Microb Ecol (1996) 32:123–147

MICROBIAL ECOLOGY

Use of Cluster and Discriminant Analyses to Compare Rhizosphere Bacterial Communities Following Biological Perturbation

G.S. Gilbert,[1] M.K. Clayton,[1,2] J. Handelsman,[1] J.L. Parke[1]

[1] Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI 53706 USA
[2] Department of Statistics, University of Wisconsin-Madison, Madison, WI 53706 USA

Received: 19 May 1995; Revised: 7 December 1995

Abstract. We present an approach to comparing the diversity and composition of bacterial communities from different habitats and for identifying which members of a community are most affected by an introduced bacterium. We use this method to explore both previously published and new data from field and growth chamber experiments in which we isolated heterotrophic bacteria from samples of root-free soil, roots of nontreated soybean seedlings, and from the roots of soybean seedlings grown from *Bacillus cereus* UW85n1–treated seeds. We characterize bacterial isolates for 40 physiological attributes, and grouped the isolates hierarchically using two-stage density-linkage cluster analysis. Multivariate analysis of variance and discriminant analysis of the relative frequencies of the clusters in the soil and rhizosphere habitats were then used to determine whether there were differences among the bacterial communities from the various habitats, and which of the clusters were most useful in discriminating among the communities. We used rarefied estimates of richness as a measure of community diversity in the various habitats. Introduction of UW85n1 affected the composition and/or diversity of rhizosphere communities in three of four experiments.

Introduction

In recent years there have been many attempts to manipulate microbial communities in the soil by introducing bacteria and fungi with properties useful for biological control of plant diseases [8, 21, 26, 27]. It is widely thought that biological control should have fewer undesirable nontarget effects than traditional chemical control systems. Advances in methods for tracking individual strains of microbes have led to a large literature on the fate of the introduced organisms themselves or on their

Present address: Department of Environmental Science, Policy and Management, 108 Hilgard Hall, University of California, Berkeley, CA 94720
Correspondence to: G.S. Gilbert.

target pathogens [2, 7, 11, 13, 16, 18, 20, 24], but there have been few attempts to determine the effects of an introduced microbe on other members of microbial communities [12, 19, 25, 29].

In a recent paper, we showed that coating soybean seeds with the biological control agent *Bacillus cereus* UW85 could significantly affect the bacterial communities that later developed on the roots, even when the introduced strain itself did not persist as a common member of the community [4]. For our descriptions of the rhizosphere and soil communities we used the proportion of isolates that expressed various (43–50) physiological traits, and then compared the communities with discriminant analysis. Using this method we were able to illustrate the impact of UW85 on the rhizosphere community, without assigning bacterial isolates to taxonomic units such as genus or species [4]. Although this is a powerful technique for identifying community-level effects, it does not permit us to assess effects on the structure of the rhizosphere communities or to determine which types of bacteria within the community warrant further study. Identifying which community members are most affected by the introduced organism is essential for evaluating the ecological importance of observed changes in the microbial community.

Although assigning taxonomic names to each bacterial isolate would provide information that could be analyzed for community diversity and the presence of indicator species, the nature of bacterial taxonomy poses several problems. First, bacterial taxonomy is still largely based on physiological characteristics of the bacteria, which results in some species representing a phylogenetically and ecologically discrete group of bacteria, whereas other species are defined much more broadly. Also, strains within a given species may have widely different ecological strategies [6, 17]. Finally, it is not clear what hierarchical level (i.e., family, genus, species, race) is appropriate for studies of bacterial communities; communities from different habitats may appear quite different from one another at one taxonomic level, but be indistinguishable at another level [1]. Consequently, in the present analysis we distinguished bacterial types based solely on physiological characteristics, without making taxonomic determinations.

In this paper we present a framework for the analysis of the structure and composition of bacterial communities and for identifying which types of bacteria, or combinations of types, are most useful as indicators of differences among rhizosphere communities. We use cluster analysis to group bacteria into physiological types or clusters, and then use discriminant analysis to determine which types of bacteria are useful indicators of different communities. By performing the analyses with the bacteria grouped into various numbers of clusters, we address how community analysis at various hierarchical levels (e.g., grouped into different numbers of clusters, varying in the degree of physiological similarity within clusters) can affect how we interpret the structure or composition of the communities. In our consideration of bacterial communities we follow a conceptual structure analogous to the hierarchical organization of parasite communities described by Esch et al. [3]. In their terminology, the "infracommunity" represents the assemblage of all populations of bacteria on a single host individual, whereas the "component community" represents all the infracommunities associated with a host population. We employ two sampling designs, one aimed at each hierarchical community level, and compare how our interpretations about the differences among rhizosphere communities might differ if the analysis were based on only a small number of

bacterial isolates from each of many root samples from the plant population (the component community sampling protocol), or if it were based on a large number of isolates from each of a small number of individual soil or root samples (the infracommunity sampling protocol). As an illustration, we employ this framework to determine which members of the bacterial community in the soybean rhizosphere are affected by coating seeds with the biological control agent *B. cereus* UW85n1.

Materials and Methods

Experimental Protocol

We examined the heterotrophic bacterial communities in root-free soil, in the rhizosphere of nontreated soybean plants, and in the rhizosphere of plants grown from seeds coated with *Bacillus cereus* UW85n1 (a spontaneous neomycin-resistant mutant of *B. cereus* UW85 that retains the bicontrol activity against damping-off disease of its parent strain) in field experiments in Wisconsin at one site in 1989 (Arlington 1989) and two sites in 1990 (Arlington 1990 and Hancock 1990). In 1990, to compare the effects on rhizosphere communities of the introduction of bacteria that have or lack biological control activity, we included a second strain, *B. cereus* UW831, which is a mitomycin C–induced mutant derived from UW85n1, and which has lost the ability to protect alfalfa seedlings against *Phytophthora medicaginis* in a growth-chamber bioassay [23]. In 1989, to determine whether results from a growth chamber experiment would reflect results from the field, we conducted a growth chamber experiment (Growth Chamber 1989) with soil collected from the Arlington field site. Soil was air-dried and sieved (2-mm mesh) before use. Soil was placed in conetainers and planted with seeds from each of the treatments (one seed per conetainer), or left fallow as a soil control. In each of the experiments, seeds were coated with a fully sporulated culture of the bacteria (grown on solid media), and air-dried. Seeds for the three plant treatments, plus designated sites or conetainers for soil cores, were completely randomized in each experiment. Seedlings were harvested 2 days after most plants had emerged from the soil (4 days in the growth chamber, 8–10 days in the field). A 1-cm segment of root (or 1-cm^3 sample of soil) collected 2–3 cm below the soil line was suspended in water, and dilutions of the resulting suspension were placed on 1/10-strength Trypticase soy agar (1/10-strength TSA; 1/10 recommended strength, containing 1.5% agar). Further details of the experimental design and sampling methods can be found in Gilbert et al. [4].

Throughout, "sample" means an individual root segment or soil collection and the bacteria isolated from it. We will refer to samples from root-free soil as "soil samples," root segments from plants grown from nontreated seeds as "nontreated root samples," segments from plants grown from seeds coated with UW85n1 as "UW85n1-treated root samples," and those from seeds coated with UW831 as "UW831-treated root samples."

Isolation and Characterization of Bacteria

We isolated a total of 2,651 bacteria from 36–50 samples per treatment in each of four experiments (Table 1a). Five (in 1989) or 6 (in 1990) isolates were randomly selected from a 3-day-old plate of each sample, purified by restreaking, and transferred to 1/10-strength TSA. We refer to this isolation procedure as the "component-community" sampling protocol, and bacteria isolated using this procedure were those described previously. This method allows good description of the collection of bacterial infracommunities across a population of plants. There were no differences within experiments in the total density of bacteria on treated and nontreated roots [4]. Additionally, in the Arlington 1990 experiment, we transferred all of the colonies that grew on one dilution plate from each of three samples of each habitat (soil, nontreated roots, UW85n1-treated roots, UW831-treated roots) to 1/10-strength TSA (35–112 isolates per sample) (Table 1b). This procedure we call the "infracommunity" sampling protocol; it provides more detailed description of the communities on individual roots, but little information about variability across plants. Some isolates from both protocols did not survive

Table 1. Number of root or soil samples, and number of bacterial isolates collected from each habitat **(a)** in each of the four experiments using the component-community sampling protocol, or **(b)** in the Arlington 1990 experiment using the infracommunity sampling protocol

a.	Arlington 1989		Growth Chamber		Arlington 1990		Hancock 1990	
Habitat	Samples	Isolates	Samples	Isolates	Samples	Isolates	Samples	Isolates
Soil	45	173	50	232	40	232	38	210
Untreated root	50	189	49	217	36	134	40	181
UW85n1-trt root	50	182	46	184	40	158	40	186
UW831-trt root	–	–	–	–	40	161	42	212

b.	No. bacteria from sample #		
Habitat	1	2	3
Soil	112	35	36
Untreated root	81	70	104
UW85n1-trt root	74	78	95
UW831-trt root	70	67	101

purification or storage procedures (22.8% overall). We then characterized each isolate by 40 physiological attributes, including Gram determination, motility, production of a range of extracellular enzymes, resistance to antimicrobial compounds, and growth on single, simple carbon sources in minimal medium (Table 4). Details of the tests were presented previously [4].

Cluster Analysis

We combined physiological data from bacteria from all four experiments in which the component-community sampling protocol was used, and subjected them to cluster analysis to classify each isolate into a hierarchical classification system of physiologically similar types. At the lowest hierarchical level, isolates within a cluster are most physiologically homogeneous; at higher hierarchical levels, there is greater heterogeneity among the physiological attributes of isolates within a given cluster. We used the nonparametric, two-stage density-linkage method of Sarle [22]. In the first stage of this method, disjoint modal clusters are formed, based on density estimates of k nearest neighbors to each observation. Large values of k produce smooth density estimates and few disjoint modes; small values produce many modes (clusters). In the second stage, the disjoint modal clusters are joined hierarchically by the single-linkage clustering method into progressively smaller numbers of clusters. In single-linkage clustering, the distance between clusters is the minimum distance between an observation in one cluster and an observation in another cluster.

We selected the number of k nearest neighbors to use in the analysis by determining the number of cluster modes created for $k = 2, 3, 4, 5, 6, 7, 8, 9$, and 10. The number of modes were 347, 114, 68, 46, 34, 24, 19, 18, and 16, respectively. We chose $k = 4$ for subsequent analyses, which permitted the analysis of communities at several hierarchical levels, with the initial 68 clusters sequentially grouped into 60, 50, 40, 30, 20, and 10 clusters. At smaller values of k, most first-stage clusters were represented by single isolates. We made no attempt to determine whether the clusters corresponded to recognized taxa.

In order to compare how component- and infracommunity sampling approaches affect our descriptions of bacterial communities on roots and in soil, we combined the data from the two sampling strategies for Arlington 1990, and subjected them to two-stage density-linkage cluster analysis at $k = 4$. We then divided the isolates into 20 or 30 clusters, and compared the apparent diversity (richness) of the bacterial communities using the two sampling protocols, as described below.

We analyzed each experiment separately, using the relative frequencies of clusters within each sample (e.g., number of isolates of cluster n/total number of isolates in that sample) from the component-community sampling protocol. Analyses were repeated with bacteria grouped into 10, 20, 30, 40, 50,

or 60 clusters (hierarchical levels). For example, at the 50-cluster hierarchical level, a sample from which we collected 3 isolates of cluster 4, 2 of cluster 16, and 1 of cluster 32 would be analyzed with relative frequencies of 0.5, 0.33, and 0.17, respectively. If the same data were then analyzed at the 20-cluster hierarchical level, and if clusters 16 and 32 were sufficiently similar to be combined into one cluster at that level, relative frequencies would then be 0.5 and 0.5. We used multivariate analysis of variance (MANOVA) to determine whether there were significant differences among the communities from the various habitats [22] (1988). MANOVAs were performed including (1) all habitats within an experiment, (2) root habitats only (excluding soil), (3) nontreated and UW85n1-treated roots only (excluding UW831-treated roots in Arlington 1990 and Hancock 1990, and soil), and (4) nontreated roots and soil only. Those experiments for which there were significant differences among communities (Wilks' Lambda, $P \leq .05$), were analyzed further using discriminant analysis.

Discriminant Analysis

We used discriminant analysis [28] to determine which combination of bacterial clusters best defined root or soil samples as belonging to their predefined groups (soil, nontreated root, etc.). As for the MANOVAs, the relative frequencies of clusters in each root or soil sample were used in the analyses. We randomly assigned half the samples from each habitat into one of two equal groups. One group was designated "training" samples, and was used to create the discriminant function. The other group, the "testing" samples, was used later to test the discriminant model. Stepwise and predictive discriminant analyses [22, 28] were used to determine which combinations of bacterial clusters (e.g., their relative frequencies within samples) were most useful in differentiating among the bacterial communities from various habitats, and to determine similarities and differences among bacterial communities. Stepwise discriminant analysis was applied to the training samples to identify those clusters that contributed most to discriminating among the communities from the various habitats in each experiment ($P \leq .15$ for a cluster to enter or remain in the function). Linear discriminant functions (LDF) were then created with those clusters selected through the stepwise procedure, using only the training samples. The LDF was then used to calculate discriminant scores for the testing samples, and each sample was classified into the habitat from which its discriminant score has the smallest squared distance. The proportion of samples that were classified correctly into their habitat of origin provides a quantitative measure of the differences among communities, and was used to compare the communities at various hierarchical levels. The entire discriminant procedure was repeated 20 times, each time randomly dividing samples into testing and training sets.

Community Diversity

The richness of a community is the number of different kinds of organisms in it. Richness is the most direct measure of the diversity of a community, and is a fundamental measure of community structure. The number of kinds of organisms found in a sample is strongly dependent on the size of the sample, and since in this study the sample size varies among habitats, we used Hurlbert's [10] nonbiased method of rarefaction to estimate the number of different clusters we would expect in a standardized sample size of 100 isolates (for the component-community sampling protocol) or 30 isolates (for the infra- and component-community sampling protocol comparison of the Arlington 1990 experiment). This method provides an intuitively accessible index of diversity while correcting for differences in sample sizes and avoiding assumptions of underlying distributions, both problems associated with most other indices of diversity.

Results

Multivariate Analysis of Bacterial Communities

Multivariate analysis of the bacterial communities was a three-step process, beginning with grouping bacteria into clusters of similar physiological characteristics,

followed by multivariate analysis of variance to test whether communities were indeed different among habitats, and finally using discriminant analysis to determine which combinations of types of bacteria are most indicative or diagnostic of the various habitats.

Using two-stage density-linkage cluster analysis we first grouped all isolates from the component-community sampling protocol initially into 68 clusters, and subsequently into 60 to 10 hierarchically nested clusters. Figure 1 represents the hierarchical relationships among clusters at these various levels, and shows the number of isolates included within each of the clusters at the lowest (60 clusters) hierarchical level.

We analyzed the relative frequencies of these clusters within samples using MANOVA, and found significant differences (Wilks' Lambda, $P \leq .05$) among the bacterial communities from various habitats for all experiments; the differences were detected at each hierarchical level of bacterial grouping (Fig. 1) from 20–60 clusters. For the 1989 experiments, there were significant differences among communities from all three habitats (soil, nontreated roots, UW85n1-treated roots). However, in the 1990 experiments, although there were significant differences between soil and nontreated root communities ($P \leq .0001$), there were no significant differences between communities on nontreated and treated roots ($P \geq .10$). Because there were significant differences among all of the habitats in the two 1989 experiments, we then used discriminant analysis to further explore the differences among the three communities from those experiments, but not from the 1990 experiments.

In order to determine which bacterial clusters (and combinations of clusters) were diagnostic of different habitats, we performed discriminant analysis at each hierarchical level (number of clusters) at which there were significant differences among habitats by MANOVA (Wilks' Lambda, $P \leq 0.05$). Analyses were performed separately for the Arlington 1989 and Growth Chamber 1989 experiments (component-community sampling protocol data). The correct classification rates and the number of clusters included in the linear discriminant functions are presented in Table 2. A total correct classification rate of 33.3% for 3 groups would be expected if samples were randomly assigned to groups (or if there were no differences among the communities). When bacteria were grouped into 10 clusters (Fig. 1), there were either no significant differences among habitats or the discriminant functions produced very high misclassification rates. When bacteria were in groups 20–60 clusters, there was only a moderate amount of variability in the total misclassification rates within an experiment, but the misclassification rates for the various habitats varied among hierarchical levels. For instance, for the Arlington 1989 experiment, the linear discriminant function using 30 clusters did a good job of correctly classifying soil and UW85n1-treated root samples, but misclassified more than 60% of the nontreated root samples. Grouping bacteria into 40 clusters, the soil and treated-root samples were less often classified correctly, but the samples from nontreated roots were correctly classified at a higher level.

We can learn a great deal about differences among the communities through inspection of the actual classification rates for each of the experiments. We probed further into the classification rates for analyses at the hierarchical level of 40 clusters, which provided the best total classification rates for Growth Chamber 1989 and the second best for Arlington 1989. The 40-cluster level was also the

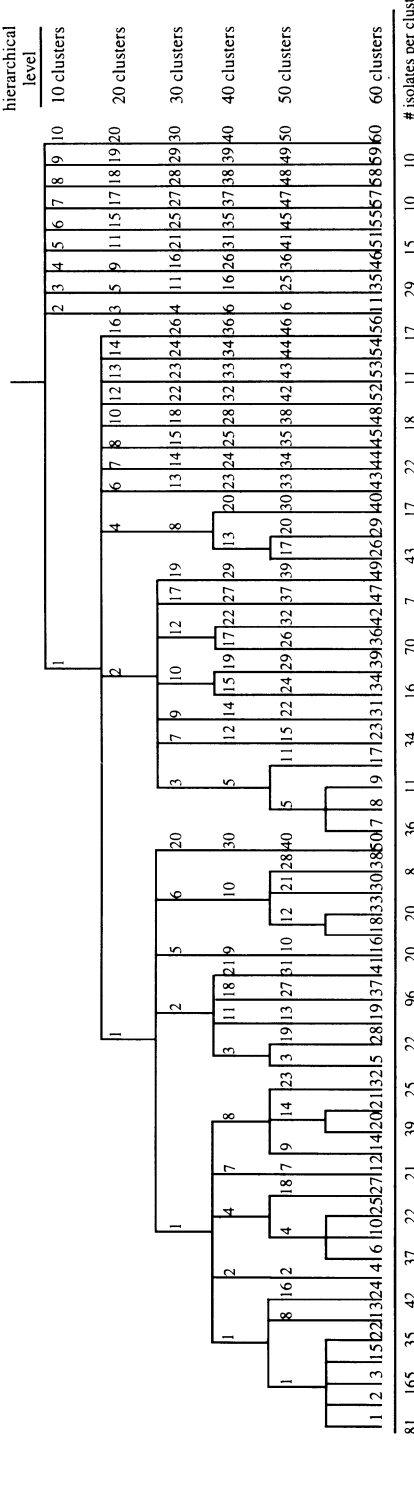

Fig. 1. Dendrogram of clusters created using two-stage density-linkage cluster analysis ($k = 4$), for all 4 experiments (component-community sampling protocol, up to 6 isolates from each of 40 soil or root samples). Each hierarchical level represents the relationships among clusters when all isolated are divided into the given number of clusters. Numbers at each node are cluster identification numbers.

Table 2. Percentage of testing samples correctly classified in discriminant analyses at each hierarchical level, for three habitats only (component-community sampling protocol)

Experiment	Hier. level[b]	% samples correctly classified[a]				No. of clusters in LDF[c]
		Soil	Nontreated root	UW85n1-treated root	Total	
Arlington 1989	10	90.9	4.0	56.0	50.3	1
	20	86.4	48.0	64.0	66.1	4
	30	72.7	36.0	80.0	62.9	6
	40	68.2	60.0	64.0	64.1	10
	50	87.3	64.0	44.0	61.8	17
	60	50.0	56.0	44.0	50.0	12
Growth Chamber 1989	10	–[d]	–	–	–	1
	20	60.0	95.8	8.7	54.8	4
	30	72.0	83.3	43.5	66.3	4
	40	92.0	70.8	65.2	76.6	8
	50	96.0	41.7	43.5	60.4	12
	60	60.0	70.8	26.1	52.3	15

[a] Percentage of testing samples correctly classified, using linear discriminant function from training samples

[b] Number of clusters into which isolates are grouped

[c] Number of clusters included in linear discriminant function, as determined by STEPDISC, $P \leq 0.15$

[d] Classification could not be performed because the cluster in the LDF was not present in two habitats

only level for which correct classification rates were ≥60% for all three treatments. Discriminant analyses for Arlington 1989 and Growth Chamber 1989 were performed 20 times, providing an average correct classification rate of 65.5% (SD = 4.3) for Arlington 1989 and 62.8% (SD = 6.3) for Growth Chamber 1989. Only the first repetition is described for each experiment (Tables 3 and 6, respectively); they are representative of the remaining repetitions.

By examining the coefficient vectors of the linear discriminant functions, and noting which clusters have coefficients that deviate furthest from zero, we can determine which clusters or combinations of clusters are particularly associated with one or another habitat. For example, in the Arlington 1989 experiment (Table 3a), the presence of clusters 3, 5, and 11 and the absence of cluster 12 strongly indicate that the sample was from root-free soil; clusters 1, 17, 29, 37, and 40 are indicative of nontreated root samples; and UW85n1-treated root samples are characterized by the presence of cluster 12 and by a lower abundance of clusters 29 and 40, which are characteristic of nontreated roots.

Examination of the physiological characteristics of these clusters indicates the kinds of compositional shifts there are among communities and can suggest the ecological changes with which they are associated. The percentage of isolates in each of the 40 clusters (isolates from both testing and training samples included) with positive responses to each of the 40 attribute tests is presented in Table 4, and the percentage of isolates from each cluster in each habitat is presented in Table 5. From these tables, we find that the clusters that best characterize soil samples are each predominated by members that produce proteolytic enzymes (GEL, CAS) and are resistant to only a few antimicrobial compounds, and that other

Table 3. Discriminant analysis of Arlington 1989 experiment

a. Linear discriminant function[a]

Cluster	Coefficient vectors[b]		
	Soil	Nontreated root	UW85n1-treated root
1	6.97	13.30	7.16
3	21.87	11.98	4.46
5	29.28	13.52	5.56
11	57.08	28.36	11.21
12	−12.21	−3.50	17.41
17	6.14	16.60	8.41
19	15.05	12.08	3.99
29	8.55	23.27	7.25
37	11.75	29.93	12.77
40	10.87	39.55	11.43
Constant	−5.61	−6.16	−1.48

b. Test-classification of samples

Habitat of origin[c]	% of samples classified into habitat			% error[d]	Number of samples
	Soil	Nontreated root	UW85n1-treated root		
Soil	68.2	0.0	31.8	31.8	22
Nontreated root	4.0	60.0	36.0	40.0	25
UW85n1-treated root	8.0	28.0	64.0	36.0	25

[a] Linear discriminant function was trained on half of the samples from each habitat

[b] Coefficient vector = $COV^{-1}\ \bar{x}_j$ and constant = $-0.5\ \bar{x}_j'\ COV^{-1}\ \bar{x}_j$, where COV = pooled covariance matrix and $\bar{x}_j$ = mean vector of sample scores for habitat j

[c] Test-classified samples were those not used in training the linear discriminant function

[d] Total error = 35.9%

attributes are quite variable among the three clusters. The clusters that characterize nontreated roots (except cluster 37) are each resistant to a range of antimicrobial compounds, can utilize a variety of simple carbon sources on minimal medium, and produce few extracellular enzymes. Those clusters that are common on nontreated roots but missing from UW85n1-treated roots are notably clusters with predominantly Gram-negative members, notable because of the well-known greater abundance of Gram-negative bacteria in the rhizosphere compared to the root-free soil [15]. Note also that *B. cereus* UW85n1 is Gram-positive and grows anaerobically, shows positive responses to BLO, GEL, CAS, LEC, LIP, STA, MOT, ZNC, NEO, AMP, and CVI, and does not grow on minimal medium (see Table 3 for explanation of test codes). None of the bacteria isolated in any of the experiments had this combination of attributes, suggesting that UW85n1 was not a common member of the rhizosphere community at the time of sampling.

We then used the LDF (created using the training samples only) to test-classify the testing samples into one of the three habitats (soil, untreated roots, UW85n1-treated roots). More than 68% of the soil samples were correctly classified (i.e., classified as being from the soil sample group) (Table 3b); all the remaining soil samples were misclassified as UW85n1-treated root samples, suggesting that soil

Table 4. Percentage of isolates with positive responses to attribute tests for each of 40 clusters, for component-community sampling protocol.[a] Clustering by two-stage density linkage method ($k = 4$)

Attribute test	Code	Percent of isolates in each cluster with positive response to attribute test							
		1	2	3	4	5	6	7	8
Gram positive	GRA	89.1	100.0	96.0	96.5	3.9	82.8	95.2	77.8
Blood haemolysis	BLO	6.2	100.0	11.8	96.5	22.5	17.2	0.0	19.0
Gelatin hydrolysis	GEL	46.7	100.0	92.8	97.8	87.6	13.8	85.7	24.6
Casein hydrolysis	CAS	5.8	100.0	83.8	95.2	84.5	10.3	0.0	6.0
Lecithinase activity	LEC	1.9	100.0	7.8	94.4	48.8	3.4	0.0	1.2
Lipase activity	LIP	0.9	5.3	3.1	67.5	74.4	3.4	0.0	6.0
Starch hydrolysis	STA	42.7	100.0	97.5	96.5	9.3	3.4	85.7	39.5
Motility	MOT	15.2	0.0	8.7	81.8	93.0	31.0	4.8	47.2
Oxidase positive	OXI	4.8	100.0	16.5	61.0	95.3	6.9	0.0	23.4
Pectin hydrolysis, pH7.8	PGA	9.2	100.0	3.1	3.9	4.7	10.3	0.0	13.7
Pectin hydrolysis, pH5.5	PLY	5.8	100.0	2.5	3.9	6.2	3.4	0.0	2.8
Pigmentation on 10% TSA	PIG	6.8	0.0	12.5	0.9	2.3	3.4	4.8	5.6
Fluorescence on King's B	FLU	0.5	0.0	0.3	0.0	24.8	0.0	0.0	0.0
NaCl (5% w/v)	NAH	89.6	0.0	83.8	47.2	20.9	93.1	4.8	18.5
$ZnSO_4$ (1 mM)	ZNC	15.7	0.0	6.5	83.5	31.0	93.1	0.0	9.3
$CuSO_4$ (1 mM)	COP	5.6	0.0	1.6	0.4	19.4	0.0	0.0	0.8
Neomycin (10 $\mu g\ ml^{-1}$)	NEO	92.6	0.0	29.6	22.5	86.0	51.7	95.2	25.0
Ampicillin (25 $\mu g\ ml^{-1}$)	AMP	33.3	5.3	71.0	93.9	97.7	55.2	9.5	42.7
Tetracycline (10 $\mu g\ ml^{-1}$)	TET	5.5	0.0	3.4	36.4	22.5	93.1	0.0	5.6
Crystal violet (0.125 $\mu g\ ml^{-1}$)	CVI	97.1	5.3	68.2	70.6	100.0	100.0	100.0	54.8
Rifampicin (10 $\mu g\ ml^{-1}$)	RIF	23.6	0.0	2.8	2.2	53.5	3.4	19.0	19.8
Vancomycin (20 $\mu g\ ml^{-1}$)	VAN	11.3	0.0	6.5	1.3	100.0	6.9	4.8	10.5
Streptomycin (20 $\mu g\ ml^{-1}$)	STR	13.5	0.0	4.7	3.9	21.7	24.1	4.8	21.0
CTAB (26 $\mu g\ ml^{-1}$)	CTB	7.0	0.0	4.0	3.5	100.0	10.3	0.0	2.8
Spectinomycin (25 $\mu g\ ml^{-1}$)	SPC	85.5	0.0	10.0	18.6	94.6	82.8	14.3	17.7
MM-glucose[b]	GLU	86.7	0.0	96.9	22.5	100.0	100.0	100.0	23.0
MM-mannose	MAN	85.0	0.0	23.4	4.3	93.8	0.0	95.2	8.5
MM-lactose	LAC	32.1	0.0	21.5	5.6	7.0	0.0	0.0	9.7
MM-xylose	XYL	68.5	0.0	90.7	3.5	24.8	6.9	4.8	10.1
MM-mannitol	MNL	86.5	0.0	98.4	27.7	99.2	100.0	100.0	35.5
MM-myo-inositol	INO	73.8	0.0	85.7	6.1	93.8	6.9	90.5	4.0
MM-rhamnose	RHA	32.6	0.0	27.1	1.7	1.6	3.4	9.5	6.9
MM-sucrose	SUC	92.8	0.0	96.9	26.0	98.4	96.6	100.0	22.2
MM-galactose	GAL	71.3	0.0	56.1	3.5	96.1	3.4	95.2	8.5
MM-azelaic acid	AZE	7.2	0.0	3.1	0.4	12.4	89.7	4.8	0.4
MM-succinic acid	SCN	76.1	0.0	22.7	5.2	97.7	100.0	90.5	5.2
MM-malonic acid	MAL	88.9	0.0	14.6	1.3	99.2	3.4	4.8	5.2
MM-sodium tartrate	TAR	64.4	0.0	7.2	0.9	84.5	3.4	0.0	1.6
MM-citric acid	CIT	94.4	0.0	56.1	22.5	97.7	93.1	100.0	13.7
MM-water	WAT	2.6	0.0	0.6	0.4	0.8	89.7	0.0	0.0
Number of isolates in cluster		585	19	321	123	129	29	21	248

Table 4. (*Continued*)

	Percent of isolates in each cluster with positive response to attribute test											
Code	9	10	11	12	13	14	15	16	17	18	19	20
GRA	0.0	5.6	93.3	32.4	13.9	15.8	0.0	0.0	77.1	91.7	11.1	5.9
BLO	0.0	2.8	96.0	11.8	4.2	42.1	25.0	24.1	15.7	87.5	83.3	5.9
GEL	10.0	2.3	82.7	20.6	8.3	89.5	93.7	89.7	28.6	69.8	63.9	23.5
CAS	5.0	2.8	89.3	8.8	2.8	100.0	93.7	89.7	15.7	60.4	63.9	5.9
LEC	0.0	0.5	0.0	2.9	1.4	31.6	56.2	3.4	5.7	3.1	8.3	0.0
LIP	0.0	1.4	61.3	2.9	2.8	47.4	25.0	24.1	8.6	44.8	52.8	0.0
STA	15.0	6.6	86.7	11.8	5.6	26.3	31.3	96.6	31.4	5.2	19.4	11.8
MOT	70.0	59.6	78.7	88.2	23.6	100.0	93.7	20.7	67.1	100.0	94.4	11.8
OXI	25.0	52.6	5.3	0.0	5.6	100.0	100.0	72.4	80.0	8.3	100.0	0.0
PGA	20.0	4.2	80.0	11.8	4.2	0.0	0.0	72.4	10.0	34.4	8.3	0.0
PLY	0.0	1.4	13.3	0.0	1.4	42.1	6.3	75.9	4.3	5.2	5.6	0.0
PIG	5.0	2.8	1.3	8.8	44.4	10.5	18.8	89.7	27.1	5.2	5.6	94.1
FLU	0.0	2.3	0.0	2.9	0.0	21.1	62.5	0.0	1.4	0.0	75.0	0.0
NAH	5.0	6.1	98.7	91.2	1.4	21.1	87.5	10.3	48.6	97.9	75.0	11.8
ZNC	80.0	96.2	8.0	88.2	15.3	15.8	31.3	31.0	77.1	87.5	41.7	88.2
COP	0.0	1.9	0.0	2.9	6.9	0.0	6.3	3.4	7.1	0.0	2.8	5.9
NEO	100.0	95.8	13.3	85.3	47.2	94.7	93.7	96.6	85.7	75.0	69.4	94.1
AMP	100.0	95.3	20.0	91.2	97.2	100.0	93.7	31.0	91.4	60.4	100.0	100.0
TET	10.0	22.5	9.3	44.1	13.9	47.4	12.5	31.0	31.4	3.1	41.7	82.4
CVI	90.0	55.9	96.0	100.0	91.7	89.5	100.0	100.0	97.1	87.5	91.7	100.0
RIF	10.0	68.5	1.3	73.5	100.0	73.7	37.5	17.2	32.9	10.4	69.4	100.0
VAN	50.0	74.2	5.3	100.0	95.8	94.7	100.0	62.1	65.7	5.2	100.0	100.0
STR	85.0	94.4	21.3	23.5	11.1	84.2	25.0	79.3	85.7	9.4	61.1	17.6
CTB	95.0	86.9	4.0	100.0	18.1	100.0	100.0	96.6	94.3	6.3	100.0	17.6
SPC	15.0	22.5	5.3	26.5	11.1	100.0	93.7	100.0	95.7	8.3	97.2	23.5
GLU	80.0	92.5	96.0	100.0	100.0	100.0	100.0	100.0	100.0	80.2	100.0	100.0
MAN	0.0	4.2	86.7	97.1	81.9	100.0	100.0	96.6	92.9	88.5	97.2	76.5
LAC	0.0	6.6	10.7	85.3	13.9	89.5	18.8	96.6	92.9	9.4	8.3	23.5
XYL	0.0	46.0	26.7	94.1	88.9	100.0	100.0	86.2	98.6	47.9	86.1	88.2
MNL	0.0	42.3	97.3	97.1	98.6	94.7	93.7	3.4	91.4	89.6	97.2	82.4
INO	0.0	3.3	40.0	94.1	43.1	100.0	93.7	3.4	94.3	21.9	47.2	0.0
RHA	5.0	1.9	2.7	73.5	44.4	94.7	18.8	20.7	95.7	5.2	11.1	23.5
SUC	50.0	64.8	89.3	100.0	95.8	100.0	100.0	93.1	97.1	76.0	94.4	82.4
GAL	30.0	71.8	1.3	100.0	94.4	100.0	100.0	100.0	100.0	80.2	100.0	94.1
AZE	75.0	79.8	0.0	2.9	40.3	15.8	25.0	3.4	15.7	0.0	8.3	23.5
SCN	5.0	57.3	8.0	88.2	83.3	100.0	100.0	10.3	87.1	34.4	100.0	82.4
MAL	70.0	66.7	0.0	52.9	61.1	100.0	100.0	0.0	37.1	0.0	100.0	23.5
TAR	20.0	36.6	0.0	2.9	34.7	89.5	6.3	0.0	38.6	2.1	5.6	17.6
CIT	10.0	45.5	14.7	88.2	29.2	94.7	100.0	13.8	22.9	79.2	97.2	82.4
WAT	0.0	0.5	0.0	0.0	0.0	0.0	6.3	0.0	1.4	0.0	0.0	0.0
# isol	20	213	75	34	72	19	16	29	70	96	36	17

(*continued*)

Table 4. Percentage of isolates with positive responses to attribute tests for each of 40 clusters, for component-community sampling protocol.[a] Clustering by two-stage density linkage method ($k = 4$) (*Continued*)

	Percent of isolates in each cluster with positive response to attribute test											
Code	21	22	23	24	25	26	27	28	29	30	31	32
GRA	100.0	72.7	100.0	31.8	97.0	5.0	100.0	88.9	20.0	100.0	13.3	100.0
BLO	87.0	22.7	82.4	27.3	42.4	10.0	0.0	22.2	37.8	4.0	60.0	0.0
GEL	87.0	18.2	100.0	18.2	90.9	90.0	0.0	0.0	24.4	72.0	100.0	6.7
CAS	65.2	4.5	100.0	13.6	87.9	90.0	0.0	0.0	6.7	0.0	100.0	6.7
LEC	0.0	4.5	100.0	9.1	42.4	5.0	0.0	0.0	2.2	0.0	33.3	0.0
LIP	30.4	0.0	17.6	13.6	18.2	20.0	14.3	22.2	2.2	8.0	53.3	0.0
STA	0.0	13.6	94.1	27.3	100.0	95.0	0.0	0.0	28.9	4.0	33.3	93.3
MOT	100.0	77.3	94.1	72.7	30.3	25.0	14.3	44.4	48.9	52.0	100.0	6.7
OXI	0.0	59.1	88.2	4.5	81.8	95.0	14.3	11.1	33.3	20.0	20.0	6.7
PGA	87.0	9.1	11.8	0.0	15.2	85.0	0.0	0.0	6.7	0.0	6.7	6.7
PLY	4.3	0.0	0.0	0.0	12.1	95.0	14.3	0.0	0.0	4.0	13.3	0.0
PIG	13.0	13.6	0.0	36.4	24.2	100.0	14.3	22.2	24.4	0.0	6.7	40.0
FLU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.6	4.0	0.0	0.0
NAH	100.0	31.8	100.0	27.3	87.9	35.0	0.0	5.6	64.4	68.0	46.7	33.3
ZNC	78.3	72.7	100.0	90.9	93.9	60.0	85.7	11.1	86.7	4.0	93.3	93.3
COP	0.0	9.1	5.9	90.9	3.0	15.0	0.0	0.0	37.8	0.0	73.3	0.0
NEO	30.4	90.9	88.2	77.3	27.3	100.0	100.0	27.8	95.6	0.0	100.0	93.3
AMP	43.5	100.0	100.0	100.0	100.0	80.0	100.0	83.3	97.8	80.0	100.0	86.7
TET	0.0	36.4	100.0	95.5	24.2	30.0	14.3	16.7	22.2	0.0	86.7	100.0
CVI	95.7	100.0	100.0	100.0	24.2	100.0	85.7	100.0	93.3	96.0	93.3	93.3
RIF	4.3	63.6	17.6	95.5	42.4	75.0	0.0	44.4	82.2	0.0	80.0	86.7
VAN	0.0	95.5	76.5	100.0	93.9	100.0	100.0	5.6	100.0	4.0	100.0	0.0
STR	0.0	95.5	23.5	95.5	81.8	95.0	0.0	5.6	95.6	4.0	100.0	73.3
CTB	0.0	81.8	29.4	100.0	90.9	100.0	85.7	50.0	88.9	0.0	93.3	0.0
SPC	0.0	100.0	52.9	100.0	100.0	100.0	57.1	38.9	93.3	8.0	93.3	93.3
GLU	82.6	100.0	94.1	95.5	100.0	90.0	85.7	94.4	100.0	100.0	0.0	100.0
MAN	56.5	100.0	94.1	100.0	84.8	90.0	28.6	94.4	95.6	8.0	0.0	100.0
LAC	4.3	100.0	0.0	9.1	36.4	90.0	85.7	100.0	93.3	12.0	0.0	73.3
XYL	4.3	100.0	11.8	72.7	90.9	90.0	85.7	94.4	97.8	44.0	6.7	86.7
MNL	60.9	100.0	88.2	95.5	97.0	95.0	100.0	77.8	97.8	92.0	20.0	100.0
INO	8.7	90.9	70.6	54.5	54.5	10.0	100.0	72.2	84.4	44.0	0.0	66.7
RHA	4.3	95.5	0.0	9.1	39.4	10.0	85.7	94.4	77.8	24.0	0.0	80.0
SUC	82.6	100.0	100.0	100.0	100.0	100.0	85.7	94.4	100.0	92.0	13.3	93.3
GAL	21.7	100.0	82.4	95.5	90.9	100.0	100.0	100.0	86.7	8.0	6.7	66.7
AZE	0.0	13.6	0.0	81.8	18.2	50.0	85.7	0.0	80.0	20.0	0.0	0.0
SCN	4.3	100.0	100.0	77.3	93.9	95.0	100.0	55.6	100.0	36.0	33.3	13.3
MAL	0.0	95.5	70.6	86.4	81.8	100.0	85.7	5.6	88.9	96.0	0.0	66.7
TAR	0.0	86.4	5.9	27.3	18.2	10.0	85.7	0.0	60.0	56.0	0.0	60.0
CIT	91.3	95.5	100.0	81.8	93.9	55.0	100.0	0.0	93.3	92.0	40.0	13.3
WAT	0.0	4.5	0.0	0.0	6.1	0.0	14.3	0.0	4.4	0.0	0.0	13.3
# isol	23	22	17	22	33	20	7	18	45	25	15	15

(*continued*)

Table 4. (*Continued*)

	Percent of isolates in each cluster with positive response to attribute test							
Code	33	34	35	36	37	38	39	40
GRA	81.8	22.2	58.3	88.2	10.0	0.0	100.0	40.0
BLO	90.9	0.0	0.0	5.9	40.0	0.0	0.0	20.0
GEL	90.9	83.3	29.2	5.9	100.0	88.9	20.0	100.0
CAS	36.4	11.1	8.3	5.9	100.0	100.0	0.0	100.0
LEC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.0
LIP	0.0	0.0	0.0	0.0	90.0	0.0	0.0	10.0
STA	63.6	77.8	100.0	94.1	90.0	100.0	10.0	90.0
MOT	100.0	88.9	100.0	0.0	60.0	0.0	10.0	80.0
OXI	9.1	88.9	45.8	0.0	100.0	44.4	100.0	10.0
PGA	100.0	5.6	62.5	5.9	60.0	33.3	100.0	0.0
PLY	81.8	0.0	0.0	0.0	40.0	100.0	100.0	10.0
PIG	0.0	5.6	0.0	29.4	100.0	100.0	100.0	80.0
FLU	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NAH	100.0	0.0	20.8	5.9	10.0	100.0	70.0	20.0
ZNC	45.5	5.6	41.7	76.5	100.0	0.0	100.0	90.0
COP	0.0	0.0	0.0	0.0	10.0	0.0	10.0	0.0
NEO	100.0	100.0	20.8	88.2	90.0	100.0	100.0	90.0
AMP	100.0	88.9	58.3	94.1	100.0	0.0	100.0	100.0
TET	0.0	5.6	4.2	88.2	90.0	11.1	100.0	0.0
CVI	100.0	100.0	75.0	100.0	90.0	100.0	70.0	80.0
RIF	54.5	0.0	0.0	82.4	20.0	33.3	100.0	20.0
VAN	45.5	100.0	4.2	23.5	30.0	66.7	90.0	90.0
STR	90.9	0.0	95.8	17.6	100.0	55.6	100.0	70.0
CTB	100.0	94.4	12.5	11.8	100.0	100.0	100.0	90.0
SPC	90.9	11.1	54.2	70.6	100.0	100.0	90.0	100.0
GLU	100.0	83.3	91.7	88.2	10.0	100.0	30.0	100.0
MAN	100.0	0.0	91.7	47.1	10.0	88.9	10.0	100.0
LAC	100.0	94.4	91.7	29.4	0.0	100.0	10.0	80.0
XYL	100.0	22.2	100.0	17.6	10.0	100.0	40.0	80.0
MNL	100.0	66.7	79.2	88.2	40.0	66.7	70.0	50.0
INO	100.0	27.8	12.5	11.8	0.0	66.7	20.0	10.0
RHA	100.0	88.9	25.0	11.8	0.0	22.2	0.0	100.0
SUC	100.0	77.8	83.3	94.1	0.0	100.0	90.0	100.0
GAL	100.0	55.6	91.7	29.4	20.0	100.0	90.0	100.0
AZE	0.0	0.0	0.0	0.0	0.0	0.0	50.0	0.0
SCN	100.0	27.8	0.0	5.9	0.0	100.0	100.0	10.0
MAL	27.3	5.6	8.3	0.0	0.0	55.6	90.0	20.0
TAR	45.5	5.6	12.5	0.0	0.0	0.0	10.0	10.0
CIT	100.0	22.2	4.2	17.6	20.0	100.0	90.0	10.0
WAT	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# isol	11	18	24	17	10	9	10	10

[a]Up to 6 isolates collected from each of 40–50 root samples (1cm) or soil samples ($1cm^3$)

Table 5. Percentage of isolates from each habitat assigned to each of 40 clusters (component-community sampling protocol[a]).

		Clusters																			
Expt.[b]	Habitat	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Arl89	Soil	9.2	0.0	35.8	19.7	3.5	0.6	0.0	9.8	0.0	1.2	1.2	0.0	0.0	1.2	0.0	1.2	0.6	0.0	4.0	0.0
	Untreated root	38.1	0.0	5.8	2.6	1.1	0.0	0.0	2.1	0.0	1.1	0.0	2.6	2.6	2.1	3.7	2.1	6.3	0.5	4.8	0.0
	UW85n1-trt root	28.6	0.0	2.2	2.7	1.1	0.0	0.5	2.2	0.0	0.5	0.0	1.1	1.1	1.1	2.2	11.0	4.4	0.5	3.8	0.0
Gch89	Soil	5.2	3.0	11.6	23.7	21.6	10.8	0.0	7.8	0.0	0.0	2.2	0.0	0.0	0.0	0.4	0.0	3.0	4.3	0.0	0.0
	Untreated root	42.4	1.8	2.8	26.3	0.0	0.0	8.3	7.4	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.0	0.0
	UW85n1-trt root	42.9	4.3	4.9	6.5	12.5	0.0	1.1	7.6	0.0	0.0	1.1	0.5	1.1	4.3	0.5	0.0	1.6	2.2	0.0	0.5
Arl90	Soil	14.7	0.0	36.6	3.9	3.9	0.9	0.0	9.5	0.0	0.4	2.6	0.0	0.0	0.0	0.0	0.4	5.2	10.8	0.0	0.0
	Untreated root	42.5	0.0	5.2	3.0	1.5	0.0	0.0	15.7	0.0	0.0	1.5	0.7	6.0	0.0	0.0	0.0	3.7	3.7	2.2	0.0
	UW85n1-trt root	21.5	0.0	13.9	3.2	1.9	0.0	0.0	13.3	0.0	1.3	1.9	2.5	5.7	0.6	0.0	0.0	0.0	8.9	1.3	0.0
	UW831-trt root	22.4	0.0	12.4	1.2	3.1	0.6	0.0	11.8	1.2	0.0	1.2	1.2	7.5	0.0	0.6	1.2	9.9	6.8	1.2	0.0
Han90	Soil	14.3	0.0	22.4	9.0	3.3	0.0	0.0	12.9	0.5	3.8	12.9	0.0	1.0	0.0	0.5	0.0	1.4	8.1	0.0	0.0
	Untreated root	16.0	0.0	4.4	7.7	3.9	0.0	0.0	12.7	0.6	26.0	4.4	3.3	4.4	0.0	0.0	0.0	0.6	1.7	2.2	4.4
	UW85n1-trt root	12.9	0.0	2.7	1.6	4.3	0.0	0.0	9.1	1.1	43.0	4.8	1.6	6.5	0.0	0.0	0.0	1.1	0.5	1.1	2.2
	UW831-trt root	8.5	0.0	3.8	3.3	2.4	0.0	0.0	11.8	6.6	33.0	3.8	4.7	5.7	0.9	0.5	0.0	0.0	0.9	0.0	1.9

[a] Up to 6 isolates collected from each of 40–50 root samples (1cm) or soil samples ($1cm^3$)

[b] Arl89, Arlington 1989; Gch89, Growth chamber 1989; Arl 90, Arlington 1990; Han90, Hancock 1990

Table 5. (*Continued*)

Expt.	Habitat	Clusters																			
		21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Arl89	Soil	0.0	0.0	0.0	0.6	9.8	0.6	0.0	0.0	0.6	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Untreated root	0.0	2.1	0.0	0.0	4.2	0.0	0.0	0.0	13.2	0.0	1.1	0.0	0.0	0.0	0.0	0.0	2.1	0.0	0.0	1.6
	UW85n1-trt root	0.0	1.1	0.5	3.3	2.2	10.4	0.5	0.0	1.6	0.0	4.9	0.0	0.0	0.0	0.0	0.0	1.6	4.9	5.5	0.0
Gch89	Soil	4.7	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.4	0.0	0.0	0.0	0.0	0.0
	Untreated root	0.9	0.9	7.4	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	UW85n1-trt root	1.1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.9	0.0	1.6	0.0	0.0	0.0	0.0	0.0
Arl90	Soil	0.9	0.0	0.0	0.0	0.4	0.0	0.0	0.4	0.0	5.2	0.0	0.9	0.0	1.3	2.2	0.0	0.0	0.0	0.0	0.0
	Untreated root	0.0	2.2	0.0	1.5	0.0	0.0	0.0	0.0	0.7	1.5	0.0	2.2	0.0	0.7	3.0	2.2	0.0	0.0	0.0	0.0
	UW85n1-trt root	1.3	1.3	0.0	2.5	0.0	0.0	3.2	0.0	0.6	1.9	0.0	2.5	0.0	3.8	1.3	5.7	0.0	0.0	0.0	0.0
	US831-trt root	0.6	1.2	0.0	0.6	0.6	0.0	0.0	0.6	2.5	1.9	2.5	0.6	0.0	4.3	1.2	0.6	0.0	0.0	0.0	0.0
Han90	Soil	1.4	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.5	0.0	0.0	1.9	0.5	1.0	0.0	0.0	3.3
	Untreated root	0.0	0.6	0.0	1.7	0.6	0.0	0.0	1.7	0.6	0.6	0.0	0.0	0.6	0.0	0.6	0.6	0.6	0.0	0.0	0.0
	UW85n1-trt root	0.0	0.0	0.0	1.1	0.5	0.0	0.5	2.7	1.1	0.5	0.0	0.0	0.0	0.0	0.5	0.5	0.0	0.0	0.0	0.0
	UW831-trt root	0.0	0.9	0.0	1.4	0.0	0.0	0.0	3.8	2.8	0.0	0.0	1.9	0.0	0.5	0.5	0.5	0.0	0.0	0.0	0.0

Table 6. Discriminant analysis of Growth Chamber 1989 experiment

a. Linear discriminant function[a]

Cluster	Coefficient vectors[b]		
	Soil	Nontreated root	UW85n1-treated root
1	6.27	17.12	15.83
2	5.90	11.72	12.90
4	7.95	15.07	10.93
7	7.31	20.84	14.08
8	8.00	16.16	15.15
14	15.00	36.20	46.84
23	5.97	21.60	12.91
33	6.85	16.94	20.61
Constant	−1.58	−7.74	−6.35

b. Test-classification of samples

Habitat of origin[c]	% of samples classified into habitat			% error[d]	Number of samples
	Soil	Nontreated root	UW85n1-treated root		
Soil	92.0	4.0	4.0	8.0	25
Nontreated root	0.0	70.8	29.2	29.2	24
UW85n1-treated root	26.1	8.7	65.2	34.8	23

[a]Linear discriminant function was trained on half of the samples from each habitat
[b]Coefficient vector = $COV^{-1}\,\bar{x}_j$ and constant = $-0.5\,\bar{x}_j'\,COV^{-1}\,\bar{x}_j$, where COV = pooled covariance matrix and $\bar{x}_j$ = mean vector of sample scores for habitat j
[c]Test-classified samples were those not used in training the linear discriminant function
[d]Total error = 24.0%

samples are more similar to UW85n1-treated root samples than to nontreated root samples. Nontreated and UW85n1-treated root samples were quite different from each other and from soil samples, with 60% and 64% of the nontreated and treated-root samples, respectively, being correctly classified.

In the Growth Chamber 1989 experiment (Table 6), none of the clusters appears to be individually characteristic of any given habitat, but in combination classify correctly 76% of the samples. This emphasizes the importance of assessing combinations of bacterial types, and not just designating individual "indicator" taxa. Soil and root samples are quite distinct from each other (Table 6b), as are communities on nontreated roots and UW85n1-treated roots. Most of the misclassified UW85n1-treated root samples were classified as soil samples, whereas misclassified nontreated root samples were always classified as treated-root samples. This once again suggests a greater similarity between communities in soil and on UW85n1-treated roots than between communities from soil and nontreated roots. These results agree with our previous results that indicated that coating seeds with UW85n1 had a significant impact on the development of the community of bacteria in the rhizosphere of treated plants [4].

Community Diversity

Component-Community Sampling Protocol. To determine whether the community diversity differed among habitats, we calculated the expected (rarefied) number of

Table 7. Diversity of bacterial communities in soil and rhizosphere habitats, as the expected number of different bacterial clusters in a standardized (rarefied) sample of 100 isolates, R(100). Estimates shown for bacteria grouped into three hierarchical levels (number of clusters) for the component-community sampling protocol[a]

Experiment	Habitat	R(100) when grouped into		
		20 clusters	40 clusters	60 clusters
Arlington 1989	Soil	5.6	13.8	23.5
	Untreated root	7.6	18.5	25.7
	UW85n1-trt root	11.2	22.4	27.7
Growth Chamber 1989	Soil	3.9	12.9	19.7
	Untreated root	2.9	9.9	18.3
	UW85n1-trt root	4.8	16.1	25.5
Arlington 1990	Soil	6.4	14.5	22.0
	Untreated root	7.7	17.9	25.7
	UW85n1-trt root	7.8	20.4	27.3
	UW831-trt root	10.4	22.7	31.2
Hancock 1990	Soil	6.3	16.2	25.2
	Untreated root	7.6	19.2	28.7
	UW85n1-trt root	6.4	17.9	27.5
	UW831-trt root	7.2	18.7	28.3

[a]Up to 6 isolates collected from each of 40–50 root samples (1cm) or soil samples ($1cm^3$)

clusters per 100 isolates for each habitat. This rarefied richness was calculated for each of three hierarchical levels: bacteria grouped into 20, 40, or 60 clusters. Rarefied richness provides a powerful index of diversity of communities, but statistical comparisons among indices from different communities are of limited reliability, and so is presented here for qualitative comparison only. Each hierarchical level of clustering provides similar results (Table 7), suggesting our assessment of diversity to be robust to the level of classification hierarchy.

In all three field experiments, the richness of the bacterial communities in the nontreated rhizosphere was always greater than in root-free soil. Furthermore, in both 1989 experiments, and to a lesser extent in the Arlington 1990 experiment, the bacterial communities on the roots of UW85n1-treated plants were richer still than communities on nontreated roots. However, no such increases in richness were associated with UW85n1 treatment at Hancock; in fact, there was a slight decrease in richness on treated roots. It is also noteworthy that the richness of the communities from all habitats in the Growth Chamber 1989 experiment was consistently less than the richness of their counterparts in other experiments. This may be due to the air-drying, sieving, and brief storage of soil prior to use, which resulted in a soil bacterial community less diverse than that found in the field soil, suggesting that the soil-handling procedures selectively reduced plant-associated bacteria from the soil community. Additionally, the Growth Chamber experiment differed from the field experiments in that the bacterial richness was greater in the soil than in the rhizosphere.

Infracommunity Sampling vs. Component-Community Sampling Protocols for 1990 Arlington. In order to evaluate the relative benefits of limited characterization of a large number of samples versus characterization of a larger number of isolates from a restricted number of samples, we analyzed the combined data from the

infra- and component-community sampling protocols for the 1990 Arlington experiment. We initially grouped the bacteria into 34 clusters, which were then hierarchically grouped further into three levels (30, 20, or 10 clusters), as shown in Fig. 2. At the 10-cluster level, most of the isolates were contained in one cluster, and this hierarchical level was not considered in further analyses.

In general, we found fairly similar frequencies of the clusters using either the infra- or component-community sampling protocols. However, there were several clusters (at the 30-cluster hierarchical level) that were isolated more frequently using one protocol or the other. A greater frequency for the infracommunity sampling protocol usually reflects local abundance on one or two samples. For instance, 32 isolates of cluster 11 were isolated from one nontreated root sample. Such local abundance is less apparent when a limited number of isolates is collected from each of a large number of samples, as was done in the component-community sampling protocol. A greater frequency for a particular cluster for the component-community sampling protocol may indicate that cluster was consistently isolated from a given habitat, regardless of whether it was encountered at very high local abundance in any of the samples processed using the infracommunity sampling protocol.

Table 8 shows the rarefied richness expected in standardized samples of 30 isolates for the 20- and 30-cluster hierarchical levels. The expected number of clusters is presented for individual plant and soil samples from the infracommunity sampling protocol, for the composite data from those same samples for each habitat, and for the composites of the component-community sampling protocol data (with few isolates from each of 36 or 40 samples). At both cluster levels, the composites of both the infra- and component-community sampling protocols suggest that the bacterial communities on either UW85n1- or UW831-treated roots have higher cluster richness than nontreated roots. The component-community sampling protocol also indicated that rhizosphere communities were more diverse than their soil counterparts. These results are in agreement with the results described above for the component-community sampling protocol. Rhizosphere communities appeared to be less diverse than soil communities in the infracommunity sampling protocol.

For root habitats, the rarefied richness estimates for composite component community samples were consistently higher than the composite infracommunity estimates (Table 8). This indicates that there is a considerable amount of variability in bacterial community composition among root samples, despite the overall similarities apparent from the discriminant analyses. In contrast, diversity in soil samples was greater for the composite infracommunity samples than for composite component communities, suggesting that the bacterial communities are relatively more homogeneous across many soil samples than among many roots.

Discussion

The combined use of cluster analysis, multivariate analysis of variance, discriminant analysis, and measures of community diversity provide an effective method for describing and comparing bacterial communities in the rhizosphere. This method permitted us to identify types of bacteria that are characteristic of each habitat, assess the effects of the introduced bacteria on both community diversity and

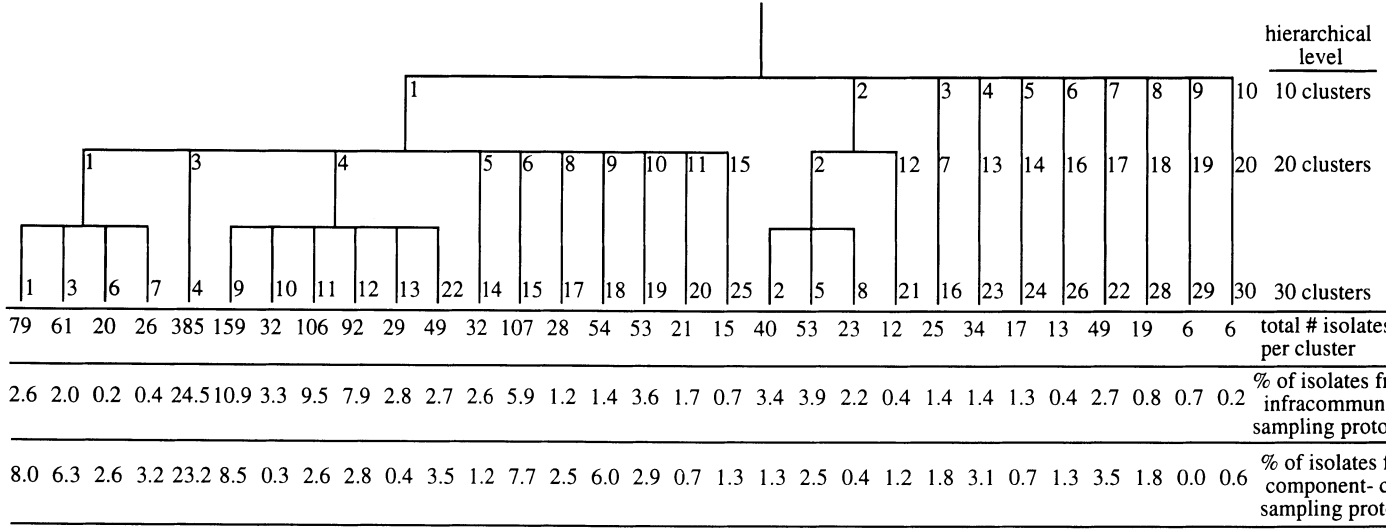

Fig. 2. Hierarchical relationships among clusters for Arlington 1990 experiment, infracommunity sampling (all colonies on one dilution plate per sample) and component-community sampling (up to 6 isolates from each of 40 soil or root samples) protocols combined. Cluster analysis performed using two-stage density-linkage method ($k = 4$).

Table 8. Bacterial community diversity of soil and rhizosphere communities, as the number of different bacterial clusters expected in standardized (rarefied) samples of 30 isolates. Estimates are shown for bacteria grouped into 20 or 30 clusters, and presented separately for the infra- and component-community sampling protocols, for the Arlington 1990 experiment

	Infracommunity sampling protocol[a]				Component-community sampling
Habitat	Sample1[b]	Sample2	Sample3	Composite[c]	Composite[d]
20 clusters					
Soil	9.8	7.7	9.6	9.5	8.9
Untreated root	7.4	5.9	3.4	6.0	9.7
UW85n1-trt root	7.1	6.3	9.1	8.4	11.3
UW831-trt root	5.3	6.8	6.1	6.7	10.3
30 clusters					
Soil	13.5	12.2	13.9	13.5	12.6
Untreated root	12.5	9.7	6.5	10.9	13.1
UW85n1-trt root	12.8	11.5	11.4	13.0	15.2
UW831-trt root	10.4	10.6	9.5	11.0	14.5

[a]Infracommunity sampling protocol = all colonies on 1 dilution plate per root segment (1 cm) or soil sample (1 cm^3); component-community sampling protocol = up to 6 isolates from each of 40 soil or root samples

[b]Individual plant or soil samples, from infracommunity sampling protocol

[c]Composites of isolates from 3 samples per habitat type, from infracommunity sampling protocol

[d]Composites of isolates from 36 (nontreated roots) or 40 (all others) samples per habitat, from component-community sampling protocol

composition, and to begin to address how sampling strategies and the grouping of bacteria into different hierarchical levels (numbers of clusters) affects our perceptions of soil and rhizosphere bacterial communities. We used this method to unravel some of the complex effects of the introduction of *B. cereus* UW85n1 on the development of specific bacterial communities in the rhizosphere.

Multivariate Analysis of Bacterial Communities

The cluster analysis used here was not intended to define groups of phylogenetically similar bacteria, but rather to create groups that were similar with respect to the physiological attributes that we tested. These attributes may or may not have ecological significance in the habitats of interest, but they are nevertheless useful as biological markers. The analysis shown here can be built upon through classical taxonomic assignment of clusters into genera or species, or through determining phylogenetic relationships among clusters. Additionally, methods can now be developed to study the ecological importance of individual types or the attributes they possess.

Overall, we could detect differences among communities from the various habitats when bacteria were grouped at a range of hierarchical levels, but discriminant analysis at intermediate numbers of clusters was more powerful than with many or few clusters (Table 2). This may be because, when isolates are clustered into a very few groups (i.e., 10 clusters), there is so much heterogeneity of bacterial types

within clusters that differences among habitats are not detectable. When isolates are divided into an increasingly large number of clusters, the relative abundance of each type necessarily must decrease, with each cluster that is present becoming increasingly rare. At some level, all clusters become so rare in each habitat that it is no longer possible to find a subset of clusters that will adequately characterize the component communities of the various habitats. We found that at the 40-cluster level the clusters were sufficiently homogeneous, and sufficiently common, to serve as good habitat discriminators.

Additionally, our assessments of community diversity based on the two sampling strategies (few isolates from each of many roots or many isolates from each of a few roots) appear to be similar, but sufficiently different to suggest that the ideal approach is to intensively sample bacteria from a larger number of samples. When time or funding is limiting, however, the decision to whether an infra- or component-community sampling approach is used should be based on whether the particular questions to be addressed pertain most to one or the other hierarchical level. For instance, when considering the effects of an introduced biocontrol agent, a key question might be how consistently, across a population of plants, the introduced organism changes the rhizosphere community in a particular way. In such a case a component-community sampling approach would be of greater value. Investigations on mechanisms of interactions between introduced and native bacteria on roots would likely be better addressed through a infracommunity approach. If isolating the maximum diversity from soil and root habitats is the goal, the inconsistent patterns of diversity for composite infra- or component communities suggests that it is more important to collect a large number of root samples (due to between root variability) than for soil samples, where a greater number of isolates from a few soil samples should yield a large bacterial diversity.

Our previous use of composite physiological attributes of the bacterial communities (from the component-community sampling protocol data included here) to discriminate among the bacterial communities from the various habitats [4] proved powerful for demonstrating differences among the communities. However, using the present analysis we can identify which types of bacteria differ among communities and how particular combinations of these types are important in describing the various habitats. This permits us to go beyond determining whether communities differ and provides information on which community member types are more or less common in a given habitat. Liljeroth et al. [14] presented a method for using cluster analysis of physiological attributes of bacterial isolates to identify types of bacteria characteristic of bacterial communities on root tips or bases, but it does not provide information about combinations of bacteria associated with particular habitats, and does not have the capabilities for classifying community samples as being similar to a particular community through discriminant analysis. This is a necessary foundation for understanding the ecological differences among the habitats, and how the introduced organism affects community composition and structure.

For the experiments presented here, we identified the types of bacteria characteristic of the soil and rhizosphere communities, and those affected by coating seeds with UW85n1. In general, it was not the presence or absence of individual types of bacteria that permitted discrimination among communities from the various habitats, but rather the combinations of clusters within samples. Nevertheless, certain diagnostic clusters were identified. Cluster 1 was much more commonly

isolated from root samples than from soil samples and was included as an important discriminator in 11 of 20 repetitions for Arlington 1989 and all 20 repetitions for Growth Chamber 1989. Clusters 3 (20/20 repetitions), 4 (19/20), and 8 (17/20) were also more or less consistently selected for inclusion in the Arlington 1989 discriminant functions. For Growth Chamber 1989, clusters 4 (20/20 repetitions), 7 (18/20), 14 (16/20), and 23 (19/20) were commonly selected. These clusters warrant further ecological and taxonomic investigation. However, none of these clusters alone is an adequate indicator taxon for a particular habitat—only through a predictive multivariate analysis such as discriminant analysis can discriminatory combinations of community members be identified.

We found it somewhat unusual, especially for the Arlington 1989 experiment, that there were no clusters in the LDFs that were unique to UW85n1-treated roots. In our previous analysis using physiological attributes for discriminant analysis [4], we found that pectolytic activity was a very strong indicator of communities on UW85n1-treated roots (40.6% of isolates from treated roots hydrolyzed pectin at pH 5.5, as compared to only 8.6% of isolates from nontreated roots and 14.3% from soil). Clusters 16 and 26 are both composed largely of pectolytic bacteria, and are much more common on UW85n1-treated roots than in the other two habitats. The two clusters differ primarily in their ability to grow on simple carbon sources, with cluster 26 having a much wider range. Both appear to belong to the genus *Cytophaga*. It seems odd that neither of these clusters was included in the LDF. Most likely, particular combinations of other clusters within a sample provided information redundant with the presence of clusters 16 and 26. In fact, when we added these two clusters to the clusters listed in Table 3a to create a linear discriminant function, there was no effect on the test-classification of any samples. Of the 10 UW85n1-treated root testing samples from which clusters 16 or 26 were isolated, only one was misclassified, indicating that the relative frequencies of other clusters in those samples was sufficient for discrimination.

There are at least two biological explanations for why the pectolytic bacteria, (*Ctyophaga*), did not contribute to the discrimination, although they were clearly characteristic of communities on UW85n1-treated roots. The unique combinations of bacteria that coexist on UW85n1-treated roots could create novel niches in the rhizosphere that can be filled by bacteria like clusters 16 and 26, which normally are not common members of the rhizosphere community. The presence of the pectolytic bacteria could therefore be symptomatic of other changes in the microbial community, which are described more directly by combinations of other types of bacteria; that is, the information in the relative frequencies of clusters 16 and 26 is redundant with information from the relative frequencies of other clusters. Alternatively, clusters 16 and 26 may be directly selected for by coating seeds with UW85n1, and their presence may modify the rhizosphere such that different combinations of bacteria can coexist in the rhizospheres of treated plants than in those of nontreated plants. It is also possible that there are no causal relationships between the pectolytic clusters and the other diagnostic clusters, but that each responds independently either to UW85n1, to a response of the plant to UW85n1, or to some other factor. Each of these explanations could also apply to other bacteria that appear to be characteristic of one habitat, but are not included in the linear discriminant function (see, for example, clusters 4, 38, and 39 in the Arlington 1989 experiment).

Community Diversity

Calculation of the rarefied richness of the component communities provides two valuable insights. The first is of logistical importance: the bacterial communities in the Growth Chamber 1989 experiment were consistently less rich than those in any of the field experiments, and only in this experiment was the richness of the nontreated root communities lower than that of the root-free soil. This suggests that soil-handling procedures (air-drying and sieving) or the controlled environment of a growth chamber selectively decreased the frequency of those soil bacteria with affinities for the rhizosphere, or the rhizosphere in its ability to support bacterial growth when plants are grown in the growth chamber. These results suggest that when addressing questions in which the microbial community plays an important role, caution is necessary when results from growth chamber experiments are used to approximate results predicted from a field experiment.

The second finding of particular interest is that in three of the four experiments (Arlington 1989, Arlington 1990, and Growth Chamber 1989), the richness of the communities on UW85n1-treated roots was higher than that of both the nontreated root and the soil communities. This may reflect a less specialized bacterial community on treated roots than on nontreated roots or indicate that the rhizosphere habitat has been altered by the introduction of UW85n1 such that bacteria that could not colonize nontreated root were able to successfully colonize treated root.

Implications for the Bacillus cereus *UW85 System*

The results of these experiments suggest a scenario for the altered development of rhizosphere communities on UW85n1-treated roots, and a potential relationship to disease suppression. When a growing root invades the soil habitat, it causes a disturbance to the microbial community around it. This disturbance, in the form of a suddenly abundant nutrient supply, provides an easily invasible habitat, and selects from the bacteria in the soil a distinct set of specific "rhizosphere" bacteria. Many root colonists (pathogens included) respond to biochemical cues from the root that stimulate chemotaxis, chemotropism, or emergence from a dormant state. The microbes that colonize the root surface utilize particular components of the root exudates, and produce metabolites of their own, so that the net biochemical appearance of a root results from the interaction of the host plant and its rhizosphere microflora. Later colonizers must interact with the early rhizosphere community. If seeds are first coated with large numbers of *B. cereus* UW85, which is characteristically a soil-inhabiting bacterium and not commonly associated with the rhizosphere, the biochemical appearance of the rhizosphere may be quite different from the appearance of the rhizosphere of a nontreated plant [5]. UW85 is known to produce Zwittermicin A, a broad-spectrum antimicrobial agent [9, 23] which, if produced in the rhizosphere may select for microbes that are resistant to the agent. As the soil microbes invade this UW85n1-disturbed habitat, these altered properties may enable bacteria that do not normally grow in the rhizosphere to colonize the root along with rhizosphere bacteria; this would result in the higher richness of bacteria on treated roots than on nontreated roots that we observed experimentally. Because early colonizers modify the appearance of the rhizosphere for later colonizers, the unique combinations of bacteria on the root surface can create niches for

non-rhizosphere bacteria (i.e., *Cytophaga*). Such a modified rhizosphere community could continue to develop in a manner unlike the communities on nontreated roots, even after UW85 is no longer a common member of the community.

The approach outlined is useful for comparisons of natural bacterial communities and for assessing the effect of introduced organisms or other disturbances on community composition and diversity. The effects of the introduction of UW85n1 on the rhizosphere communities points to the importance of community-level studies in understanding the impact of introduced microorganisms.

Acknowledgments. We thank Brenda Scholz and Lynn Jacobson for excellent technical assistance, and Douglas Emlen for critical evaluation of the manuscript. This work was supported by the University of Wisconsin College of Agricultural and Life Sciences and National Science Foundation grant number DCB-8819401.

References

1. Allen TFH, Starr TB (1982) Hierarchy—perspectives for ecological complexity. University of Chicago Press, Chicago
2. Bull CT, Weller DM, Thomashow LS (1991) Relationship between root colonization and suppression of *Gaeumannomyces graminis* var. *tritici* by *Pseudomonas fluorescens* strain 2-79. Phytopathology 81:954–959
3. Esch GW, Shostak AW, Marcogliese DJ, Goater TM (1990) Patterns and process in helminth parasite communities: an overview. In: Esch G, Bush A, Aho J (eds) Parasite communities: patterns and process. Chapman and Hall, New York, pp 1–19
4. Gilbert GS, Parke JL, Clayton MK, Handelsman J (1993) Effects of an introduced bacterium on bacterial communities on roots. Ecology 74:840–854
5. Gilbert GS, Handelsman J, Parke JL (1994) Root camouflage and disease control. Phytopathology 84:222–225
6. Gonzalez CF, Vidaver AK (1979) Bacteriocin, plasmid and pectolytic diversity in *Pseudomonas cepacia* of clinical and plant origin. J Gen Microbiol 110:161–170
7. Halverson LJ, Clayton MK, Handelsman J (1993) Population biology of *Bacillus cereus* UW85 in the rhizosphere of field-grown soybeans. Soil Biol Biochem 25:485–493
8. Handelsman J, Raffel S, Mester EH, Wunderlich L, Grau CR (1990) Biological control of damping-off of alfalfa seedlings with *Bacillus cereus* UW85. Appl Environ Microbiol 56:713–718
9. He H, Silo-Suh LA, Handelsman J, Clardy J (1994) Zwittermicin A, and antifungal and plant protection agent from *Bacillus cereus*. Tetrahedron Lett 35:2499–2502
10. Hurlbert SH (1971) The nonconcept of species diversity: an critique and alternative parameters. Ecology 52:577–586
11. Juhnke ME, Mathre DE, Sands DC (1989) Relationship between bacterial seed inoculum density and rhizosphere colonization of spring wheat. Soil Biol Biochem 21:591–595
12. Kloepper JW, Schroth MN (1981) Relationship of in vitro antibiosis of plant growth-promoting rhizobacteria to plant growth and the displacement of root microflora. Phytopathology 71: 1020–1024
13. Kloepper JW, Schroth MN, Miller TD (1980) Effects of rhizosphere colonization by plant growth-promoting rhizobacteria on potato plant development and yield. Phytopathology 70:1078–1082
14. Liljeroth E, Burgers SLGE, Van Veen JA (1991) Changes in bacterial populations along roots of wheat (*Triticum aestivum* L.) seedlings. Biol Fertil Soils 10:276–280
15. Lochhead AG (1940) Qualitative studies of soil microorganisms. III. Influence of plant growth on the character of the bacterial flora. Can J Res 18:42–53
16. Loper JE, Haack C, Schroth MN (1985) Population dynamics of soil pseudomonads in the rhizosphere of potato (*Solanum tuberosum* L.). Appl Environ Microbiol 49:416–422

17. McArthur JV, Kovacic DA, Smith MH (1988) Genetic diversity in natural populations of a soil bacterium across a landscape gradient. Proc Natl Acad Sci USA 85:9621–9624
18. Okon Y, Kapulnik Y (1986) Establishment of inoculated *Azospirillum* spp. in the rhizosphere and in roots of field grown wheat and sorghum. Plant Soil 90:1–3
19. Orvos DR, Lacy GH, Cairns Jr J (1990) Genetically engineered *Erwinia carotovora:* survival, intraspecific competition, and effects upon selected bacterial genera. Appl Environ Microbiol 56:1689–1694
20. Parke JL (1990) Population dynamics of *Pseudomonas cepacia* in the pea rhizosphere in relation to biocontrol of *Pythium.* Phytopathology 80:1307–1311
21. Parke JL, Rand RE, Joy AE, King EB (1991) Biological control of Aphanomyces root rot and Pythium damping-off of peas by *Pseudomonas cepacia* and *Pseudomonas fluorescens* applied to seeds. Plant Dis 75:987–992
22. SAS Institute, Inc. (1988) SAS/STAT user's guide, release 6.03 edition. SAS Institute, Inc., Cary, NC
23. Silo-Suh LA, Lethbridge BJ, Raffel SJ, He H, Clardy J, Handelsman J (1994) Biological activities of two fungistatic antibiotics produced by *Bacillus cereus* UW85. Appl Environ Microbiol 60: 2023–2030
24. Suslow TV, Schroth MN (1982) Rhizobacteria of sugar beets: effects of seed application and root colonization on yield. Phytopathology 72:199–206
25. Weller DM (1983) Colonization of wheat roots by a fluorescent pseudomonad suppressive to take-all. Phytopathology 73:1548–1553
26. Weller DM (1988) Biological control of soilborne plant pathogenes in the rhizosphere with bacteria. Annu Rev Phytopathol 26:379–407
27 Weller DM, Cook RJ (1983) Suppression of take-all of wheat by seed treatments with fluorescent pseudomonads. Phytopathology 73:463–469
28. Williams BK (1983) Some observations on the use of discriminant analysis in ecology. Ecology 64:1283–1291
29. Yuen GY, Schroth MN (1986) Interactions of *Pseudomonas fluorescens* strain E6 with ornamental plants and its effect on the composition of root-colonizing microflora. Phytopathology 76:176–180