INFECTION AND IMMUNITY, Dec. 1984, p. 720–726
0019-9567/84/120720–07$02.00/0

Vol. 46, No. 3

Variation in Periodontal Floras

W. E. C. MOORE,[1]* L. V. HOLDEMAN,[1] E. P. CATO,[1] I. J. GOOD,[2] E. P. SMITH,[2] R. R. RANNEY,[3] AND K. G. PALCANIS[3]

Departments of Anaerobic Microbiology[1] and Statistics,[2] Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061; and School of Dentistry, Virginia Commonwealth University, Richmond, Virginia 23298[2]

Received 13 June 1984/Accepted 5 September 1984

Statistical analyses indicated (i) that the floras of individual samples taken from the depth of sulci with nickel-plated Morse 00 scalers were highly reproducible and representative of the flora present at any given time, (ii) that the different compositions of floras of different people with similar clinical signs were statistically highly significant, and (iii) that floras of different affected sites may differ significantly in some (two of three) people at any one time or may differ from week to week in other people (one of three). Thus the flora composition of individual sites appears to be in dynamic flux, probably in response either to environmental changes or to host responses. There was no evidence that double sampling per se (two single passes with 00 scalers) changed the composition of the flora. Repeat samples taken after 1 week were slightly more similar to the initial samples than were samples taken after 3 weeks.

The experimental design and statistical analyses of studies of periodontal floras have been difficult because estimates of the variability attributable to samples, sites, time, and persons were unknown. To determine which factors may be responsible for observed variations, we conducted a minimal experiment in which duplicate samples were taken in a modified Latin square design from three affected sites on each of three patients with moderate (chronic) periodontitis.

MATERIALS AND METHODS

Subjects were three men (35 to 52 years of age) with generalized moderate (chronic) periodontitis. Duplicate samples were taken from three sites affected with periodontitis (probeable depth, 4 to 6 mm) from each person. Two sites were resampled in duplicate after 1 and 3 weeks according to the schedule in Table 1. Clinical measurements of these sites are shown in Table 2.

Sample sites were isolated with sterile cotton rolls. The supragingival area coronal to the sample sites was dried with sterile swabs and cleaned with sterile toothpicks. Samples were taken with individual sterile nickel-plated Morse 00 detachable-tip scalers inserted to the depth of the sulcus.

Each scaler tip was placed in prereduced anaerobically sterilized dilution broth (9) with 100-μm glass beads in a tube flushed with oxygen-free CO_2. The samples were dispersed by vibrating the stoppered tubes on a Vortex mixer for 10 to 12 s. The original suspensions were serially diluted and cultured on prereduced anaerobically sterilized brain heart i nfusion agar supplemented with vitamin K, hemin, powdered yeast extract, fresh yeast extract, formate, fumarate, serum, and thiamine pyrophosphate (BHIA-D4 agar [9]) in roll tubes and spread on plates of BHIA-D4 blood agar. Plates were incubated under 10% H_2–10% CO_2–80% N_2, and tubes were incubated under 3% H_2–10% CO_2–87% N_2 as described previously (8–10).

After 5 days of incubation at 37°C all colonies were counted, and 15 from each sample were picked in a random manner from plates; another 15 were picked from roll tubes. Each isolate, grown in BHI-D5 broth (9), was streaked to agar medium and repicked to assure purity. If the streaked isolates produced more than one colony type, representatives of each colony type were picked and identified.

Isolates were identified by Gram stain, electrophoretic pattern of soluble cellular proteins (7), chromatographic analysis of acid and gaseous products, and 30 or more biochemical and cultural tests as required for individual taxa (4).

The data were analyzed by Good's L (or lambda) test (3, 9) on the basis of the geometric mean of the percent concentrations of the taxa shared by each of the samples being compared and on the basis of the minimum percent similarity of the shared taxa. For example, if species X were 3% of the isolates from sample A and 12% of isolates from sample B, then the geometric-mean similarity is the square root of (3 × 12) = 6%, and the minimum similarity (of these two samples, for this species) is 3%, which is the actual percentage of the flora that the two bacteriological samples have in common. The summations of such values for all taxa shared by the two samples are called the geometric-mean similarity g and the minimum similarity g', respectively, of the two samples.

Briefly, for L analysis (3, 9) the similarity of each sample compared with every other sample is determined. Then the mean similarity ($\bar{g}$) between the two subgroups is divided by the mean similarity within the two subgroups to obtain an observed ratio, L. The probability that two subgroups having the observed mean within-subgroup similarity and the observed (usually lower) mean between-subgroup similarity (i.e., two groups that are this distinct) could occur among these samples by chance alone is then determined by dividing the total population of samples into two subgroups of the same size at random 1,000 times and calculating L each time. If no random assignment produces an equally low (or lower) L ratio, then the probability that the observed L could occur by chance alone is less than 0.001.

RESULTS

In this study, 132 taxa among 1,337 isolates were identified to species (or subspecies or serotype where possible). The distribution of species that comprised at least 1% of the flora in any one person is given in Table 3.

* Corresponding author.

TABLE 1. Site sample schedule

Person	Wk	Site[a] and sample[b]
1	0	A1, A2, B1, B2, C1, C2
	1	A3, A4, B3, B4
	3	B5, B6, C3, C4
2	0	A1, A2, B1, B2, C1, C2
	1	B3, B4, C3, C4
	3	A3, A4 C5, C6
3	0	A1, A2, B1, B2, C1, C2
	1	A3, A4, C3, C4
	3	A5, A6, B3, B4

[a] A, tooth 30 (military numbering system); B, tooth 3; and C, tooth 14 of each subject. Mesial sites were sampled on persons 1 and 3, and distal sites were sampled on person 2.
[b] Samples 1 and 2, 3 and 4, and 5 and 6 are replicate samples.

The results of the analyses of the similarity of duplicate samples, between sites within each person, and between times (weeks) within persons is shown in Table 4. Similarity values within each subset and between subsets, as calculated by both minimum and geometric-mean similarity, are listed. High probability values indicate similarity, and low probability values indicate dissimilarity, of the subgroups. The numbers of taxa shared between weeks in each person are shown in Table 5.

The distribution of dissimilarity as calculated from minimum similarity values is shown in Table 6, and the distribution of dissimilarity as calculated from geometric-mean similarity values is shown in Table 7. The individual dissimilarity values used to obtain the data in these tables were obtained by subtracting each similarity value from 100 (because 100 would be the similarity value if there were no difference between the two samples).

DISCUSSION

Duplicate samples. Although 132 taxa among 1,337 isolates were identified (to subspecies and serotype levels where possible), the duplicate samples taken from any one site at any one time were extremely similar as shown by a probability (P) of 0.994 (minimum) or 0.970 (geometric) when all first samples were compared with all second samples (Table 4). If each pair of first and second samples had been identical, the probability would have been 0.999 as shown by theory and actual test. (The maximum probability of 1.000 is not reached because Laplace's law of succession is applied to the L test, where 1 is added to the number of occurrences of equal or lesser L values [obtained by assigning the

TABLE 2. Clinical measurements of affected sites sampled in three subjects with moderate periodontitis

Person[a]	Tooth no.[b]	Sulcus depth (mm)	Wk 0			Wk 3		
			PI[c]	GI[d]	BI[e]	PI	GI	BI
1	30 M[f], A[g]	6	1	2	1	1	1	0
	3 M, B	4	1	1	1	1	0	0
	14 M, C	5	1	1	1	0	0	1
2	30 D, A	5	2	1	1	2	2	1
	3 D, B	5	2	1	0	1	2	1
	14 D, C	5	2	1	1	2	1	1
3	30 M, A	6	2	2	1	2	2	1
	3 M, B	6	2	2	1	2	2	1
	14 M, C	6	2	2	1	2	2	1

[a] Subject 1 was a 38-year-old black man, subject 2 was 52-year-old white man and subject 3 was a 37-year-old black man. Whole mouth measurements were taken on week 0 immediately after bateriological sampling. Whole mouth measurements also were taken only on subject 2 immediately after sampling on week 1. Measurements were taken on all subjects at the completion of the experiment. No change in pocket depths was observed.
[b] Military tooth numbering system.
[c] Plaque index of Silness and Löe (11).
[d] Gingival index of Löe and Silness (5).
[e] Bleeding index: 0, absent; 1, present after probing to the depth of the pocket.
[f] M, mesial surface; D, distal surface.
[g] Tooth designation used in Table 1.

TABLE 3. Distribution of predominant species in people

Species	% of isolates in person:		
	1	2	3
Actinomyces sp. WVA963[a]	0	0.7	1.1
Actinomyces israelii (−)[b]	0.2	1.4	0.4
Actinomyces israelii I	0.2	1.1	0
Actinomyces meyeri	2.5	0	0
Actinomyces meyeri (−)	1.6	0.4	0
Actinomyces naeslundii (−)	0.2	0.7	1.3
Actinomyces naeslundii I	0.2	0.2	3.1
Actinomyces naeslundii III	1.1	1.6	2.2
Actinomyces sp.NV	14.2	0.2	0.2
Actinomyces odontolyticus (−)	0.2	0	1.3
Actinomyces odontolyticus I	1.1	2.7	0.2
Actinomyces viscosus II	0.5	1.8	0.2
Bacteroides buccalis	0	1.1	0.4
Bacteroides sp. D10B	0	0.2	2.4
Bacteroides sp. D28	0.9	0.2	1.1
Bacteroides denticola	0.9	8.4	1.5
Bacteroides gingivalis	0	0.2	12.2
Bacteroides gracilis	0	3.6	0
Bacteroides intermedius 8944	0	4.7	0.4
Bacteroides veroralis	0	1.6	0
Bifidobacterium dentium	0	1.1	0
Capnocytophaga gingivalis	0	1.1	0
Capnocytophaga ochracea	0.2	1.1	0.2
Eubacterium alactolyticum	0.2	0.4	1.3
Eubacterium brachy	5.3	2.5	1.3
Eubacterium nodatum	1.6	1.8	9.4
Eubacterium timidum	3.4	0.2	1.3
Eubacterium sp. D06	0	0.2	2.0
Eubacterium sp. D08	0.7	3.2	2.0
Eubacterium sp. D11	0	0.7	3.9
Eubacterium sp. D13	1.1	0	0
Fusobacterium nucleatum	25.2	5.4	4.4
Lactobacillus minutus	1.1	2.3	2.0
Lactobacillus sp. D02	1.8	0.9	2.0
Propionibacterium acnes	3.2	1.4	0.4
Peptostreptococcus anaerobius	0.5	2.7	2.8
Peptostreptococcus anaerobius II	0.2	0	1.1
Peptostreptococcus micros	10.5	2.3	4.4
Rothia dentocariosa	0	1.1	0
Streptococcus anginosus	0.9	2.3	3.9
Streptococcus morbillorum	0	1.8	0
Streptococcus sanguis I	1.1	2.3	0.7
Streptococcus sanguis II	1.6	1.1	0.9
Streptococcus sanguis III	1.6	0	0
Selenomonas sputigena	0	1.6	2.6
Selenomonas sp. D04	0.9	0.7	1.5
Selenomonas sp. D06	0	1.1	0.2
Veillonella parvula	1.8	1.8	2.6
Wolinella recta	2.3	5.4	2.6
Did not survive through identification	1.4	2.5	3.5
Total	437	443	457

[a] Underscribed species are designated by letters or numbers.
[b] (−) indicates strains that have the phenotypic properties of the species listed, but that did not react with monospecific fluorescent antisera.

TABLE 4. *L* analyses of the distribution of bacterial species[a]

Subsets analyzed	Sample times[b]	Subset size	Minimum mean similarity[c]				Geometric mean similarity			
			W'n	W'n	B'n	*P*	W'n	W'n	B'n	*P*
All first vs all second samples[d]	All	21-21	17	17	18	0.994	25	26	27	0.970
Person 1 vs 2	All	14-14	26	21	14	<0.001	40	27	22	<0.001
Person 1 vs 3	All	14-14	26	26	15	<0.001	40	37	22	<0.001
Person 2 vs 3	All	14-14	21	26	16	<0.001	27	37	22	<0.001
Person 1 site A vs B	All	4-6	35	26	25	0.21	49	40	38	0.19
Person 1 site A vs B	Same	4-4	35	30	27	0.21	49	45	40	0.11
Person 1 site B vs C	All	6-4	26	21	27	0.59	40	33	41	0.71
Person 1 site B vs C	Same	4-4	28	21	28	0.66	46	33	43	0.69
Person 1 site A vs C	All	4-4	35	21	26	0.38	49	33	38	0.37
Person 2 site A vs C	All	4-6	39	24	17	0.005	48	32	24	0.005
Person 2 site A vs C	Same	4-4	39	22	15	0.03	48	33	22	0.03
Person 2 site B vs C	All	4-6	25	24	15	0.005	33	32	20	0.008
Person 2 site B vs C	Same	4-4	25	25	18	0.03	33	31	24	0.03
Person 2 site A vs B	All	4-4	39	25	24	0.03	48	33	29	0.03
Person 3 site A vs B	All	6-4	27	31	21	0.04	37	38	31	0.06
Person 3 site A vs B	Same	4-4	25	31	23	0.12	35	38	33	0.18
Person 3 site A vs C	All	6-4	27	37	21	0.01	37	51	31	0.01
Person 3 site A vs C	Same	4-4	32	37	22	0.03	45	51	31	0.03
Person 3 site B vs C	All	4-4	31	37	33	0.30	38	51	45	0.46
Mean between sites, same wks			25	28	22	0.18	42	38	32	0.18
Person 1 wk 0 vs 1 (A and B)	Same	4-4	40	26	26	0.12	57	37	40	0.09
Person 1 wk 0 vs 3 (B and C)	Same	4-4	35	31	22	0.09	49	48	36	0.05
Person 2 wk 0 vs 1 (B and C)	Same	4-4	20	26	19	0.20	29	33	25	0.18
Person 2 wk 0 vs 3 (A and C)	Same	4-4	22	23	20	0.29	31	31	29	0.22
Person 3 wk 0 vs 1 (A and C)	Same	4-4	24	26	29	0.89	38	36	39	0.72
Person 3 wk 0 vs 3 (A and B)	Same	4-4	20	29	26	0.63	30	38	35	0.63
Mean between wks same sites			27	27	24	0.37	39	37	34	0.32

[a] The mean percent similarity of the 21 duplicate pairs was 37 (minimum mean similarity [$\bar{g}$]) and 49 (geometric mean similarity [$\bar{g}'$]).
[b] In this column the word same means that the sites compared were sampled at the same sample times.
[c] W'n, mean within subset; B'n, mean between subsets; *P*, probability. The two columns headed W'n correspond to the first and second subsets.
[d] All first samples in one subset were compared with all second samples in the second subset as if they were replicate trials. Nearly identical results were obtained, which indicates that there was no immediate effect of sampling per se on the flora composition.

samples at random to two groups of the same size] and 2 is added to the number (1,000) of random test comparisons: 1,001/1,002 = 0.999.) The observed results give assurance that single samples are representative of the flora sampled, with an overall error among 21 pairs of 30-isolate samples of only 0.6% (minimum similarity) or 3.0% (geometric-mean similarity).

The highly representative nature of individual samples may surprise some readers, who will note the seemingly low (37% minimum and 49% geometric-mean) similarity among duplicate samples. However, this is to be expected because there are several scores of bacterial species in most sites at any one time, and it is impossible to detect all of them among 30 isolates taken at random from a single sample. Thus, multiple samples are required to differentiate, with statistical accuracy, between the floras of people, sites within people, times within sites, or, for that matter, between different disease states. Although larger samples (more isolates per sample) show greater similarity, the large proportion of variation that occurs between people and within disease classification indicates that more can be learned about the causes of periodontitis by taking more samples (e.g., more people, sites, or weeks) than can be learned by identifying more isolates from few samples. As we have emphasized earlier (9), the variation among samples within sites is more apparent than real.

Comparisons between persons. Fourteen 30-isolate samples per person were sufficient to show a statistically highly significant difference of flora compositions in different people (Table 4). If there are species that frequently are agents of tissue destruction, this real difference between floras of people with the same disease (in which many species are present) should help to pinpoint bacterial species or groups of species that are primary agents of tissue destruction.

There was little difference in the results when either the minimum similarity or the geometric-mean similarity was used in the comparisons. For differences between people, the observed *L* values were 7.3 standard deviations from the mean random *L* value for geometric-mean similarity and 6.8 standard deviations from the mean random *L* value for minimum similarity calculations (data not shown). Therefore, the geometric-mean calculations were very slightly more sensitive.

Sites within people and weeks within people. Similarity in composition of the floras of different sites within people depended upon the person. The floras of the three sites

sampled in person 1 were not significantly different. (P ranged from 0.21 to 0.66 [minimum] and 0.11 to 0.69 [geometric] [Table 4].)

Variation between affected sites was greatest in person 2. Among duplicate samples taken from any two of three sites on the same 2 days the probability that the observed difference would occur by chance alone was 1/35. That value represents the maximum L analysis sensitivity that can be attained with two groups, each of four samples, because the same sets of four samples will be assigned at random to two groups with a frequency of 1/35. When all six C samples were compared with the four available A or B samples, the probability was 0.005, which is the limit of sensitivity of the L analysis for groups with four and six samples. Although this comparison may have been confounded with week-to-week variation (because there were two unmatched C samples in each comparison) the floras of different sulci in person 2 may differ as much as the floras of different people. These results in themselves indicate that the flora of each site in this patient remained relatively constant as compared with the different flora compositions among sites. That is, the similarity within sites (from time to time) would have to remain high to obtain the observed level of significance with such a small number of samples.

In person 3 the flora of sample A differed from that of the other two sites, but there was no detectable change in composition from week to week.

The site-to-site variation in flora composition may explain the irregular patterns of active tissue destruction within individuals. It probably relates to different environments of different sulci and may correlate with Eh measurements as reported by Loesche et al. (6).

The data indicate that, although all of the sites were probed in each of the patients before these trials, the floras of individual sites remained distinct. Thus, probing with same instrument did not appear to distribute a uniform flora to all periodontal sites in each patient.

Week-to-week variation also is perhaps not the same among persons. It appeared that the greatest change (although it was not statistically significant) was in person 1, the only person in which the flora of the test sites did not differ significantly. The data suggest that the variation observed during 3 weeks usually was less than that observed among different affected sites at any one time in the same subject. Although the minimum and geometric-mean similarity calculations produced similar results for differences between sites within people, in five of six comparisons between weeks within people the geometric-mean calculations showed slightly greater sensitivity (lower probability). This might indicate that the week-to-week variation had a component of substantial change in relative numbers of certain species in one or more sites. The number of species shared from week to week is shown in Table 5.

Distribution of dissimilarity. The relative importance of variation in composition of the flora of different people, different periodontal sites at any given time within people, week-to-week variation in the same sulci, and duplicate samples is shown in Tables 6 and 7. In these calculations, nonindependence of samples was taken into account (see below). The difference between people was a major source of sample variation, even though the pocket depths and disease classification were comparable in all three subjects. In person 1 the floras of the three different sites were relatively similar (contributed less dissimilarity), but the change in the flora during the 3-week period was an important source of variation. Apparently, the change was reasonably similar among the three sites.

The data in Tables 4, 6, and 7 indicate that duplicate samples show relatively little variation, indicating that individual samples are reasonably representative of the flora at any given site and time; site-to-site variation in some people is a major source of variance (Table 4, person 2 and site A of person 3; Tables 6 and 7, person 3), week-to-week variation is a major source of variance in other people (person 1, Tables 4, 6, and 7); the compositions of different sites within individuals change differently with time (interactions within people, Tables 4, 6, and 7). In subjects 2 and 3, the difference in the floras of different sites at any given time contributed more variation than the week-to-week changes in flora composition of the periodontal sites (as might be expected from L analysis of the data [Table 4]). The high interaction in person 2 indicates that his flora changed differently in different sites between sample times.

These results suggest that there is a continuing dynamic flora-host relationship that might explain why periodontal destruction periodically occurs at only a few sites within a person. One might envision that the host is trying to control a menagerie of bacteria; as soon as one or more are

TABLE 5. Taxa shared between weeks within people

Person	Wk	No. of samples	No. of isolates	No. of taxa	% coverage[a]	No. of taxa shared		
						Wks 0 and 1	Wks 0 and 3	Wks 1 and 3[b]
1	0	6	184	28	92	14	11	13
	1	4	123	30	91			
	3	4	130	32	89			
2	0	6	192	64	85	30	29	27
	1	4	127	48	81			
	3	4	124	52	80			
3	0	6	196	52	88	31	22	21
	1	4	136	51	81			
	3	4	125	37	90			

[a] As determined by formula 9 of Good (2). This value indicates that the observed number of taxa accounted for the listed percentage of the total viable cells in the samples.

[b] Comparisons of weeks 1 and 3 are based on different sites. Therefore, these results are confounded by sites and weeks. In person 1, where the floras of different sites were not significantly different, the value for this 2-week period is intermediate between values for the the 1- and 3-week periods (as might be expected). In persons 2 and 3, there was a greater difference between the floras of different sites (Table 3). This difference may have decreased the (apparent) number of taxa shared in succeeding weeks.

TABLE 6. Dissimilarity distribution (calculated by minimum similarity)[a]

Source of dissimilarity	N comparisons[b]	Mean similarity[c] ($\bar{g}'$)	Sum of dissimilarity values	Half-width of 95% confidence interval	Total dissimilarity			Dissimilarity within person(s)		
					% of total	Expected % (if random)[d]	Observed % of expected	% of person(s)	Expected % (if random)	Observed % of expected
Total	861	0.177	70,834	1,411	100	100	100			
Between people	588	0.146	50,189	709	70.85 ± 1.00[e]	68.29	103.75 ± 1.47			
Within people	273	0.244	20,645	665	29.15 ± 0.94	31.71	91.93 ± 2.96	100	100	100
Between sites within wks	60	0.257	4,459	232	6.29 ± 0.33	6.97	90.24 ± 4.72	21.60 ± 1.12	21.98	98.27 ± 5.11
Between wks within sites	60	0.253	4,485	202	6.33 ± 0.28	6.97	90.82 ± 4.07	21.72 ± 0.97	21.98	98.82 ± 4.43
B'n duplicates W'n sites and wks[f]	21	0.367	1,329	124	1.88 ± 0.18	2.44	77.05 ± 7.38	6.44 ± 0.60	7.69	83.75 ± 7.80
Interaction within people	132	0.214	10,372	518	14.64 ± 0.73	15.33	95.50 ± 4.47	50.24 ± 2.50	48.35	103.91 ± 5.18
Person 1	91	0.263	6,703	477	9.46 ± 0.67	10.57	89.50 ± 6.36	100	100	100
Between sites within wks	20	0.323	1,354	189	1.91 ± 0.27	2.32	82.33 ± 11.47	20.20 ± 2.81	21.98	91.90 ± 12.80
Between wks within sites	20	0.209	1,582	130	2.23 ± 0.18	2.32	96.12 ± 7.79	23.60 ± 1.94	21.98	107.37 ± 8.84
B'n duplicates W'n sites and wks	7	0.450	385	97	0.54 ± 0.14	0.81	66.67 ± 17.29	5.74 ± 1.45	7.69	74.64 ± 18.86
Interaction within person	44	0.231	3,382	349	4.77 ± 0.48	5.11	93.35 ± 9.48	50.46 ± 5.20	48.35	104.36 ± 10.77
Person 2	91	0.207	7,212	309	10.18 ± 0.44	10.57	96.31 ± 4.15	100	100	100
Between sites within wks	20	0.208	1,584	89	2.24 ± 0.12	2.32	96.55 ± 5.29	21.96 ± 1.24	21.98	99.91 ± 5.63
Between wks within sites	20	0.261	1,478	109	2.09 ± 0.16	2.32	90.09 ± 6.83	20.49 ± 1.51	21.98	93.23 ± 6.85
B'n duplicates W'n sites and wks	7	0.314	480	68	0.68 ± 0.10	0.81	83.95 ± 12.35	6.66 ± 0.94	7.69	86.61 ± 12.22
Interaction within person	44	0.166	3,670	232	5.18 ± 0.33	5.11	101.37 ± 6.49	50.89 ± 3.21	48.35	105.25 ± 6.64
Person 3	91	0.260	6,730	331	9.50 ± 0.47	10.57	89.88 ± 4.41	100	100	100
Between sites within wks	20	0.240	1,521	111	2.15 ± 0.16	2.32	92.67 ± 6.81	22.60 ± 1.65	21.98	102.82 ± 7.49
Between wks within sites	20	0.288	1,425	123	2.01 ± 0.17	2.32	86.64 ± 7.38	21.17 ± 1.83	21.98	96.31 ± 8.31
B'n duplicates W'n sites and wks	7	0.337	464	66	0.66 ± 0.09	0.81	81.48 ± 11.11	6.89 ± 0.98	7.69	89.60 ± 12.74
Interaction within person	44	0.245	3,320	278	4.69 ± 0.40	5.11	91.78 ± 7.76	49.33 ± 4.13	48.35	102.03 ± 8.54

[a] The similarity values in the matrix of all samples compared against all other samples were each subtracted from 100% to obtain the percent dissimilarity. Appropriate subsets of dissimilarity values were added to give the amount of dissimilarity contributed to the total dissimilarity by each measured variable. The 95% confidence interval of each of these subtotals was determined by: Subtotal $\pm t_{N-1,0.975} \times \sqrt{N} \times$ standard deviation. The calculation of the standard deviation is explained in the text.

[b] $861 = \binom{42}{2}$; $588 = 3 \times 14^2$; $91 = \binom{14}{2}$; $20 = [\binom{6}{2} - 3] + [\binom{4}{2} - 2] + [\binom{4}{2} - 2]$; $44 = 91 - 20 - 20 - 7$; $273 = 3 \times 91$; $60 = 3 \times 20$; $21 = 3 \times 7$; $132 = 3 \times 44$. Here 7 is the number of pairs of bacteriological samples per person.

[c] The calculation of the mean similarity $\bar{g}'$ is exemplified by $1 - (708.34/861) = 0.177$.

[d] If all of the variation were random, then each subset of samples should contribute dissimilarity in proportion to the number of samples in the subset.

[e] ±, half-width of 95% confidence interval (see the text for calculations).

[f] See footnote c of Table 4.

controlled, others overpopulate and require specific attention. The information may be useful for designing suitable experiments, but it does not tell us whether tissue destruction results from a specific flora (or floras) or whether changes in host resistance allow tissue destruction by types of flora that cause no measurable destruction at other times.

APPENDIX A

Variance of the estimated mean similarity. Suppose we have two multinomial samples, of sample sizes M_1 and M_2, each with t categories. Let the sampled frequencies be denoted by (m_i) and (n_i) $(i = 1, 2, \cdots, t)$ where $\Sigma\, m_i = M_1$, $\Sigma\, n_i = M_2$. Let the population frequencies be called P_i and Q_i. Then m_i/M_1, denoted by p_i, is an

TABLE 7. Dissimilarity distribution (calculated by geometric-mean similarity)[a]

Source of dissimilarity	N comparisons	Mean similarity ($\bar{g}$)	Sum of dissimilarity values	Half-width of 95% confidence interval	Total dissimilarity			Dissimilarity within person(s)		
					% of total	Expected % (if random)	Observed % of expected	% of person(s)	Expected % (if random)	Observed % of expected
Total	861	0.261	63,599	1,805	100	100	100			
Between people	588	0.223	45,672	997	71.81 ± 1.56	68.29	105.15 ± 2.89			
Within people	273	0.343	17,927	786	28.19 ± 1.24	31.71	88.90 ± 3.91	100	100	100
Between sites within wks	60	0.355	3,873	250	6.09 ± 0.40	6.97	87.37 ± 5.70	21.60 ± 1.40	21.98	98.27 ± 6.36
Between wks within sites	60	0.352	3,888	235	6.11 ± 0.37	6.97	87.66 ± 5.36	21.69 ± 1.30	21.98	98.68 ± 5.93
B'n duplicates W'n sites and wks	21	0.488	1,076	143	1.69 ± 0.22	2.44	69.26 ± 9.02	6.00 ± 0.80	7.69	78.02 ± 10.40
Interaction within people	132	0.311	9,090	653	14.29 ± 1.02	15.33	93.22 ± 6.66	50.71 ± 3.64	48.35	104.88 ± 7.53
Person 1	91	0.396	5,497	517	8.64 ± 0.82	10.57	81.74 ± 7.76	100	100	100
Between sites within wks	20	0.446	1,109	183	1.74 ± 0.29	2.32	75.00 ± 12.47	20.17 ± 3.32	21.98	91.77 ± 15.11
Between wks within sites	20	0.333	1,335	166	2.10 ± 0.26	2.32	90.52 ± 11.19	24.29 ± 3.01	21.98	110.51 ± 13.72
B'n duplicates W'n sites and wks	7	0.609	274	91	0.43 ± 0.14	0.81	53.09 ± 17.28	4.98 ± 1.66	7.69	64.76 ± 21.59
Interaction within person	44	0.368	2,779	387	4.37 ± 0.61	5.11	85.52 ± 11.84	50.55 ± 7.04	48.35	104.55 ± 14.56
Person 2	91	0.274	6,606	372	10.39 ± 0.59	10.57	98.30 ± 5.58	100	100	100
Between sites within wks	20	0.286	1,428	94	2.25 ± 0.14	2.32	96.98 ± 6.24	21.62 ± 1.42	21.98	98.36 ± 6.43
Between wks within sites	20	0.344	1,313	123	2.06 ± 0.19	2.32	88.79 ± 8.30	19.88 ± 1.87	21.98	90.45 ± 8.51
B'n duplicates W'n sites and wks	7	0.391	426	73	0.67 ± 0.11	0.81	82.72 ± 13.58	6.45 ± 1.11	7.69	83.88 ± 14.43
Interaction within person	44	0.218	3,439	298	5.41 ± 0.47	5.11	105.87 ± 9.21	52.06 ± 4.52	48.35	107.67 ± 9.35
Person 3	91	0.360	5,824	389	9.16 ± 0.61	10.57	86.66 ± 5.79	100	100	100
Between sites within wks	20	0.332	1,336	139	2.10 ± 0.21	2.32	90.52 ± 9.24	22.94 ± 2.38	21.98	104.37 ± 10.84
Between wks within sites	20	0.380	1,240	144	1.95 ± 0.23	2.32	84.05 ± 9.84	21.29 ± 2.47	21.98	96.86 ± 11.22
B'n duplicates W'n sites and wks	7	0.463	376	75	0.59 ± 0.12	0.81	72.84 ± 14.81	6.46 ± 1.29	7.69	84.01 ± 16.78
Interaction within person	44	0.347	2,872	330	4.52 ± 0.52	5.11	88.45 ± 10.10	49.31 ± 5.66	48.35	101.99 ± 11.69

[a] See footnotes *a* through *f* of Table 6 for explanations.

estimate of P_i and $n_i/M_2 = q_i$ is an estimate of Q_i. The "geometric-mean" measure of similarity in the pair of populations is defined as $\Sigma \sqrt{P_iQ_i}$ (1). It is estimated by $\Sigma \sqrt{p_iq_i}$, which we have called g although it is only an estimate.

A formula for the variance of g was given by van Belle and Ahmad (12), namely:

$$\text{var } (g) \approx (1/4)\ (1/M_1 + 1/M_2)(1 - g^2) \qquad (1)$$

It can be proved (Good and Smith, submitted for publication) that:

$$\text{cov } (g_{pq},g_{qr}) \approx (4M_2)^{-1}\ (g_{pr} - g_{pq}g_{qr}) \qquad (2)$$

where the notations are self-explanatory.

In the present paper we have computed a confidence interval for various subtotals of dissimilarities. Each such calculation used formulas (1) and (2) with the g's all taken as equal to the mean similarity of all the pairs of bacteriological samples in the subset corresponding to the subtotal.

As an initial estimate of the variance of g (or the dissimilarity $1 - g$) for a given subtotal, we computed $s^2 = \Sigma_i(g_i - \bar{g})^2/(N - 1)$ where g_i is the estimated similarity for one pair of bacteriological samples. This estimate does not allow for the covariances between pairs of similarity measures. We therefore adjusted it by multiplying by $1 + [(n_c \times \text{covariance})/(n_v \times \text{variance})]$ where n_c is the number of nonzero covariance terms, and $n_v = N$ is the number of variance terms. The ratio of covariance to variance is approximated by $(4M_2)^{-1}\ (g_{pr} - g_{pq}g_{qr})/[(1/4)\ (M_1^{-1} + M_2^{-1})(1 - g^2)]$. Since we have assumed that g_{pr}, etc., are all equal to g, and since the M's are all equal to 30, we may write this approximation as $(1/2)(g - g^2)/(1 - g^2) = (1/2)g/(1 + g)$. We assume a corresponding result for g'.

As an example of the calculation of n_v and n_c, consider the total similarity. There are 42 bacteriological samples in all, giving $\binom{42}{2} = 861$ pairs. The number of covariance terms is equal to the number of pairs of pairs that have a bacteriological sample in common and is 3 $\binom{42}{3} = 34440$.

ACKNOWLEDGMENTS

We thank Pauletta C. Atkins, Luba S. Fabrycky, Leesa R. Miller, and Sue C. Smith for bacteriological analyses; Faye M. Scott for glassware preparation; Phyllis V. Sparks for media preparation; Karen S. Loferski for electrophoretic analyses and computer programming; and Kimberly Hardison for patient logistics and record management. We thank the California State Department of Health Services for cooperative studies with the actinomyces and for the monovalent fluorescent antiactinomyces conjugates.

The work was financed by Public Health Service grant DE 05054 from the National Institute of Dental Research and by project 2025790 from the Commonwealth of Virginia.

LITERATURE CITED

1. **Bhattacharyya, A.** 1946. On a measure of divergence between two multinomial populations. Sankhyā **7**:401–406.
2. **Good, I. J.** 1953. The population frequencies of species and the estimation of population parameters. Biometrika **40**:237–264.
3. **Good, I. J.** 1982. An index of separateness of clusters and a permutation test for its significance. J. Statist. Computat. Simulat. **15**:81–84.
4. **Holdeman, L. V., E. P. Cato, and W. E. C. Moore (ed.).** 1977. Anaerobe laboratory manual. Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg.
5. **Löe, H., and J. Silness.** 1963. Periodontal disease in pregnancy. I. Prevalence and severity. Acta Odontol. Scand. **21**:533–551.
6. **Loesche, W. J., F. Gusberti, G. Mettraux, and S. Syed.** 1983. Relationship between oxygen tension and subgingival bacterial flora in untreated human periodontal pockets. Infect. Immun. **42**:659–667.
7. **Moore, W. E. C., D. E. Hash, L. V. Holdeman, and E. P. Cato.** 1982. Polyacrylamide slab gel electrophoresis of soluble proteins for studies of bacterial floras. Appl. Environ. Microbiol. **39**:900–907.
8. **Moore, W. E. C., L. V. Holdeman, E. P. Cato, R. M. Smibert, J. A. Burmeister, and R. R. Ranney.** 1983. Bacteriology of moderate (chronic) periodontitis in mature adult humans. Infect. Immun. **42**:510–515.
9. **Moore, W. E. C., L. V. Holdeman, R. M. Smibert, I. J. Good, J. A. Burmeister, K. G. Palcanis, and R. R. Ranney.** 1982. Bacteriology of experimental gingivitis in young adult humans. Infect. Immun. **38**:651–667.
10. **Moore, W. E. C., L. V. Holdeman, R. M. Smibert, D. E. Hash, J. A. Burmeister, and R. R. Ranney.** 1982. Bacteriology of severe periodontitis in young adult humans. Infect. Immun. **38**:1137–1148.
11. **Silness, J., and H. Löe.** 1964. Periodontal disease in pregnancy. II. Correlation between oral hygiene and periodontal condition. Acta Odontol. Scand. **22**:112–135.
12. **van Belle, G., and I. Ahmad.** 1974. Measuring affinity of distributions, p. 651–668. *In*: F. Proschan and R. J. Serfling (ed.), Reliability and biometry. Society of Industrial and Applied Mathematics, Philadelphia.