Classification of Microbial Defects in Milk Using a Dynamic Headspace Gas Chromatograph and Computer-Aided Data Processing. 1. Principal Component Similarity Analysis

Yasumi Horimoto,* Karoline Lee, and Shuryo Nakai

Department of Food Science, The University of British Columbia, 6650 Northwest Marine Drive, Vancouver, BC, V6T 1Z4 Canada

There is a need for reliable objective tests for the detection of off-flavors of milk, some of which are caused by bacteriological growth. The objective of the study was to distinguish abnormal from normal milk samples using a low-cost headspace gas chromatograph and principal component similarity analysis. UHT-sterilized milk was inoculated with *Pseudomonas fragi, Pseudomonas fluorescens, Bacillus subtilis, Enterobacter* a*erogenes, Lactococcus lactis,* and a mixed culture (*P. fragi*:*E. aerogenes*:*L. lactis* = 1:1:1) to approximately 4.0 $\log_{10}$ CFU mL^{-1}. The samples were stored at 4 °C for 10 days for *P. fragi* and *P. fluorescens* and 30 °C for 24 h for the remaining bacteria. A new multiavariate analysis technique, principal component similarity analysis, was capable of distinguishing milk samples inoculated with *P. fragi, P. fluorescnes, L. lactis, E.* a*erogenes*, and a mixed culture from uninoculated samples.

INTRODUCTION

In the dairy industry, milk off-flavors present serious problems. Off-flavors may develop in milk for many reasons, including bacteriological growth and the action of native milk or bacterial enzymes as well as chemical changes catalyzed by light or heavy metals (Badings and Neeter, 1980).

The presence or absence of abnormal quantities of microbial metabolites in spoiled food may show whether or not the spoilage was caused by microorganisms. Gas chromatographic analysis of biologically produced volatiles in the headspace vapor of several foods has been used for the characterization and the identification of several bacterial species (Eyles and Adams, 1986; Guarino and Krammer, 1969; Lee et al., 1979; Schafer et al., 1982).

The feasibility of using GC as a means of studying bacteria in milk has been shown by several researchers. Bawdon and Bassette (1966) studied the effect of growth of *Escherichia coli* and *Aerobacter aerogenes* on milk quality by direct analysis of headspace vapor containing volatile compounds produced during growth. Bassette et al. (1967) also measured volatile compounds produced at different periods of bacterial growth in milk to characterize various microorganisms. They discussed the advantages of gas chromatographic methods in studying bacterial metabolism.

Possible origins of contamination of raw milk are numerous, but the main source of psychrotrophic microflora can be attributed to the inadequate disinfection of milking equipment (Cousin, 1982). Growth of psychrotrophic organisms can develop the fruity off-flavor in milk. Several researchers have analyzed the fruity off-flavor in milk using headspace gas chromatography

(Hosono et al., 1973; Pierami and Stevenson, 1975; Reddy et al., 1967; Toan et al., 1965; Wellnitz-Ruen et al., 1982).

A new multivariate analysis technique, principal component similarity (PCS) analysis, has been developed (Vodovotz et al., 1993). It combines principal component analysis (PCA) and pattern similarity computation (Aishima et al., 1987). PCS was applied to gas chromatographic data to classify mango samples (Vodovotz et al., 1993) and HPLC data of cheddar cheese to identify quality defects (Furtula et al., 1994a,b).

Recently a low-cost gas chromatograph has appeared on the market. This portable gas chromatograph (SRI Model 8620, SRI Instruments, Inc., Las Vegas, NV) contains a built-in purge and trap system and uses a computer as its integrator. Its flexibility and low cost make it attractive for quality control purposes. In conjunction with more powerful computer programs, it is now feasible to construct an objective flavor evaluation system for routine quality control purposes.

The research was undertaken because of the need to develop objective methods for quality control of raw milk. The objective of the research was to distinguish samples with microbial defects from control UHT milk using a low-cost headspace gas chromatograph and principal component similarity analysis.

MATERIALS AND METHODS

Ultrahigh Temperature (UHT)-Sterilized Milk. One case of UHT milk (2% milk fat) in 12 1-L tetrapak cartons of the same lot number was obtained from a local dairy (Dairyworld Foods, Burnaby, BC) and stored at 4 °C. The cartons were opened aseptically with flamed scissors after the carton mouth was wiped with 70% ethanol. The procedures for preparation and sterilization of low-density polyethylene carboys fitted with sampling spigots (4 L capacity, Sybron/Nalgene, Rochester, NY) were conducted as described by Skura et al. (1986). Three carboys with magnetic stirring bars were sterilized by filling them with a 200 ppm sodium hypochlorite solution and letting them stand for 2 h at room temperature. Then they were rinsed throughly with sterile distilled water.

* To whom correspondence should be addressed. Telephone: (604) 822-3404. FAX: (604) 822-3959. E-mail: yasu@unixg.ubc.ca.

734 *J. Agric. Food Chem.,* Vol. 45, No. 3, 1997

Horimoto et al.

Milk (2 L) was aseptically transferred to each carboy. To mimic the highly oxygenated state of raw milk, air (Pacific Medigas Ltd, Vancouver, BC) continuously overlaid the milk which was stirred at 300 rpm on a VMR mini-stir magnetic table (VMR Scientific, Inc., San Francisco, CA) and equilibrated for 36 h before the inoculation. The stirring bar remained in the carboy throughout the experiment. Air was sterilized with 0.3-μm bacterial air vent filters (Gelman Sciences, Ann Arbor, MI) and controlled with a flowmeter (400 mL/min) (Series 150, Linde Union Carbide, Somerset, NJ). The dissolved O_2 content was measured in parts per million (ppm) using a calibrated portable Clark electrode O_2 probe (YSI, Yellow Springs, OH).

Microorganisms. The following cultures were used for this experiment: *Pseudomonas fragi* ATCC 4973, *Pseudomonas fluorescens* biotype A ATCC 17397, *Lactococcus lactis* ATCC 29146, *Enterobactor aerogenes* ATCC 13048, and *Bacillus subtilis* ATCC 6460 (American Type Culture Collection, Rockville, MD). The mixed culture consisted of three bacteria (*L. lactis, E. aerogenes,* and *P. fragi*) in a ratio of 1:1:1.

P. fragi and *P. fluorescens* were maintained by monthly transfer on trypticase soy agar (TSA) slants (Becton Dickinson Microbiology Systems, Lockeysville, MD) kept at 4 °C. *L. lactis* was maintained by weekly transfer on a brain heart infusion (BHI) agar slant (Difco, Detroit, MI) kept at room temperature. *E. aerogenes* and *B. subtilis* were also maintained by weekly transfer on nutrient agar slants (Difco, Detroit, MI) kept at room temperature.

P. fragi and P. fluorescens as Inocula. Erlenmeyer flasks (250 mL) containing 25 mL of trypticase soy broth (TSB) (Becton Dickinson Microbiology Systems, Lockeysville, MD) were inoculated with a loopful of each bacterium from a slant. Cultures were incubated at 21 °C for 18 h in a water-bath shaker at 160 rpm (Lab-Line, Melrose Park, IL). A 10-mL sample of the culture was centrifuged at 10000*g* for 15 min at 4 °C in a RC2-B Sorvall Superspeed centrifuge (Dupont Sorvall, Newtown, CT). The centrifugation was repeated after the supernatant was removed and the pellet resuspended in 10 mL of 0.1% peptone (Difco, Detroit, MI). The population of the pellet was estimated by measuring optical density at 660 nm with a Shimadzu UV-160 double-beam spectrophotometer (Tekscience, Oakville, ON). UHT milk (2 L) was inoculated with *P. fragi* and *P. fluorescens* to a population of approximately 10^4 CFU mL^{-1}. Samples were collected for plating to provide day 0 cell numbers after 30 min. The samples were stored at 4 °C for up to 10 days and taken at 2-day intervals for the determination of bacterial cell numbers. Samples of 30 mL were also collected into amber threaded vials (Fisher Scientific, Ottawa, ON), frozen at −20 °C, and analyzed by GC within 24 h. UHT milk without bacteria was included as a control. One trial was conducted.

L. lactis, E. aerogenes, B. subtilis, and a Mixed Culture of L. lactis, E. aerogenes, and P. fragi as Inocula. Brain heart infusion (BHI) broth (25 mL) was inoculated with a loopful of *L. lactis* and incubated at 30 °C for 20 h in a shaking water bath at 160 rpm (Lab-Line, Melrose Park, IL). Each nutrient broth (Difco, Detroit, MI) (25 mL) was inoculated with a loopful of *E. aerogenes* and *B. subtilis,* respectively, and incubated in the same way as for *L. lactis.* The centrifugation and washing procedures were carried out as described in *P. fragi* and *P. fluorescens* as Inocula.

The populations for these bacteria were estimated in the same way as for *P. fragi* and *P. fluorescens.* Since these mesophiles have very long generation times at 4 °C, 1.5 L of UHT-sterilized milk in a 2-L flask was inoculated with an initial population of approximately 10^4 CFU mL^{-1} and incubated at 30 °C. There was no air flow over the milk during this experiment. Instead, milk samples were incubated in a shaking water bath at 140 rpm (SW-20C, Julabo Labortechnik GmbH, Lab Equipment, Seelbach, Germany). Samples, except *B. subtilis* (trial 2), were collected and analyzed every 4 h up to 12 h, once at 24 h, and then terminated. Milk samples inoculated with *B. subtilis* (trial 2) were terminated at 48 h. The methodology for GC used with *P. fragi* and *P. fluorescens*

**Table 1. Summary Data for Random Centroid Optimization for the SRI Gas Chromatograph**

experiment no.	sample volume (mL)	purge temp (°C)	purge time (min)	response
1	8.3	35	28.29	2226.66
2	8.3	44	8.06	125.67
3	11.4	31	11.77	77.92
4	10.9	36	18.11	221.52
5	11.4	47	15.87	387.16
6	6.1	56	5.44	85.77
7	11.4	56	25.90	354.37
8	7.9	57	18.86	681.06
9	14.4	39	10.83	244.69
10	11.9	51	7.37	349.97
11	14.4	48	26.49	753.96
12	7.2	44	23.55	298.07
13	7.4	30	22.56	185.44
14[a]	11.4	52	18.90	3118.70
15[a]	11.3	52	21.21	2699.07

[a] Centroid points.

was also used for the individual and mixed bacteria. Two trials for each strain and one trial for a mixed culture were conducted.

Enumeration. Growth of *P. fragi* and *P. fluorescens* was monitored by dropping 20 μL on TSA plates in quadruplicate. Serial dilution was made with sterile 0.1% peptone. The psychrotrophic count was obtained after incubation of plates at 21 °C for 25 h. Growth of *L. lactis* was monitored by dropping 20 μL on BHI agar plates and incubated at 30 °C for 24 h. Growth of *E. aerogenes* and *B. subtilis* was monitored by dropping 20 μL on nutrient agar plates and incubated at 30 °C for 24 h. Colony morphology and growth for a mixed culture were observed on selective media under different incubation conditions. These conditions were determined by performing plate counts after various incubation periods and temperatures. The final conditions were as follows: crystal violet tetrazolium agar (Speck, 1976) for *P. fragi* at 7 °C for 5 days, MacConkey agar (Difco, Detroit, MI) for *E. aerogenes* at 35 °C for 24 h, and *Lactobacillus* MRS agar (Speck, 1976) for *L. lactis* at 30 °C for 24 h.

Dynamic Headspace Gas Chromatographic Analysis. A portable gas chromatograph (SRI Model 8610, SRI Instruments, Inc., Las Vegas, NV) equipped with a purge and trap system and a flame ionization detector was used for analysis of volatile components. To obtain the best results after conducting the minimum number of experiments, random centroid optimization (Dou et al., 1993; Nakai, 1990) was used to establish purge conditions. Simplicity of operation, high search efficiency, and reduced chance of arriving at local optima make this method an ideal choice for many applications (Girard and Nakai, 1991; Lee et al., 1990; Vallejo-Cordoba and Nakai, 1994; Vodovotz et al., 1993). The purge conditions which included purge time, purge temperature, and sample size were optimized by assessing the number of peaks and their areas. The range for each parameter was determined by literature search and preliminary experiments (Vallejo-Cordoba and Nakai, 1994; Vodovotz et al., 1993). The search space for optimum was confined to sample size ranging from 5 to 15 mL, sample temperature varying from 30 to 55 °C, and purge time ranging from 5 to 30 min. The selection for a column, a trap, an internal standard, and analytical conditions for GC were determined on the basis of previous analyses of milk (Vallejo-Cordoba and Nakai, 1994) and the analysis of mango using the SRI gas chromatograph (Vodovotz et al., 1993). A DB-624 megabore column, 75 m long, 0.56 mm diameter, and 3.0 μm film thickness (J & W Scientific, Folsom, CA), was installed along with a Tenax Trap (SRI Instruments, Inc., Las Vegas, NV). A 2.5 ppm solution of 4-methyl-2-pentanone (10 μL) (99.5% HPLC grade, Sigma, St. Louis, MO) was used for the internal standard. 1-Tetradecanol (10 mg) (Sigma, St. Louis, MO) was used to avoid foaming during the purging. The helium carrier gas (UHP grade, Pacific Medigas Ltd., Vancouver, BC) flow rate was set to 4.6 (±0.2) mL/min. The hydrogen (UHP grade, Pacific Medigas Ltd., Vancouver,

Principal Component Similarity Analysis

J. Agric. Food Chem., Vol. 45, No. 3, 1997 **735**
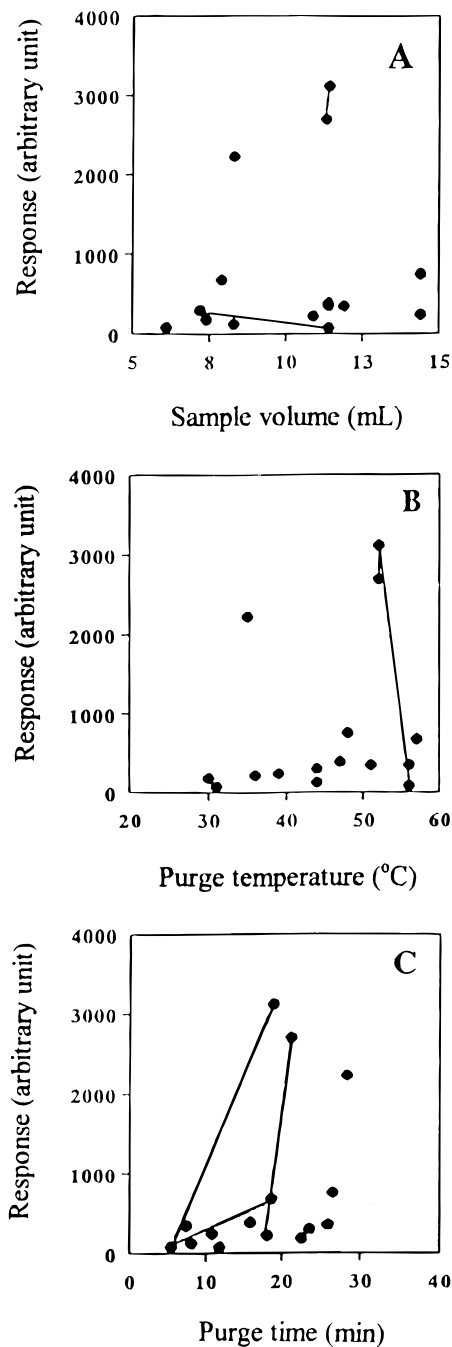
Figure 1. Mapping result of the first cycle of experiments generated by random centroid optimization technique: A, sample volume; B, purge temperature; C, purge time.

BC) and air flow rates for the detector were 20 and 250 mL/min, respectively. Frozen samples were warmed to 30 °C in a water bath (SW-20C, Julabo Labortechnik GmbH, Lab Equipment, Seelbach, Germany) for 15 min and analyzed. Purging conditions were optimized by the random centroid optimization technique. Purging conditions will be discussed in the Results and Discussion. The trap was dry purged for 10 min after purging the milk to eliminate water. Volatiles on the Tenax trap were thermally desorbed at 185 °C for 5 min and transferred into the column. The temperature program was decided by trial and error for the best chromatogram. It began at 30 °C for 5 min, was then ramped at 3 °C/min until 180 °C, and was subsequently ramped at 5 °C/min to 220 °C.

Data Manipulation. Data for each gas chromatogram were stored in an IBM-PC compatible 386/25 MHz personal computer interfaced with the GC system (Peak 2, Version 2.0, SRI Instruments, Inc., Las Vegas, NV). ASCII files were created for peak areas and retention time for each sample.

Table 2. Repeatability of Analytical Conditions for the SRI Gas Chromatograph

peak no.	retention time (min)	mean ($n = 4$)	SD[b]	CV (%)[c]
1	53.5	1.161[a]	0.034	2.92
2	57.3	0.143	0.003	1.84
3	57.8	0.109	0.002	1.71
4	60.2	0.054	0.001	2.72
5	67.4	0.057	0.007	12.32
6	75.4	0.061	0.016	26.34

[a] Peak area/peak area of internal standard. [b] Standard error of deviation. [c] Coefficient of variation.
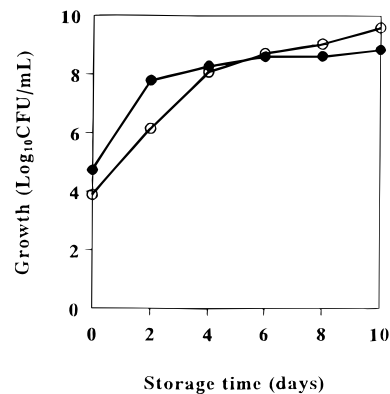
Figure 2. Growth of *P. fragi* and *P. fluorescens* in UHT-sterilized milk: ●, *P. fragi*; ○, *P. fluorescens*. Incubation temperature is 4 °C.

Each gas chromatogram was divided into several windows on the basis of retention time (Arteaga et al., 1994; Vallejo-Cordoba and Nakai, 1994). The selection of the numbers was based on the chromatogram which had the maximum peak numbers. Standardization of data was performed by dividing peak areas by the area of an internal standard. Each window was used as an independent variable. An automated peak recognition procedure was applied by using a program written in Basic for a personal computer (Vallejo-Cordoba and Nakai, 1994a). These data were transferred to a SYSTAT file. All statistical computations were made with an IBM-PC compatible 486/66 MHz personal computer.

The data set for *P. fragi* and *P. fluorescens* consisted of 30 milk samples. Each gas chromatogram was divided into 24 windows. In the cases of *L. lactis, E. aerogenes, B. subtilis,* and a mixed culture, the data set consisted of 104 milk samples. Each gas chromatogram was divided into 38 windows.

Principal Component Analysis (PCA). PCA was performed by a SYSTAT statistical software (FACTOR) program (Wilkinson, 1990) on the correlation matrix of the sample with no rotation.

Principal Component Similarity Analysis (PCS). Appendix 1 shows the PCS procedure. At first, PCA was calculated in a SYSTAT (FACTOR) program on the correlation matrix of the sample with no rotation. Then principal component scores (eigen value ≥ 1) were transferred to ASCII files. Linear regression analysis was then carried out to compute the coefficient of determination and slope of principal component scores of samples against a reference in the PCS program written in Quick Basic (Vodovotz et al., 1993).

RESULTS AND DISCUSSION

Dynamic Headspace Gas Chromatographic Analysis. Parameters for purge were optimized by using random centroid optimization (RCO), and the results are summarized in Table 1. The RCO consists of three steps in a cycle. They are random search, centroid search, and mapping process. If the optimum point is not obtained, the cycle is repeated until the optimum point is obtained. The total area of peaks of
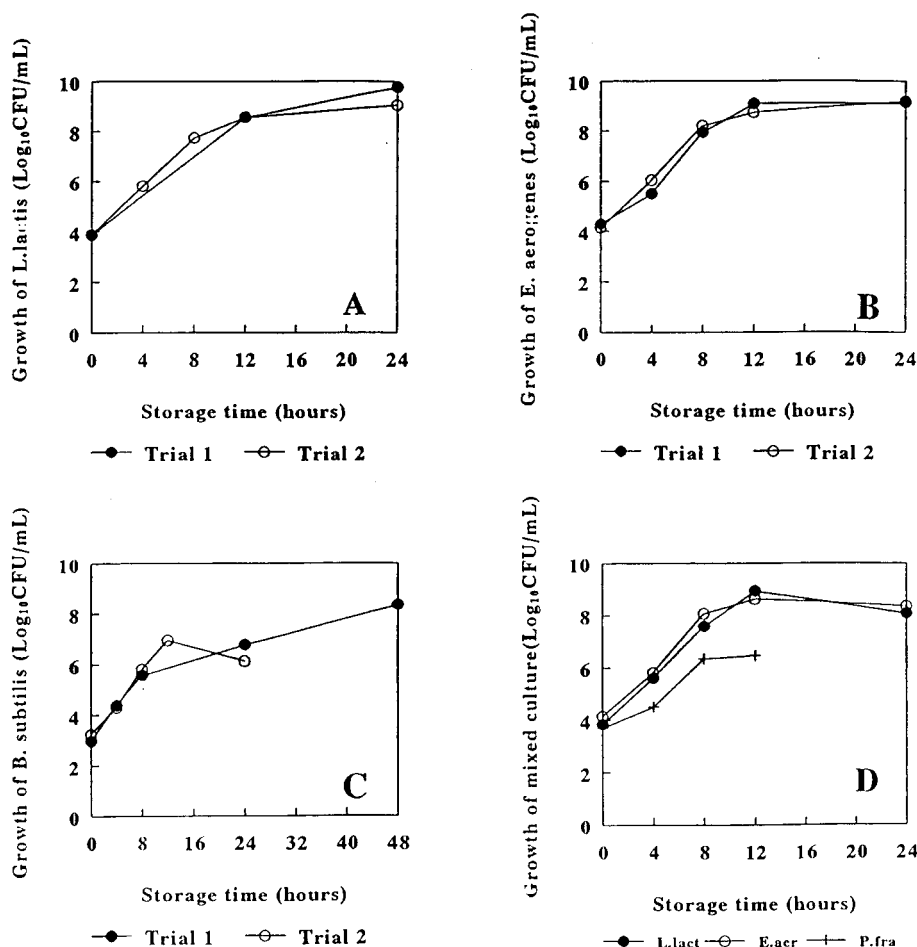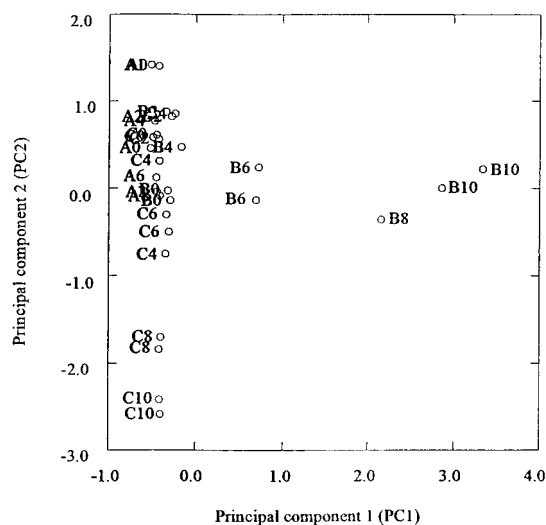
Figure 3. Growth of *L. lactis* (A), *E. aerogenes* (B), *B. subtilis* (C), and a mixed culture (*L. lactis:E. aerogenes:P. fragi* = 1:1:1) (D) in UHT-sterilized milk. Incubation temperature is 30 °C.
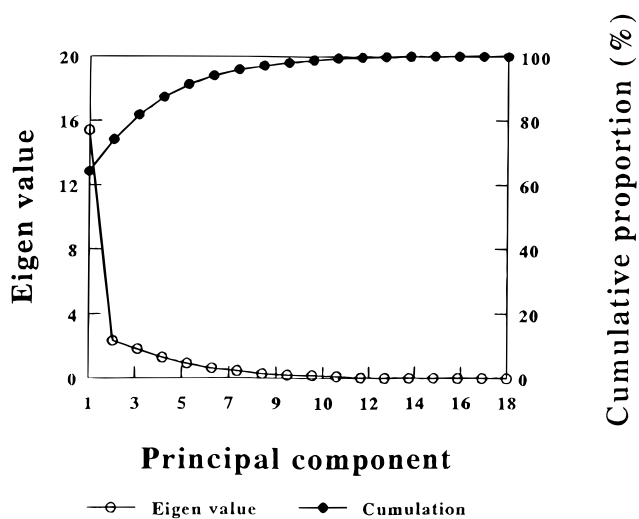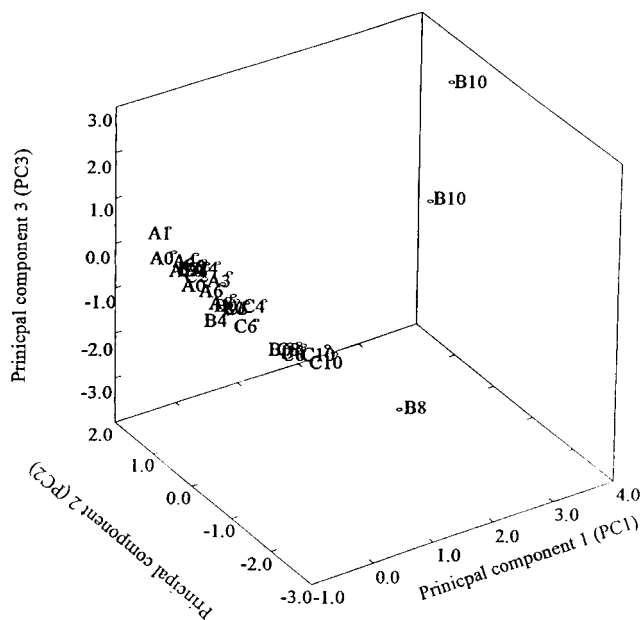

Figure 4. Eigenvalues and cumulative proportion in total variance on the basis of gas chromatographic peak areas for UHT milk inoculated with *P. fragi* and *P. fluorescens.*

Figure 5. Principal component analysis plot (two-dimensional) from gas chromatographic data for *P. fragi* and *P. fluorescens*: A, negative control; B, *P. fragi*; C, *P. fluorescens.* Numbers after each letter represent storage time (days).

gas chromatograms was calculated as a response value. The response value was maximized. The results of the first 13 experiments were used to generate the two centroid experiments. The centroid is the average of the parameters for *n* (number of experiments) experiments excluding the worst response in *n* + 1 experiments (Dou et al., 1993; Nakai, 1990). The heater may not be completely satisfactory for heating the entire tube, and the temperature sensor may be somewhat less accurate than optimum. This could cause somewhat

inconsistent results in Table 1. The coefficient of variation was 5−7%. Then all of these data were mapped to aid in narrowing the search space for the optimum (Figure 1). The mapping process aided in visualization of the experimental response surface, indicating the trend of the data (Nakai et al., 1984). The optimum was located by applying a curve-fitting program to link data points. From Figure 1, the optimum was reached for each parameter after 15 experiments.

Principal Component Similarity Analysis

J. Agric. Food Chem., Vol. 45, No. 3, 1997 **737**

Figure 6. Principal component analysis plot (three-dimensional) from gas chromatographic data for *P. fragi* and *P. fluorescens*: A, negative control; B, *P. fragi*; C, *P. fluorescens*. Numbers after each letter represent storage time (days).
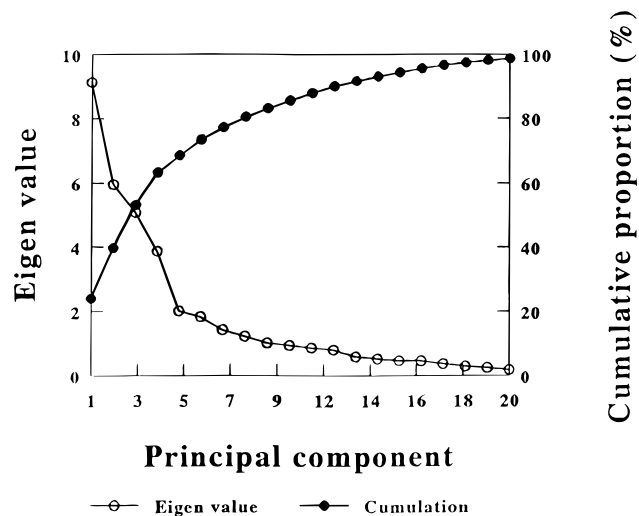

Figure 7. Eigenvalues and cumulative proportion in total variance on the basis of gas chromatographic peak areas from UHT milk inoculated with *L. lactis, E. aerogenes, B. subtilis,* and a mixed culture and incubated at 30 °C.

The parameters were sample size 11.4 mL, purging times 18 min and 54 s, and purging temperature 52 °C.

Coefficients of variation (CV) were calculated to investigate the conditions for repeatability (Table 2). Four chromatograms were run with the optimum conditions obtained from the RCO. Six major peaks were picked up and compared for their repeatability. Peaks 1−4 showed good repeatability (CV < 10%). When the retention time was longer (peaks 5 and 6), the CV increased.

Growth of Microorganisms. The dissolved oxygen tension of milk was 6.59 ppm before the inoculation. *P. fragi* and *P. fluorescens* showed the fastest growth rate within the first 4 days (Figure 2). By day 6, the population of *P. fragi* reached the stationary phase but the *P. fluorescens* continued to grow until day 10. *P. fragi* had a shorter generation time than *P. fluorescens*. Populations of both bacteria reached 8 $\log_{10}$ CFU mL^{-1}
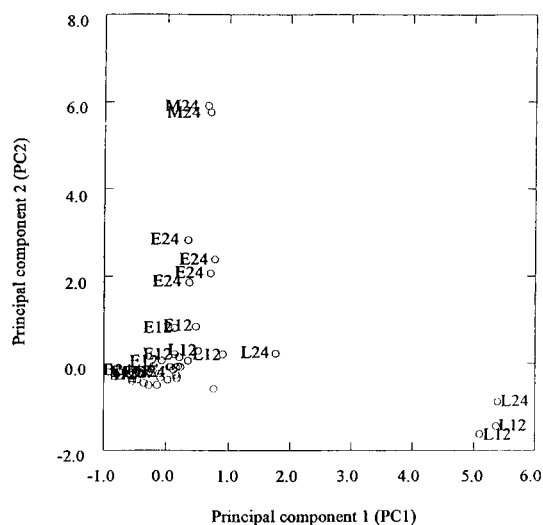
Figure 8. Principal component analysis plot (two-dimensional) from gas chromatographic data for *L. lactis, E. aerogenes, B. subtilis,* and a mixed culture: S, negative control; L, *L. lactis*; E, *E. aerogenes*; B, *B. subtilis*; M, a mixed culture. Numbers after each letter represent storage time (hours).
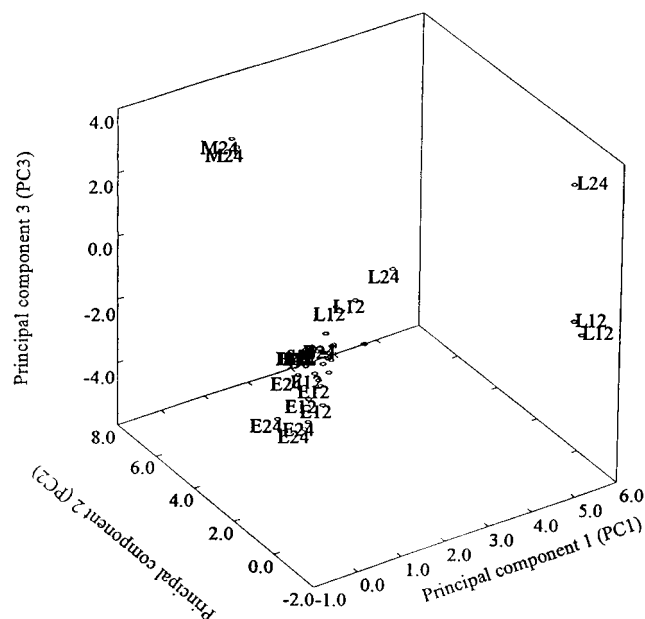
Figure 9. Principal component analysis plot (three-dimensional) from gas chromatographic data for *L. lactis, E. aerogenes, B. subtilis,* and a mixed culture: S, negative control; L, *L. lactis*; E, *E. aerogenes*; B, *B. subtilis*; M, a mixed culture. Numbers after each letter represent storage time (hours).

Table 3. Results of Cumulative Eigenvalues for *P. fragi* and *P. fluorescens*

principal component	eigenvalue	variance	cumulation (%)
1	15.42	64.257	64.257
2	2.364	9.848	74.105
3	1.847	7.697	81.802
4	1.326	5.527	87.329

after 4 days, which is the generally accepted numbers for spoilage (Griffiths and Phillips, 1986).

Growth curves for *L. lactis, E. aerogenes, B. subtilis,* and a mixed culture are shown in Figure 3. In each trial, the initial population of each strain was approximately 4 $\log_{10}$ CFU mL^{-1}. *L. lactis* grew to a high density, exceeding 8 $\log_{10}$ CFU mL^{-1} after 12 h. *E. aerogenes* reached the stationary phase after 12 h. The generation time of *B. subtilis* in trial 1 was slower than

738 *J. Agric. Food Chem.,* Vol. 45, No. 3, 1997

Horimoto et al.

Table 4. Curve-Fitting Equations for Each Bacterial Specie and a Mixed Culture

| bacterial species	curve-fitting equation	$r^{2\ a}$	F value
P. fragi	$y = 0.001 - 0.2327x + 5558.42x^2$	0.81	10.69
P. fluorescens	$y = 436.49 - 1414.44x + 1529.76x^2 - 550.70x^3$	0.86	14.65
L. lactis	$y = -41.64 + 189.18x - 256.72x^2 + 110.27x^3$	0.94	6.99
E. aerogenes	$y = 18.65 - 62.79x + 75.57x^2 - 30.51x^3$	0.60	40.96
mixed culture	$y = 9.43 - 18.18x + 9.75x^2$	0.99	419.41

^{a} Coefficient of determination.
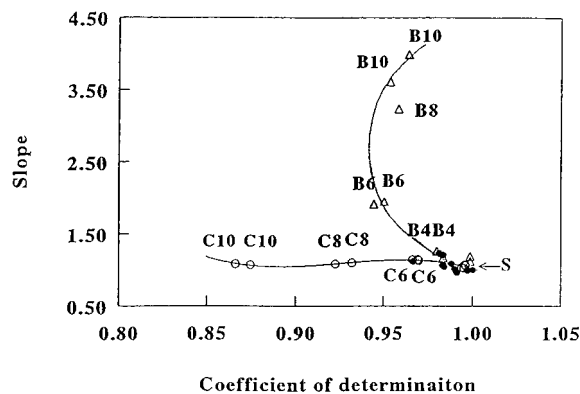
Figure 10. Principal component similarity analysis plot of gas chromatographic data for *P. fragi* and *P. fluorescens* using negative control (day 0) as a reference: ●, negative control; B, *P. fragi* (△); C, *P. fluorescens* (○). Numbers after B and C represent storage time (days). S = reference. The coordinates for the reference are (1, 1).

those of *L. lactis* and *E. aerogenes* and reached 8 $\log_{10}$ CFU mL^{-1} at 48 h. The population of *B. subtilis* in trial 2 began to decline after 12 h. In the mixed culture, population densities of *E. aerogenes* and *L. lactis* were similar for the first 12 h. After 12 h the populations of both bacteria began to decrease. *P. fragi* did not increase as much at 30 °C as at 4 °C. Consequently, the dilution scheme used for the 30 °C at 24 h sampling was too high to enumerate the population density of *P. fragi*. The total populations, by the end of the experiment, were lower when the organisms were grown in the mixed culture.

Principal Component Analysis (PCA): Effects of Inoculating Milk with *P. fragi* and *P. fluorescens.* Principal component (PC) scores were extracted among 24 peaks. Figure 4 shows eigen values and their cumulative proportion. The eigen value for a principal component indicates the variance that it accounts for out of the total variances (Manly, 1986). The cumulative proportion up to the fourth PC (eigen value ≥ 1.0) accounted for approximately 87%. Therefore, the information was extracted into four principal components with about 13% loss of the information. Figure 5 shows a plot of the values for the first two PCs, which account for 74% of the variation in the data. All samples until day 4 were not clearly classified. *P. fragi* had a larger PC1 with more storage days. On the other hand, *P. fluorescens* had a smaller PC2. Figure 6 shows a three-dimensional PCA plot. The first three PCs account for approximately 82%. When PC scores were plotted in three dimensions, there was no clear difference between the negative control (UHT-sterilized milk without inoculation) and *P. fluorescens.* The milk samples inoculated with *P. fragi* (after 8 days) were plotted far from the rest of the samples. Two- and three-dimension scattergrams are frequently examined for unsupervised classification. However, ignoring other PC scores by using only two or three PC scores may result in ignoring portions of the variation present in the original data;
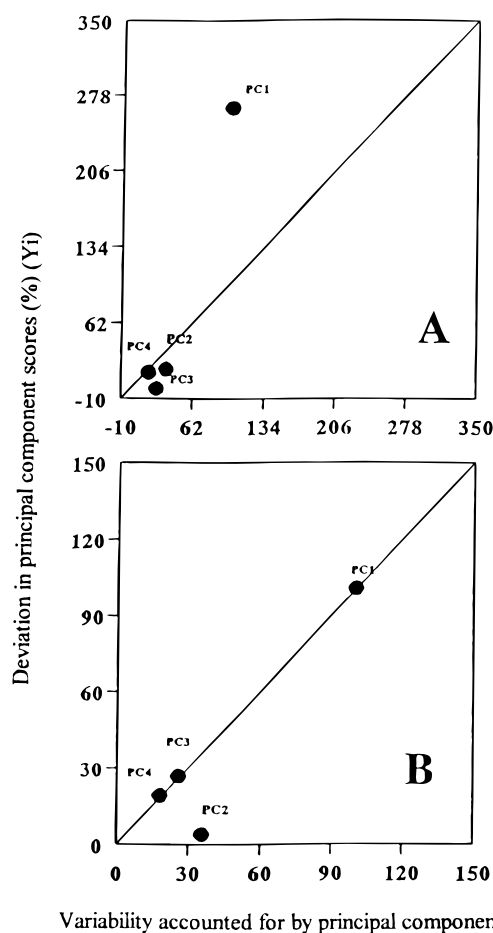

Figure 11. Adjusted principal component scores of UHT-sterilized milk inoculated with *P. fragi* after 8 days (A) and *P. fluorescens* after 8 days (B). Adjusted principal component scores were computed according to steps 1−3 in Appendix 1. The 45° line represents the match with PC scores of the reference.

this may be a reason PCA is not usually categorized as a classification method (Vodovodz et al., 1993).

Principal Component Analysis (PCA): Effects of Inoculating Milk with *L. lactis, E. aerogenes, B. subtilis,* and a Mixed Culture. Eigenvalues and their cumulative proportion are shown in Figure 7. The first nine PCs had eigenvalues greater than 1.0 and accounted for about 83% of the total variance. The first two PCs, which account for about 40% of the total variance, are plotted in Figure 8. Before 12 h all bacteria were not clearly differentiated from the negative control. PC1 for *L. lactis* increased after 12 h. PC2 for *E. aerogenes, B. subtilis,* and the mixed culture increased after 12 h. The first three PC scores are plotted in Figure 9. Compared to Figure 8, differences among *E. aerogenes, B. subtilis,* and the mixed culture are clear. PC3 scores for *E. aerogenes* decreased in value after 12 h. PC3 scores for the mixed culture were larger after 12 h. PC3 scores for *B. subtilis* were close
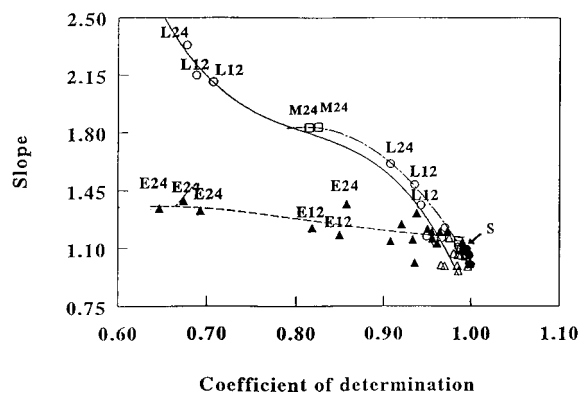
Figure 12. Principal component similarity analysis plot of gas chromatographic data for *L. lactis, E. aerogenes, B. subtilis,* and a mixed culture using a negative control (0 h) as a reference: ●, negative control; L, *L. lactis* (○); E, *E. aerogenes* (▲); B, *B. subtilis* (△); M, mixed culture (*L. lactis:E. aerogenes: P. fragi* = 1:1:1) (□). Numbers after L, E, and M represent storage time (hours). S = reference. The coordinates for the reference are (1, 1).

to that of the negative control. PC3 scores for the mixed culture were large.

Principal Component Similarity (PCS) Analysis: Effects of Inoculating Milk with *P. fragi* and *P. fluorescens.* In PCS analysis, PCA is conducted first and then PC scores are used to compute the coefficient of determination and slopes between adjusted PC scores of samples and those of a reference. According to the rule of thumb, an eigenvalue which is lower than 1.0 can be discarded (Vodovotz et al., 1993). PCA yielded four PCs with eigenvalues greater than 1.0, which accounted for approximately 87% of the total variation (Table 3). Thus the first four PC scores were used to conduct PCS. Figure 10 shows the PCS plot of gas

chromatographic peak areas. UHT-sterilized milk without inoculation (day 0) was used as a reference. For both *P. fragi* and *P. fluorescens,* there was little difference in the chromatograms between the reference and the inoculated samples until day 4. The chromatograms for *P. fragi* and *P. fluorescens* showed a substantial difference after day 6. The slope of *P. fragi* became greater with longer storage time, while the coefficient of determination stayed between 0.95 and 1.0. On the other hand, the slope of *P. fluorescens* was approximately 1.0, but the coefficient of determination became smaller with longer storage time.

The best curve-fitting equation was examined to predict the relationship between a quantitative response and explanatory variables (slope and coefficient of determination) for samples with *P. fragi* and *P. fluorescens.* SlideWrite software (Curve Fitter) (SlideWrite Plus, 1995) was used for this purpose. Among several functions, linear and polynomial groups were used. The equations which gave the best coefficient of determination for each bacterium and a mixed culture are shown in Table 4. The coefficient of determination for each equation was 0.81 for *P. fragi* and 0.86 for *P. fluorescens.* These equations can be used to predict the direction of each species. When the plot of an unknown sample is close to one of these lines, the bacterial species can be classified.

To display the difference between the sample inoculated with bacteria and the reference, adjusted PC scores of the sample (day 8) were plotted against the adjusted PC scores of reference (day 0) in Figure 11. This figure shows that points which fall on the diagonal line are similar to those obtained from the reference sample. As the sample approaches the reference, the deviation decreases, thus approaching the diagonal line. The PC1 of *P. fragi* deviated greatly from the PC1 of
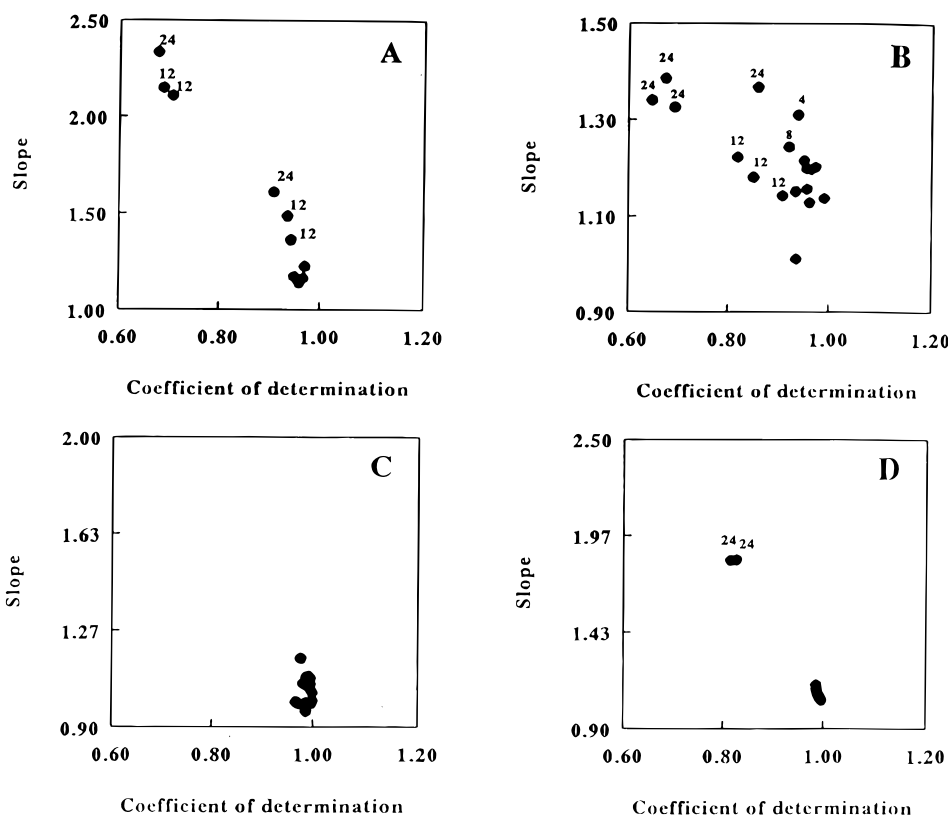
Figure 13. Individual principal component similarity analysis plots of gas chromatographic data for *L. lactis* (A), *E. aerogenes* (B), *B. subtilis* (C), and a mixed culture (D) using a negative control (0 h) as a reference. Numbers in plots represent storage time (hours). The coordinates for the reference are (1,1).
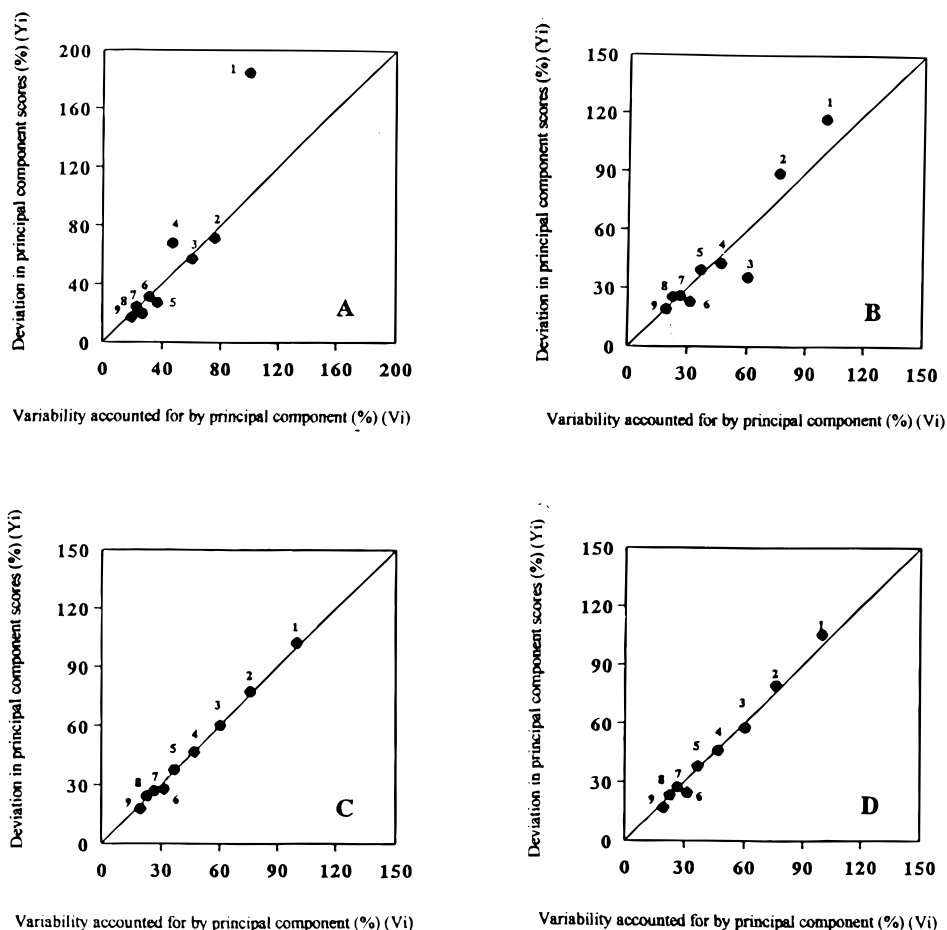
Figure 14. Adjusted principal component scores of UHT-sterilized milk inoculated with *L. lactis* (A), *E. aerogenes* (B), *B. subtilis* (C), and a mixed culture (D) after 12 h. Adjusted principal component scores were computed according to steps 1−3 in Appendix 1. The 45° line represents the match with principal component scores of the reference.

Table 5. Factor Loadings (Pattern) from Principal Component Analysis of Data for *P. fragi* and *P. fluorescens*

| peak no. | principal component (PC) | | | |
	1	2	3	4
1	−0.242	0.477	0.400	0.628
2	0.950	0.004	−0.197	0.030
3	0.991	0.010	0.074	−0.012
4	−0.522	0.594	0.217	−0.447
5	0.976	0.034	−0.128	0.103
6	0.881	0.002	0.207	−0.056
7	0.977	0.031	0.178	−0.005
8	0.932	−0.014	−0.022	−0.089
9	0.820	0.173	−0.205	0.336
10	0.963	−0.143	0.144	0.052
11	0.988	0.046	−0.059	0.061
12	0.819	0.046	0.364	0.081
13	0.927	0.027	0.335	−0.009
14	0.375	0.074	−0.826	−0.027
15	0.158	0.731	−0.140	0.035
16	0.803	0.045	0.153	−0.296
17	0.983	0.033	−0.106	0.088
18	0.921	−0.010	0.018	−0.081
19	0.986	0.018	0.124	0.002
20	0.689	−0.038	−0.575	0.026
21	−0.327	0.644	−0.057	0.450
22	0.637	0.298	0.242	−0.375
23	0.981	0.078	0.019	0.094
24	0.003	−0.819	0.236	0.358

the reference. The component loadings of the first four PCs are shown in Table 5. Peaks 3, 11, 17, 19, and 23 were found to have high loadings (≥0.98) in PC1. They may be used for classifying milk samples possibly contaminated with *P. fragi*. While the PC1 of *P.*

Table 6. Results of Cumulative Eigenvalues for *L. lactis, E. aerogenes, B. subtilis,* and a Mixed Culture

principal component	eigenvalue	variance	cumulation (%)
1	9.138	24.047	24.047
2	5.961	15.687	39.734
3	5.089	13.392	53.126
4	3.877	10.202	63.328
5	2.023	5.324	68.562
6	1.839	4.840	73.402
7	1.434	3.773	77.175
8	1.230	3.237	80.412
9	1.031	2.713	83.125

fluorescens was similar to the PC1 of the reference, PC2 showed a clear difference (Figure 11).

Principal Component Similarity (PCS) Analysis: Effects of Inoculating Milk with *L. lactis, E. aerogenes, B. subtilis,* and a Mixed Culture. PCS was carried out using the first nine PCs, which accounted for 83% of the total variance (Table 6). The PCS plot is shown in Figure 12. PCS plots for each bacterium and the mixed culture are shown individually in Figure 13. A clearer direction for each bacterium can be found in Figures 12 and 13. The chromatograms of milk inoculated with bacteria deviated from those of the negative control during prolonged storage time. In the cases of *E. aerogenes* and *L. lactis*, the samples with longer storage times (after 12 h) appeared farther away from the position of the reference (coefficient of determination = 1, slope = 1). All plots for *B. subtilis* were close to those of the reference. The mixed culture showed differences from milk samples inoculated with

Principal Component Similarity Analysis

J. Agric. Food Chem., Vol. 45, No. 3, 1997 **741**

Table 7. Factor Loadings (Pattern) from Principal Component Analysis of Data for *L. lactis, E. aerogenes, B. subtilis,* and a Mixed Culture

peak	principal component (PC)								
	1	2	3	4	5	6	7	8	9
1	0.831	−0.149	0.155	0.042	−0.180	0.167	−0.360	0.027	−0.004
2	0.113	0.330	−0.593	−0.205	−0.083	0.089	0.203	0.014	0.203
3	0.094	0.423	−0.657	−0.248	0.049	0.253	−0.133	−0.112	−0.121
4	0.122	0.461	−0.732	−0.259	0.059	0.147	−0.139	−0.011	−0.093
5	0.116	0.478	−0.790	−0.260	0.083	0.108	−0.113	0.010	−0.020
6	−0.029	−0.058	0.214	−0.058	−0.469	0.081	−0.312	0.136	0.362
7	0.379	0.750	0.383	0.135	−0.126	0.112	−0.138	0.037	0.010
8	0.114	0.574	−0.553	−0.209	0.104	0.197	−0.231	−0.072	−0.246
9	0.079	0.835	0.405	0.230	0.140	−0.110	0.083	0.054	0.105
10	0.645	−0.093	0.263	−0.674	0.124	−0.027	0.006	−0.083	0.003
11	0.724	−0.113	−0.162	0.379	−0.374	0.243	0.005	−0.001	−0.072
12	0.112	0.037	0.122	0.021	−0.580	0.163	−0.571	0.165	−0.042
13	0.087	0.932	−0.055	0.040	0.138	0.183	0.054	−0.034	0.001
14	0.529	0.537	0.579	−0.269	−0.001	0.043	−0.045	0.000	0.067
15	0.800	−0.135	0.251	−0.376	0.111	0.101	−0.065	−0.053	0.008
16	0.765	−0.127	−0.191	0.556	0.135	−0.005	−0.064	−0.023	0.095
17	0.524	0.071	−0.081	−0.028	−0.651	−0.070	0.301	0.114	−0.234
18	0.421	−0.109	−0.081	−0.160	−0.287	−0.184	0.039	−0.251	0.035
19	−0.363	−0.253	0.041	0.008	0.011	0.522	0.164	0.176	0.381
20	0.091	0.635	0.451	0.293	−0.237	0.088	0.116	−0.026	0.026
21	0.115	0.895	0.337	0.196	0.112	−0.012	0.087	0.006	0.067
22	0.563	−0.069	0.346	−0.723	0.103	0.038	0.007	−0.068	0.006
23	0.740	−0.189	−0.172	0.568	0.123	0.081	−0.045	−0.061	0.070
24	0.552	−0.088	0.322	−0.698	0.130	−0.008	0.066	−0.099	0.009
25	0.699	0.056	0.031	−0.491	−0.254	−0.231	0.191	0.125	0.040
26	0.394	−0.155	−0.076	0.391	0.116	−0.026	0.042	−0.248	−0.179
27	0.094	0.818	0.453	0.243	0.111	−0.041	0.113	0.015	0.105
28	0.016	0.048	−0.137	−0.031	−0.487	0.098	0.546	−0.335	−0.036
29	0.428	0.132	−0.584	0.142	−0.174	−0.117	0.269	0.077	0.179
30	0.742	−0.124	−0.322	0.499	0.147	−0.005	−0.030	0.014	0.088
31	0.208	0.232	0.124	0.246	−0.074	−0.593	−0.103	0.117	−0.451
32	0.184	0.299	−0.722	−0.194	−0.050	−0.217	0.035	0.235	0.188
33	0.331	0.018	−0.292	−0.102	0.106	−0.670	−0.070	0.414	0.257
34	0.917	−0.204	−0.023	0.126	0.200	0.062	0.016	−0.087	0.084
35	0.715	−0.204	0.043	0.074	0.210	0.013	0.095	0.110	−0.079
36	0.488	−0.219	0.048	0.159	0.055	0.285	0.207	0.362	−0.121
37	0.911	−0.227	0.042	0.131	0.189	0.068	−0.001	−0.116	0.071
38	0.039	−0.109	0.156	−0.133	0.155	0.401	0.245	0.650	−0.329

a single strain. Most plots of the mixed culture before 12 h were scattered around the reference. Between 12 and 24 h milk inoculated with the mixed culture produced different chromatograms. The best curve fitting equation was examined for *L. lactis, E. aerogenes,* and the mixed culture. The method was conducted as described for *P. fragi* and *P. fluorescens.* It was difficult to find the best curve fitting equation. For example, there was little difference between coefficients of determination in linear and polynomial equations in the case of the mixed culture. The equations which gave the best coefficient of determination are shown in Table 4. The coefficients of determination from equations for *L. lactis* and the mixed culture were 0.94 and 0.99, respectively. The coefficient of determination for *E. aerogenes* was 0.60. This indicates that it would be difficult to determine the direction of *E. aerogenes*.

The adjusted PC scores of each bacterium and the mixed culture are plotted in Figure 14. The behavior of the PC1 of *L. lactis* was quite different from the PC1 of the reference. The PC3 of *E. aerogenes* behaved differently from the PC scores of the reference. The adjusted PC scores of *B. subtilis* and the mixed culture were similar to those of the reference. The component loadings of the first nine PCs are shown in Table 7. Peaks 1, 15, 34, and 37 had high loadings (≥0.8) in PC1. These peaks are important in classifying milk samples.

CONCLUSIONS

A new multivariate analysis technique, principal component similarity analysis (PCS), was applied to peak areas in chromatograms from milk samples inoculated with several species of bacteria which cause off-flavors. The method was capable of detecting samples inoculated with selected bacterial species. Since PCS is an unsupervised method, it does not need preliminary grouping for the samples.

APPENDIX 1: PROCEDURE FOR PRINCIPAL COMPONENT SIMILARITY ANALYSIS (PCS) (VODOVOTZ ET AL., 1993)

Step 1. Apply PCA to the original data for *k* variables and *n* samples to derive *k* PC scores and *k* eigenvalues *E* and then steps 2−5 are followed for each sample.

$$P_i = E_i / \sum E_k$$

Step 2. Compute independent variables *V* as follows:

$$V_i = 100(1 - \sum P_j)$$

where *i* is the principal component number. (According to the rule of thumb, E_i lower than 1.0 can be discarded. However, it is recommended to include some smaller E_i, e.g., greater than 0.5, in the case of flavor analysis as it is not unusual that minor peak area variation plays an important role.)

Step 3. Compute dependent variables *Y* as follows:

$$Y_i = V_i + M(PC_i - PCq)$$

where *q* is the reference and *M* is percentage of variance of each principal component factor.

742 *J. Agric. Food Chem.,* Vol. 45, No. 3, 1997

Horimoto et al.

Step 4. Carry out linear regression analysis for *Y* versus *V* for each sample at each time to compute correlation coefficient (r^2) and slope (*S*).

Step 5. Plot *S* (slope) versus r^2.

LITERATURE CITED

Aishima, T.; Wilson, D. L.; Nakai, S. Application of simplex algorithm to flavour optimization based on pattern similarity of GC profile. In *Flavour Science and Technology*; Martens, M., Dallen, G. A., Russwurm, H., Eds.; John Wiley & Sons Ltd.: Chichester, U.K., 1987; pp 501−508.

Arteaga, G. E.; Luebsen, B. E.; Horimoto, Y.; Nakai, S. Prediction of cheese variety by applying artificial neural networks to GC data. *Abstracts of Papers*; Annual Meetings of the Institute of Food Technologists, Atlanta, GA; Institute of Food Technologists: Chicago, IL, 1994; 79C-11.

Badings, H. T.; Neeter, R. Recent advances in the study of aroma compounds of milk and dairy products. *Neth. Milk Dairy J.* **1980**, *34*, 9−30.

Bassette, R.; Bawdon, R. E.; Claydon, T. J. Production of volatile materials in milk by some species of bacteria. *J. Dairy Sci.* **1967**, *50*, 167−171.

Bawdon, R. E.; Bassette, R. Differentiation of *Escherichia coli* and *Aerobacter aerogenes* by gas liquid chromatography. *J. Dairy Sci.* **1966**, *49*, 624−627.

Cousin, M. A. Presence and activity of psychrotrophic micro-organisms in milk and dairy products: a review. *J. Food Prot.* **1982**, *45*, 172−207.

Dou, J.; Toma, S.; Nakai, S. Random-centroid optimization for food formulation. *Food Res. Int.* **1993**, *26*, 27−37.

Eyles, M. J.; Adams, R. F. Detection of microbial metabolites by gas chromatography in the examination of spoiled canned foods and related products. *Int. J. Food Microbiol.* **1986**, *3*, 321−330.

Furtula, V.; Nakai, S.; Amantea, G. F.; Laleye, L. Reverse-phase HPLC analysis of cheese samples aged by a fast-ripening process. *J. Food Sci.* **1994a**, *59*, 528−532, 567.

Furtula, V.; Nakai, S.; Amantea, G. F.; Laleye, L. Reverse-phase HPLC analysis of reference cheddar cheese samples for assessing accelerated cheese ripening. *J. Food Sci.* **1994b**, *59*, 533−538.

Girard, B.; Nakai, C. Static headspace gas chromatographic method for volatiles in canned salmon. *J. Food Sci.* **1991**, *56*, 1271−1274.

Griffiths, M. W.; Phillips, J. B. The application of the pre-incubation test in commercial dairies. *Aust. J. Dairy Technol.* **1986**, *41*, 71−79.

Guarino, P. A.; Krammer, A. Gas chromatographic analysis of head-space vapors to identify micro-organisms in foods. *J. Food Sci.* **1969**, *34*, 31−37.

Hosono, A.; Elliott, J. A.; Mcgugan, W. A. Production of ethylesters by some lactic acid and psychrotrophic bacteria. *J. Dairy Sci.* **1973**, *57*, 535−539.

Lee, M. L.; Smith, D. L.; Freeman, L. R. High-resolution gas chromatographic profiles of volatile organic compounds produced by micro-organisms at refrigerated temperatures. *Appl. Environ. Mirobiol.* **1979**, *37*, 85−90.

Lee, Y. Z.; Wong, R. M.; Nakai, S. Preparation of colorless globulin from bovine hemoglobin using sodium alginate. *J. Food Sci.* **1990**, *55*, 577−578.

Manly, R. F. J., Ed. Principal component analysis. In *Multivariate Statistical Methods*; *A Primer*; Chapman and Hall: London, 1986; Chapter 10.

Nakai, S. Computer-aided optimization with potential application in biorheology. *J. Jpn. Soc. Biorheology.* **1990**, *4*, 143−150.

Nakai, S.; Koide, K.; Eugster, K. A new mapping super-simplex optimization for food product and process development. *J. Food Sci.* **1984**, *49*, 1143−1148, 1170.

Pierami, R. M.; Stevenson, K. E. Detection of metabolites produced by psychrotrophic bacteria growing in milk. *J. Dairy Sci.* **1975**, *59*, 1010−1015.

Reddy, M. C.; Bills, D. O.; Lindsay, R. C.; Libbey, L. M. Ester production by *Pseudomonas fragi*. 1. Identification and quantification of some esters produced in milk cultures. *J. Dairy Sci.* **1967**, *51*, 656−659.

Schafer, M. L.; Peeler, J. T.; Bradshaw, J. G.; Hamilton, C. H.; Carver, R. B. A rapid gas chromatographic method for the identification of sporeformers and nonsporeformers in swollen cans of low-acid foods. *J. Food Sci.* **1982**, *47*, 2033−2037.

Skura, B. J.; Craig, C. and McKellar, R. C. Effect of disruption of an N$_2$-overlay on growth and proteinase production in milk by Pseudomonas fluorescens. *Can. Inst. Food Sci. Technol. J.* **1986**, *19*, 104−106.

SlideWrite Plus. Advanced Graphics Software, Inc., Sunnyvale, CA. 1995.

Speck, M. L., Ed. *Compendium of Methods for Microbiological Examination of Foods*; American Public Health Association, Inc.: Washington, DC, 1976.

Toan, T. T.; Bassette, R.; Claydon, T. J. Methyl sulfide production by *Aerobacter aerogenes* in milk. *J. Dairy Sci.* **1965**, *48*, 1174−1178.

Vallejo-Cordoba, B.; Nakai, S. Keeping-quality assessment of pasteurized milk by multivariate analysis of dynamic headspace gas chromatographic data. 1.Shelf life prediction by principal component regression. *J. Agric. Food Chem.* **1994**, *42*, 989−993.

Vodovotz, Y.; Arteaga, G. E.; Nakai, S. Principal component similarity analysis for classification and its application to GC data of mango. *Food Res. Int.* **1993**, *26*, 355−363.

Wellnitz-Ruen, W.; Reineccius, G. A.; Thomas, E. L. Analysis of the fruity off-flavour in milk using headspace concentration capillary column gas chromatography. *J. Agric. Food Chem.* **1982**, *30*, 512−514.

Wilkinson, L. *SYSTAT: The System for Statistics*, SYSTAT, Inc., Evanston, IL, 1990.