O/V_E ratio was higher for the town plots than for the field plots, and higher for *P. coccineus* than for *P. vulgaris* (Table 4). Plots in which both bean species were sampled had higher V_O/V_E ratios and mean number of mismatches than plots in which only *P. vulgaris* was sampled (Table 4); suggesting again that the *P. coccineus* plants nodulate with a wider range of *R. etli* bv. *phaseoli* genotypes than *P. vulgaris*.

In all cases, the V_O/V_E ratio was significantly larger than 1, indicating significant linkage disequilibrium at all the analysed hierarchical levels. These results show that *R. etli* bv. *phaseoli* has a clonal population structure at the different spatial and host levels.

Cluster analysis

The genetic relatedness of the 126 ETs are depicted in a UPGMA dendrogram (Fig. 2). The first division of the dendrogram is at 0.9 genetic distance (measured as the proportion of mismatches) separating the *R. etli* bv. *phaseoli* isolates from the *R. meliloti* strain. The *R. etli* bv. *phaseoli* cluster can be divided at a genetic distance of 0.7 in three clusters (Fig. 2). Cluster I contains 95 isolates, cluster II contains only the *R. etli* bv. *phaseoli* reference strains, while cluster III contains 387 isolates.

The compositions of clusters I and III are explained in Table 5. Cluster I is composed of 19.7% of the total isolates, mainly from the town plots and the *P. coccineus* plants. Of the *P. coccineus* isolates 53.9% were within this cluster, while only 10.3% of *P. vulgaris* isolates were represented

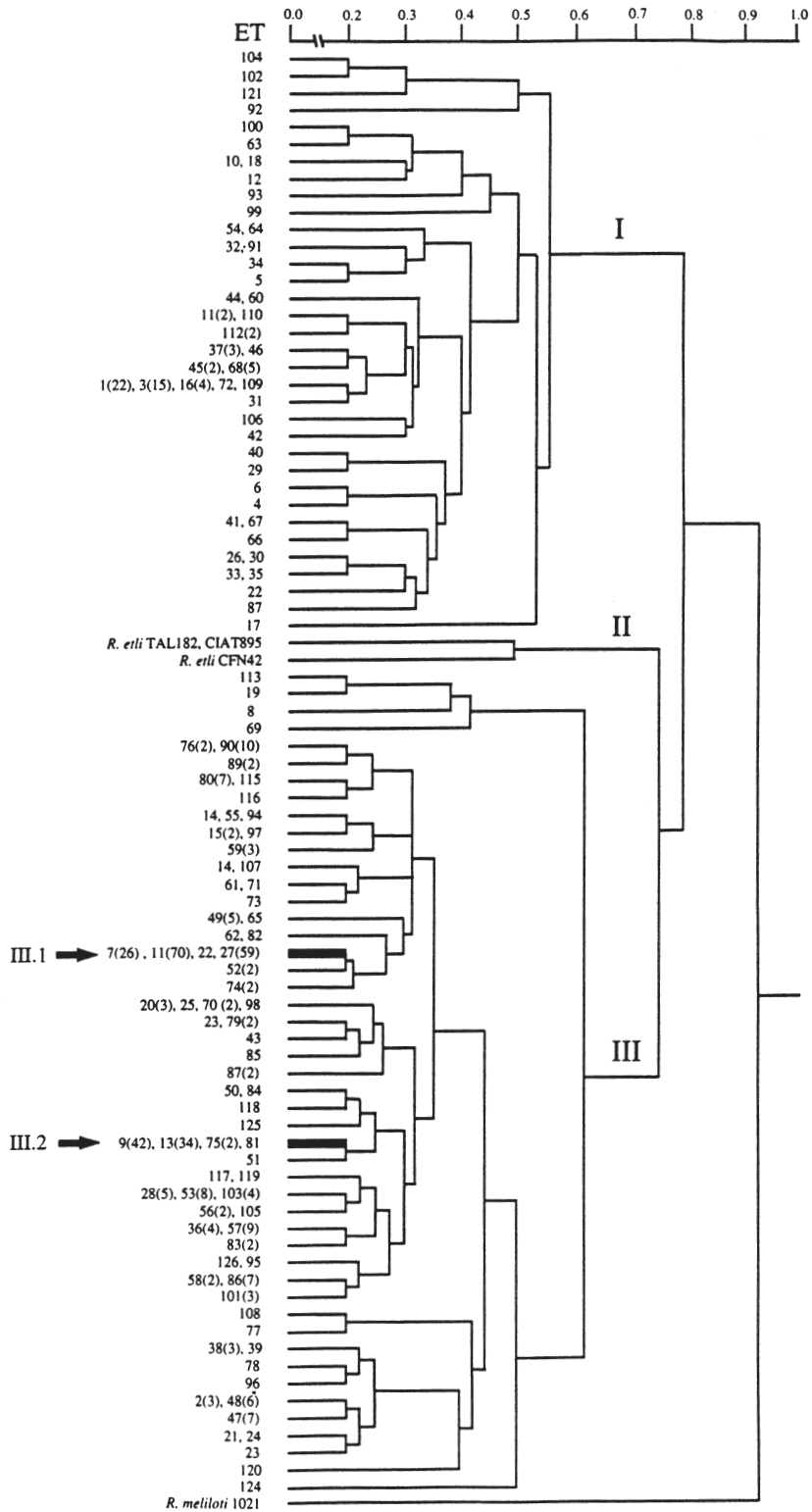


Fig. 2 Genetic relatedness among 126 electrophoretic types (ETs) of *Rhizobium etli* bv. *phaseoli* based on 10 polymorphic loci, generated by distance method UPGMA. The genetic distances were estimated as the proportion of allelic mismatches. Only genetic distances above 0.2 are shown. The terminal codes represent the electrophoretic type (ET) number and in parentheses the number of isolates for multiple ETs. The *R. etli* bv. *phaseoli* clusters I, II and III are indicated. The bold lines and black arrows refer to subclusters III.1 and III.2.

herein. More than half of the isolates in this cluster belong to plot A (62.1%; Table 5). Even though this cluster is not the largest one, it contains a high proportion of unique ETs (of the 48 ETs in this cluster, 40 were unique),

representing 83.3% of the cluster ETs and 47% of the unique ETs in the total sample. This clustering reflects the finding that the genotypes from plot A and from *P. coccineus* plants are the most distinct populations.

Table 5 Distribution of *Rhizobium etli* isolates belonging to clusters I and III, as defined in Fig. 2

Data set	Cluster I			Cluster III		
	<i>n</i> *	% pop.†	% clu‡	<i>n</i>	% pop.	% clu
A	59	62.8	62.1	35	37.2	9.0
B	3	4.3	3.2	66	95.7	17.1
C	15	15.6	15.8	81	84.4	20.9
D	2	2.6	2.1	76	98.7	19.6
E	15	20.8	15.8	56	77.8	14.5
F	1	1.4	1.1	73	98.6	18.9
Town	77	29.7	81.1	182	70.3	47.0
Field	18	8.1	18.9	205	91.9	53.0
<i>P. v.</i> §	39	10.3	41.0	339	89.7	87.6
<i>P. c.</i> ¶	56	53.8	59.0	48	46.2	12.4
Total	95	19.7	100.0	387	80.3	100.0

*Number of isolates in the corresponding population.

†Percentage of the corresponding population.

‡Percentage of the cluster.

§*Phaseolus vulgaris*.

¶*P. coccineus*.

Cluster III comprises the largest part of both the total isolates (80.3%) and the total ETs (61.9%). Almost all the field locality isolates and the *P. vulgaris* isolates are included here (Table 5). This cluster had a more homogeneous plot distribution than cluster I, because most of the isolates of each plot were found herein, with the exception of plot A (Table 5). Of the 78 ETs in this cluster, only 45 were unique, representing approximately half (57.7%) of the ETs in the cluster. The five common and widespread ETs were found herein, grouped in two smaller subclusters, named subclusters III.1 and III.2 (Fig. 2).

Subcluster III.1 contains the widespread ETs 11, 7 and 27 (with 70, 26 and 59 isolates, respectively), and the unique ET 22 (Fig. 2), comprising 32.4% of the total isolates. Subcluster III.2 contains the widespread ETs 13

and 9 (with 34 and 42 isolates, respectively), ET 75 with two isolates, and the unique ET 81 (Fig. 2), comprising 16.4% of the total population. The genetic distance among the subclusters was small. These results suggest that these abundant and widespread genotypes are closely related, and could be the result of the diversification of an original successful genotype.

We performed genetic diversity and linkage disequilibrium analyses for both the genotypes and the isolates within clusters I and III (Table 6). Cluster I showed a higher genetic diversity, mean number of mismatches, and V_O/V_E ratios than cluster III, both for ETs and isolates (Table 6). The isolates' V_O/V_E ratios for both clusters showed significant linkage disequilibrium, but the ETs' V_O/V_E ratios were not significantly different from 1 (Table 6), showing random allele association between loci (Maynard Smith *et al.* 1993). These results indicate that frequent mixis occurs among the genotypes within the two main genetic divisions.

Discussion

Genetic diversity and differentiation

Bacteria that nodulate beans in San Miguel are very diverse, globally and within each plot. The level of genetic variation found for the total population ($H = 0.531$) is similar to that reported for other *Rhizobium* species (e.g. $H = 0.493$ for *R. leguminosarum* bv. *viciae* (Gordon *et al.* 1995), $H = 0.426$ for *R. leguminosarum* bv. *trifolii* (Hagen & Hamrick 1996a)) and for other *R. etli* bv. *phaseoli* populations associated with cultivated *Phaseolus vulgaris* ($H = 0.499$ in 1987 and $H = 0.407$ in 1988 (Souza *et al.* 1994), $H = 0.513$ (Segovia *et al.* 1991)).

Sullivan *et al.* (1995) showed that genetic diversity in *Rhizobium* can be generated by recombination events between local bacteria and an inoculant strain. They proposed that these new nodulating strains may account for the rapid diversification of rhizobia in the field. In the

Table 6 Genetic diversity and multilocus linkage disequilibrium estimates for *Rhizobium etli* within clusters I and II, as defined in Fig. 2

Cluster	No. of isolates	No. of ETs	Mean no. of alleles	H^*	Isolates			ETs		
					mism†	V_O/V_E ‡	P §	mism	V_O/V_E	P
I	95	48	3.4	0.400 (0.082)	2.8	1.87	<0.001	4.1	1.15	N. S.
III	387	78	3.6	0.347 (0.067)	2.0	1.18	<0.001	3.4	0.75	N. S.
Total	482	126	4.2	0.531 (0.066)	3.8	3.93	<0.001	5.4	2.27	<0.001

ET, electrophoretic type.

*Mean genetic diversity, values in parenthesis represent \pm standard error (SE).

†Mean number of mismatches.

‡Observed variance/expected variance of the mismatch distribution.

§Probability of rejecting by chance alone the null hypothesis that $V_O = V_E$.

case of *R. etli* populations, Segovia *et al.* (1991) found a proportion of at least one symbiotic strain to 40 nonsymbiotic strains in the soil, and suggest the participation of these nonsymbiotic populations in the generation of new symbiotic strains with different adaptive traits. In San Miguel, we found a high proportion of genotypes represented by only one isolate (67.3% of the ETs or 17.6% of the isolates), which could be persisting rare genotypes or the outcome of mutation and/or recombination events within the *R. etli* genetic pool. Nevertheless, these unique genotypes may be outcompeted by other genotypes preferentially selected by the host plants. Among the preferred strains are the five widespread and genetically related genotypes that account for almost half of the isolates. These few successful genotypes may be adapted to the global environmental conditions of the site to survive during the drought season, and be selected by a wide range of the bean landrace genotypes. However, competitive differences among genotypes under specific ecological conditions could account for their differential abundance between the two localities, probably due to differences in soil composition. For example, ET 1 with 22 isolates, was the third most abundant among town plots, but this ET was not represented in the field plots. It is interesting that this ET belongs to cluster I, as do most of the genotypes from town plots. We found a slightly higher average genetic diversity in town plots than in field plots, perhaps indicating that the town environmental conditions allow the development of a wider range of *R. etli* *bv. phaseoli* genotypes. A high content of soil organic matter has been correlated with the maintenance of higher rhizobial populations, and maybe higher genetic diversity (E. Martínez-Romero, personal communication). The association between isolate distribution and plot of origin suggests an ecotypic structure, with clones adapted to the local conditions (Maynard Smith 1991; Haubold & Rainey 1996).

The G_{ST} related values were significant at the higher hierarchical levels (total and locality), although not significantly different from 0 within each plot or at the bean species level. These results indicate that a high proportion of the genetic diversity is found within individual plants. On the basis of the high ET/isolates ratio, mean genetic diversity, G_{ST} value, mean number of mismatches, and V_O/V_E ratio, we proposed that the *P. coccineus* plants nodulate with a wider range of *R. etli* *bv. phaseoli* genotypes than the *P. vulgaris* plants. The plot pairwise genetic identity values were very high, indicating that most alleles are shared among the different plots. Identity values and the percentage of shared ETs indicate that plot A is the most differentiated population. An explanation for the different genotypes found in plot A could be the differences in soil conditions, as a steeper inclination of this terrain can result in higher erosion and lower water retention.

The variance component estimation over the total genetic diversity showed that the main part of the differences was found among the isolates that nodulate the plants, whereas only a small fraction of the variability was explained by differences among plots or plants. This shows that the founder effect of the isolates that nodulate a plant and the genotype-specific selection of the host are important processes that are shaping the local population structure of this symbiotic nitrogen-fixing bacteria. This finding is in agreement with the results observed for *R. trifolii* by Hagen & Hamrick (1996a). They suggest that the gene flow acts at a larger scale, while the founder effect is important at local scales. Our data support a high degree of gene flow among plots and localities, but also among most plants. This may reflect the soil movement due to tillage practices, in contrast with the noncultivated roadside *R. trifolii* populations analysed by Hagen & Hamrick (1996a). All Nm values obtained for the sampled populations were higher than 1, and most were higher than 4, suggesting that the sampled plots may experience sufficient gene flow to prevent substantial genetic divergence and may belong to a single evolutionary unit (Wright 1951).

The total *R. etli* *bv. phaseoli* sampled population is in linkage disequilibrium, indicating clonality. But, as pointed out by Maynard Smith *et al.* (1993), linkage disequilibrium may arise by geographical or ecological isolation. In the present study, the linkage disequilibrium analysis of the ETs within the two main genetic clusters showed evidence of frequent mixis among genetically related genotypes. Because genotypes of both clusters were present in all the sampled plots, these results could explain why significant linkage disequilibrium was found at all the hierarchical levels analysed. This, together with the high gene flow detected in the population, suggests that geographical separation may not be an important source of linkage disequilibrium, reinforcing the observation that the lack of recombination found in the sampled populations of *R. etli* *bv. phaseoli* in San Miguel was due to the coexistence of two main genetic lineages sexually isolated. This suggests a reticulate genetic structure (Maynard Smith *et al.* 1993). On the other hand, the fact that the clusters linkage disequilibrium analysis performed with isolates showed clonality, and when performed only with ETs showed mixis, suggests an epidemic genetic structure (Maynard Smith *et al.* 1993). These two features brought together point towards a reticulated-epidemic population genetic structure (Fig. 3) for *R. etli* *bv. phaseoli* in San Miguel.

Evidence of a reticulate genetic structure has been reported for *R. meliloti* and *R. tropici* (Eardly *et al.* 1990; Martínez-Romero *et al.* 1991). As pointed out by Gordon *et al.* (1995) and Lenski (1993), conclusions derived from the analysis of subgroups will be much stronger when

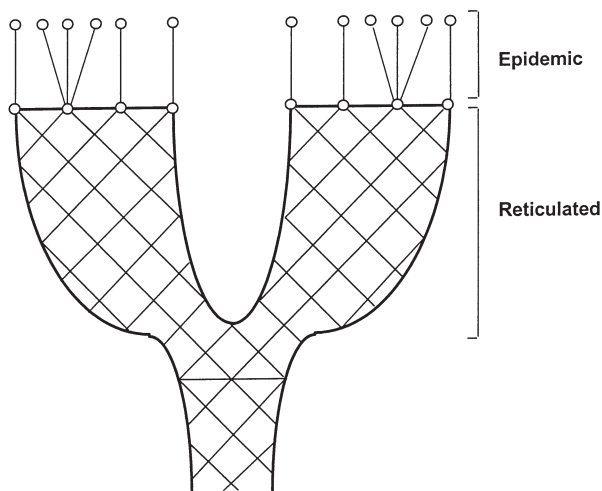


Fig. 3 Representation of the reticulated-epidemic genetic structure proposed for *Rhizobium etli*. The structure is basically reticulated, where recombination occurs between the members of the two main lineages, but is restricted among them. Occasionally a successful genotype increases in frequency to produce epidemic clones.

there is independent evidence that the subgroups are biologically meaningful entities. In the case of our *R. etli* *bv. phaseoli* isolates, it would be interesting to have additional information that would support the existence of these lineages and establish if these genetic divisions are part of the *R. etli* genetic pool or if they are different species.

This study strongly suggests that in San Miguel the local conditions have shaped a complex genetic structure. Within the *R. etli* genetic pool, mutation and recombination within the two lineages can generate diversity. On the one hand, we found ecotypes locally adapted, and on the other, we observed widespread genotypes adapted to the global conditions. The selection imposed by the bean landraces coupled with the founder effect of the different ecotypes produces an epidemic structure in the nodules, while gene flow prevents genetic divergence within the lineages and maintains the populations as a single evolutionary unit.

Acknowledgements

We thank the people of San Miguel, especially the Aguilar family, for their kindness and help in the field work. We thank Jordan Golubov, Valérie Bouchet, Jenny Bain, Antonio Cruz, Carlos Llorens and Gerardo Alvarado for field and laboratory assistance. We especially thank Aldo Valera for technical assistance and discussion. To Frans de Bruijn and David Romero for the reference strains. We also thank T. Whittam for the ETDIV and ETCLUS programs. We are grateful to Lorenzo Segovia, Esperanza Martínez, David Romero, Guillermo Dávila, Daniel Piñero, Valérie Bouchet, Santiago Elena and an anonymous reviewer for the critical comments of the manuscript. This research was supported by a MacArthur fellowship to V.S., and C.S. received a studentship from Consejo Nacional de Ciencia y Tecnología (México).

References

- Alcorn JB (1990) Indigenous agroforestry systems in the Latin American tropics. In: (Altieri MA, Hecht SB eds) *Agroecology and Small Farming Development*, pp. 203–220. CRC Press, Boston.
- Altieri MA, Merrick LC (1987) In situ conservation of crop genetic resources through maintenance of traditional farming systems. *Economic Botany*, **41**, 86–96.
- Bottomley PJ, Cheng HH, Strain SR (1994) Genetic structure and symbiotic characteristics of a *Bradyrhizobium* population recovered from a pasture soil. *Applied and Environmental Microbiology*, **60**, 1754–1761.
- Brockwell J, Bottomley P, Thies JE (1995) Manipulation of rhizobia microflora for improving legume productivity and soil fertility: a critical assessment. *Plant and Soil*, **174**, 143–180.
- Brown ADH, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics*, **96**, 523–536.
- Bush S (1986) Genetic diversity and conservation in traditional farming systems. *Journal of Ethnobiology*, **6**, 151–167.
- Crow JF, Aoki K (1984) Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proceedings of the National Academy of Sciences of the USA*, **81**, 6073–6077.
- Demezas DH, Reardon TB, Strain SR, Watson JM, Gibson AH (1995) Diversity and genetic structure of a natural population of *Rhizobium leguminosarum* *bv. trifolii* isolated from *Trifolium subterraneum* L. *Molecular Ecology*, **4**, 209–220.
- 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.
- Geniaux E, Laguerre G, Amarger N (1993) Comparison of geographically distant populations of *Rhizobium* isolated from root nodules of *Phaseolus vulgaris*. *Molecular Ecology*, **2**, 295–302.
- Gliessman SR (1986) Plant interactions in multiple cropping systems. In: (Francis CA ed.) *Multiple Cropping Systems*, pp. 82–95. Macmillan, New York.
- Gordon DM, Wexler M, Reardon TB, Murphy PJ (1995) The genetic structure of *Rhizobium* populations. *Soil Biology and Biochemistry*, **27**, 491–499.
- Hagen MJ, Hamrick JL (1996a) A hierarchical analysis of the population genetic structure in *Rhizobium leguminosarum* *bv. trifolii*. *Molecular Ecology*, **5**, 177–186.
- Hagen MJ, Hamrick JL (1996b) Population level processes in *Rhizobium leguminosarum* *bv. trifolii*: the role of founder effects. *Molecular Ecology*, **5**, 707–714.
- Haubold B, Rainey PB (1996) Genetic and ecotypic structure of a fluorescent *Pseudomonas* population. *Molecular Ecology*, **5**, 747–761.
- Hebert PDN, Beaton MJ (1993) *Methodologies for Allozyme Analysis using Cellulose Acetate Electrophoresis*. Helena Laboratories, Beaumont, TX.
- Istock CA, Duncan KE, Freguson N, Zhou X (1992) Sexuality in a natural population of bacteria – *Bacillus subtilis* challenges the clonal paradigm. *Molecular Ecology*, **1**, 95–103.
- Laguerre G, Fernandez MP, Edel V, Normand P, Amarger N (1993) Genomic heterogeneity among French *Rhizobium* strains isolated from *Phaseolus vulgaris* L. *International Journal of Systematic Bacteriology*, **43**, 761–767.
- Lenski RE (1993) Assessing the genetic structure of microbial populations. *Proceedings of the National Academy of Sciences of the USA*, **90**, 4334–4336.

- Leung K, Strain SR, De Bruijn F, Bottomley PJ (1994) Genotypic and phenotypic comparisons of chromosomal types within an indigenous soil population of *Rhizobium leguminosarum* bv. *trifolii*. *Applied and Environmental Microbiology*, **60**, 416–426.
- 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.
- Maynard Smith J (1991) The population genetics of bacteria. *Proceedings of the Royal Society of London, Series B*, **245**, 37–41.
- 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.
- 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.
- Segovia L, Piñero D, Palacios R, Martínez E (1991) Genetic structure of soil populations 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.
- Sokal RR, Michener CD (1958) A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin*, **28**, 1409–1438.
- Souza V, Eguiarte L, Avila G *et al.* (1994) Genetic structure of *Rhizobium etli* biovar *phaseoli* associated with wild and cultivated bean plants (*Phaseolus vulgaris* and *Phaseolus coccineus*) in Morelos, Mexico. *Applied and Environmental Microbiology*, **60**, 1260–1268.
- Souza V, Nguyen TT, Hudson RR, Piñero D, Lenski RE (1992) Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for localized sex? *Proceedings of the National Academy of Sciences of the USA*, **89**, 8389–8393.
- Souza V, Silva C, Bain J *et al.* (in press) do agricultural practices modify the population genetic structure of the nitrogen fixing bacteria *Rhizobium etli* biovar *phaseoli*? *Journal of Ethnobiology*, in press.
- Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW (1995) Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proceedings of the National Academy of Sciences of the USA*, **92**, 8985–8989.
- Weir BS (1996) Genetic data analysis. *Methods for Discrete Population Genetic Data*, pp. 184–186. Sinauer Associates, Sunderland, MA.
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Whittam TS (1990) ETDIV and ETCLUS programs. Pennsylvania State University.
- Wise MG, Shimkets LJ, McArthur JV (1995) Genetic structure of a lotic population of *Burkholderia (Pseudomonas) cepacia*. *Applied and Environmental Microbiology*, **61**, 1791–1798.
- Workman PL, Niswander JD (1970) Population studies on southwestern Indian tribes. II. Local genetic differentiation in the Papago. *American Journal of Human Genetics*, **22**, 24–49.
- Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

The research presented is part of the master degree dissertation of C. Silva at the Facultad de Ciencias, UNAM. These studies were completed under the direction of Dr V. Souza and Dr L. E. Eguiarte, who are working at the Departamento de Ecología Evolutiva of the Instituto de Ecología (UNAM) on the evolution of bacterial populations of agricultural and medical importance, as *Rhizobium etli* and *Escherichia coli*.
