Morphometrical parameters of gut microflora in human volunteers

B. C. MEIJER*. G. J. KOOTSTRA AND M. H. F. WILKINSON

*Laboratory for Medical Microbiology, State University of Groningen,
Oostersingel 59, 9713 EZ Groningen, The Netherlands*

SUMMARY

The morphology of faecal microflora of nine healthy human volunteers was studied by digital image analysis of microscopic slides. Weekly specimens were collected during an 8-week period. Seven morphometrical parameters were derived: the means and medians of components 1, 2 and 3, and morphometrical entropy. Statistically significant differences among subjects were found for means of components 1 and 2, medians of components 1, 2 and 3, and entropy. The stability in normal circumstances provides an excellent basis for the detection of pathological change in gut flora balance.

INTRODUCTION

The microflora of the human digestive tract has been the subject of physiological and pathological study. Resistance to colonization and infection with intestinal pathogens depends on the integrity of the resident microflora [1, 2]; details of this colonization resistance have not yet been fully clarified. Some antibiotics have strong effects on gut flora; due to practical problems with existing methods the consequences of antibiotic therapy for colonization resistance cannot yet be predicted from routine measurements of patient specimens. Therefore, detailed study of gut microflora is needed.

Methods in use for studying gut microflora include aerobic and anaerobic culturing of faeces, and various chemical techniques aimed at detection in faeces of substances associated with integrity or otherwise of the flora [3–6]. In principle, the culturing methods can provide detailed insight in the microbial species present. In view of the large number of anaerobic species in normal gut flora (around 400; [7]) and the fact that most of them grow slowly and are difficult to identify, anaerobic culture is of limited practical use. Aerobic culture is less laborious but still takes between 2 and 4 days for full results on the aerobic segment of the microbial population [8]. The chemical Microflora Associated Characteristics are used primarily in research. So far, the only method in routine clinical use is – chiefly qualitative – aerobic culture.

Digital analysis of microscopic images (micromorphometry) has recently been used in environmental and medical microbiology [9–11] and shows promise for

* Present address: Regional Laboratory for Public Health, van Ketwich Verschuurlaan 92, 9721 SW Groningen, the Netherlands.

summarizing properties of complex microbial ecosystems [12]. Details of methods can be found in the textbooks by Gonzalez and Wintz [13] and Serra [14]. The methods employed in this study were described in more detail by Meijer and co-workers [15]. Interestingly, Baquero and colleagues [16] have found that on microscopic examination of faecal flora from antibiotic-treated subjects, morphological diversity was less than normal.

The present study is part of a research programme to find morphometrical parameters suitable for routine monitoring of faecal flora in patients. Sensitivity to disturbance of ecological balance, and stability in the absence of such disturbance are required of such parameters. To provide a basis for further work, normal variation of these parameters among humans and in time will have to be quantitated. Furthermore, data analysis should permit quantitation of morphological diversity.

MATERIALS AND METHODS

During the 8-week period, weekly faecal samples were collected from nine healthy human volunteers, who should not have taken antibiotics in the previous 4 weeks. From these samples, microscopic slides for phase contrast microscopy were prepared. First, 0.5 g of faeces were suspended in 4.5 ml water with 0.25% Tween 80. This suspension was centrifuged for 10 min at 15 g in order to eliminate debris; the supernatant was centrifuged again, now at 10000 g for 15 min, to pellet the bacteria. Bacteria were resuspended in water with 0.25% Tween 80, spread on a clean slide, and fixed with ethanol 70%. A drop of Tris-(hydroxymethyl)-methylamine (TRIS) was added and a cover-slip put on with nail polish. Slides were examined using an Olympus BH2 microscope with phase contrast optics (100 times magnification, N.A. 1.30 objective), and a Fairchild CCD 5000 video camera, connected to a PC/AT compatible microcomputer.

Details of image processing procedures have been given elsewhere [15]. A Matrox PIP-1024 digitizer with four $512 \times 512 \times 8$ bit image stores was used. After correction for pixel squareness the distance between adjacent pixels corresponded to $0.13\ \mu$m in the object. Segmentation of the image in objects and background was performed according to Kittler and co-workers [17]; the objects were analysed one by one. Four measurements were taken: surface area (A), perimeter (P), moment of inertia (I), and area of convex hull (H). The moment of inertia was computed from the second-order central moments in x and y (see [13], p. 356):

$$I = \mu_{20} + \mu_{02}$$

with

$$\mu_{20} = \sum_i (x_i - \bar{x})^2$$

and

$$\mu_{02} = \sum_i (y_i - \bar{y})^2$$

in which the x_i are the x coordinates of the pixels belonging to the object, and $\bar{x}$ is their mean; likewise for y_i and $\bar{y}$.

From these measurements, four variables were derived:

$$a = \log A,$$

the logarithm of the area;

$$f_1 = 2 \log P - \log A,$$

the first form factor;

$$f_2 = \log I - 2 \log A,$$

the second form factor;

$$c = \log H - \log A,$$

the concavity.

Natural logarithms were used. Next, coefficients obtained from a representative sample ($N = 1000$) of faecal bacteria were used to compute principal component (PC) score coefficients. This set of coefficients was kept unchanged during this study and will be used in further experiments. Three principal component scores (x_1, x_2 and x_3) were then computed for each bacterium. PC analysis is treated extensively in texts on multivariate statistics, for example by Mardia and colleagues [18]. It should be stressed that the procedure of principal components was used here with a more orthodox purpose than in our work on pure cultures [15]. Briefly, the transformation from original variables to principal components consists of an axis transformation: rotation and scaling. The transformation is chosen such that the coordinates of the measurement points on each of the PC axes (the principal component scores) are linearly uncorrelated. The new axes are at right angles to one another. The PC scores are linear combinations of the original variables; conversely, the original variables can be computed as linear combinations of the PC scores. In this computation, each PC imparts some variance to the original variables. It is said to 'account for' or 'explain' this amount of variance. The number of PCs that can be derived from a data set is at most equal to the number of variables. When the PCs that explain least variance are discarded, a model is left that fits the data optimally (in terms of variance accounted for) in as few dimensions as possible. In the work described here, PC analysis was used to reduce dimensionality and to obtain uncorrelated, standardized PC scores.

At least 500 bacteria were examined per specimen, and seven further statistical parameters were computed: sample means and medians of PCs 1, 2 and 3 (means 1, 2 and 3, and medians 1, 2 and 3), and morphometrical entropy, which measures the variety of bacterial form. It is computed by dividing each axis of the cube ($-5 \leqslant x_1 \leqslant 5$, $-5 \leqslant x_2 \leqslant 5$, $-5 \leqslant x_3 \leqslant 5$) in 16 parts, so that the cube is divided in 4096 smaller cubes. For each cube we compute

$$\hat{p}_i = \frac{n_i}{N},$$

the estimated probability that a bacterium from the specimen will yield a measurement point falling into that cube (the n_i are the numbers of bacteria in each cell; N is the total number of bacteria). The entropy is now estimated by

$$S = -\sum_i \hat{p}_i \log \hat{p}_i.$$

Analysis of variance was performed on four separate groups of variables: means of a, f_1, f_2 and c; medians of a, f_1, f_2 and c, and means and medians of the three

Table 1. *Principal component analysis*

Component	Variance explained	Percentage of total variance
1	3·21	80·1
2	0·54	13·4
3	0·21	5·4
4	0·04	1·1

Variables used: log area, form factor 1, form factor 2, concavity. Variables were standardized prior to principal component analysis.

Table 2. *Among-subjects and pooled within subject standard deviations of means, medians and entropy*; F- *and* P-*values*

	Within subject SD df = 49	Among subjects SD df = 8*	F	P
Mean log area	0·163	0·333	4·144	0·001
Mean log concavity	0·009	0·019	4·104	0·001
Mean log f_1	0·022	0·069	9·872	0·000
Mean log f_2	0·029	0·102	11·912	0·000
First discriminant function†	1	3·998	15·982	0·000
Mean log area	0·141	0·371	6·943	0·000
Mean log concavity	0·007	0·017	5·903	0·000
Mean log f_1	0·022	0·070	10·219	0·000
Mean log f_2	0·037	0·123	10·978	0·000
First discriminant function‡	1	3·399	11·553	0·000
Mean of PC 1	0·112	0·309	7·610	0·000
Mean of PC 2	0·218	0·337	2·391	0·029
Mean of PC 3	0·088	0·220	6·305	0·000
First discriminant function§	1	3·951	15·609	0·000
Median of PC 1	0·108	0·354	10·618	0·000
Median of PC 2	0·199	0·366	3·398	0·000
Median of PC 3	0·094	0·259	7·513	0·001
First discriminant function‖	1	4·633	21·461	0·000
Entropy	0·105	0·328	9·811	0·000

† Based on means of a, c, f and f.
‡ Based on medians of a, c, f and f.
§ Based on principal component means.
‖ Based on principal component medians.
* df: degrees of freedom.

components derived from those variables. As an indication of discriminating power we determined F-ratios for the first discriminant components found from each of the four groups. Homogeneity of within-subject variances was tested by Bartlett's method.

RESULTS

A total of 58 (80 %) of a possible 72 faecal specimens was received for analysis. One subject provided 8 specimens, 3 provided 7, 4 provided 6, and 1 provided 5.

Table 1 presents the results of principal component analysis on a representative sample of faecal bacteria ($N = 1000$). Most of the original variance is explained by
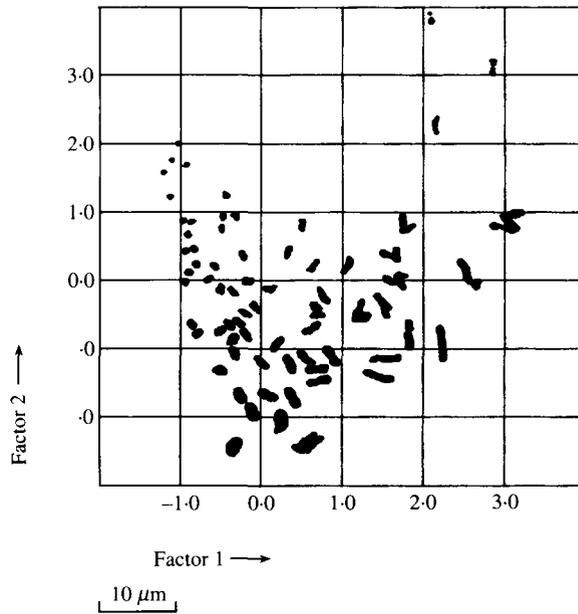
Fig. 1. Subject 7: modified scatter plot of bacteria from faecal flora: first two prin-
cipal components.
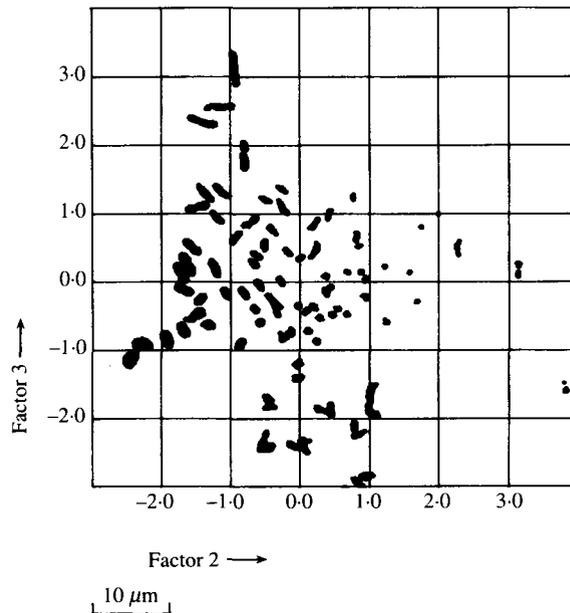
Fig. 2. Subject 7: modified scatter plot of bacteria from faecal flora: principal com-
ponents 2 and 3.

the first three components. We decided to discard the fourth. Figs. 1 and 2 are
examples of modified scatter plots in which, instead of points or symbols, the
bacteria themselves have been drawn on the spot corresponding to their principal
component scores. Such plots provide a sorted overview of the morphology of
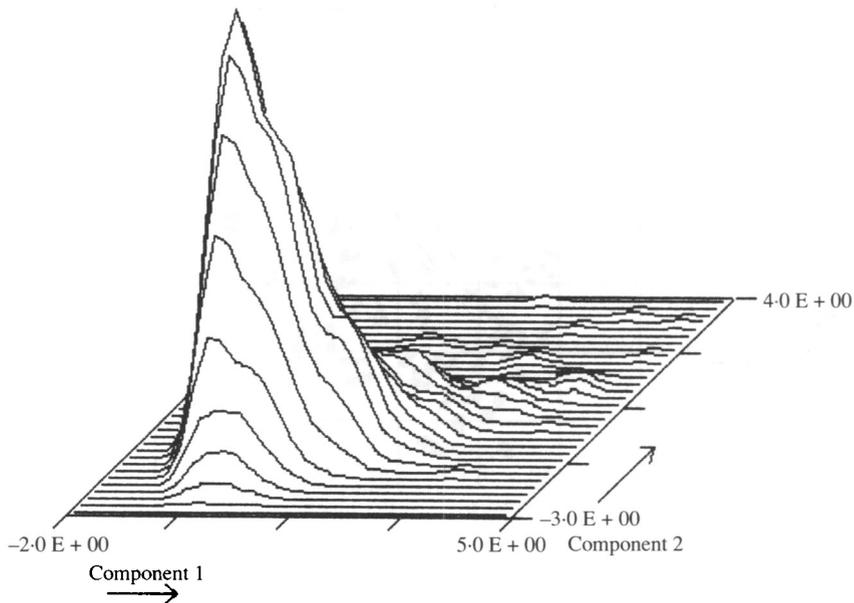
Fig. 3. Subject 7: hidden line plot of estimated density of bacteria.
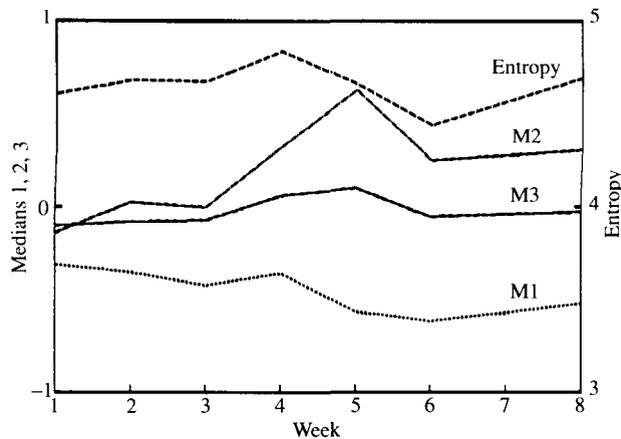
Fig. 4. Subject 1: medians of principal components 1, 2 and 3, and entropy as functions of time.

bacteria present in the flora. In addition, they show the morphological aspects on which the three components depend: elongatedness, size, and convexity, respectively. Fig. 3 is a 'hidden line' graph in which the estimated density of bacteria is plotted against the first two principal components. In Figs. 4 and 5 the three medians and entropy are plotted against time for two typical subjects. Table 2 summarizes the analysis of variance for morphometrical difference between subjects. The F-values are obtained by dividing variance among subjects (signal) by pooled variance within subjects (noise) and indicate the usefulness of a variable. In this sense, the median and the mean of component 1, and the median and mean of form factors f_1 and f_2, and entropy, are the most useful. Discriminant
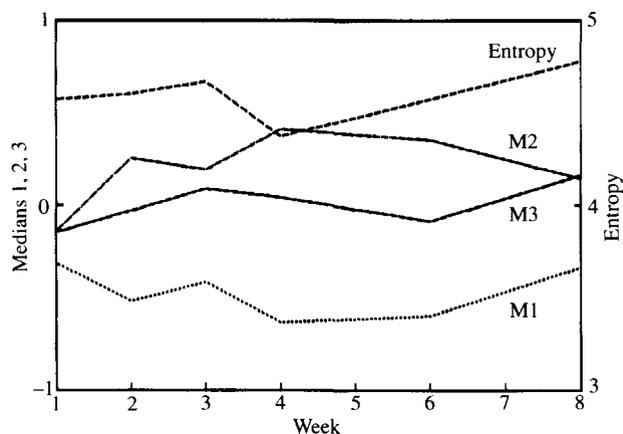
Fig. 5. Subject 9: medians of principal components 1, 2 and 3, and entropy as functions of time.

functions are linear combinations of variables, computed with the aim of maximizing the F-ratio for this combination. The F-value associated with the first discriminant function derived from a set of variables gives an indication of the discriminatory power in that set. The means of the first three principal components discriminate only slightly worse than the means of the original variables. The tests for homogeneity of variance yielded non-significant results, except for the median of component 1 $(0.1\% < p < 0.5\%)$.

DISCUSSION

The four measurements: log area, form factors 1 and 2, and concavity, were chosen from a total of 12, including perimeter, size by repeated erosion, skeleton length, and projection lengths on the X and Y axes [14]. For the variables eliminated, it proved possible either to predict them accurately (by linear regression) from those retained or to establish that remaining inaccuracy was due to measurement errors in the eliminated variable. The results of principal-components analysis show that one more variable could plausibly have been eliminated: the fourth component explains very little variance.

Subsequent transformation from four variables to three PCs was done to reduce dimensionality. In the present experiment, no loss of discriminatory power was found. On the contrary, when sample medians of both original viables and PCs were used to distinguish between subjects, the latter gave better results (Table 2). At first sight this is a strange effect. It can only be explained when the non-linearity of the operation of taking the median, and the non-normality of the distributions involved, are taken into account.

The choice between means or medians to characterize a sample is not straightforward. It depends on how one assumes change in morphometrical distribution is caused by e.g. antibiotic treatment. As stated earlier [15], each species in the flora may have its morphology characterized by a species mean and covariance. In our model, overall changes in flora morphology are primarily caused by variation in relative numbers of several species, and less by shifts in the

morphological characteristics of the new species themselves. The overall mean of a mixed population depends linearly on the means of constituent species. This linearity may be advantageous when we correlate flora variations, documented by culturing, with results of morphometry. The median has no such property, but sample medians are more robust against outliers than are means, and so will provide a better signal-to-noise ratio. In this experiment, therefore, both means and medians have been determined, so that reference values of both will be available for future work.

The morphometrical parameters chosen are stable in time, and temporal variation has been quantitated. This allows differences between subjects to be distinguished. Furthermore, it means that, given one specimen of normal gut flora from an individual, a 'normal' region can be constructed for the morphometrical parameters of future specimens from this individual. If later samples yield parameter values outside this region, the gut flora is seen to be disturbed. Strictly, median 1 cannot be used for this purpose, because it does not have the same degree of variation in time for every individual.

The morphological composition of human gut flora shows differences between individuals. Our method can be used to document such differences, and this estimate of normal variation between humans will be used in further work on the influence of antibiotics, diet, and disease on flora morphology. The method would, however, not suffice to identify single humans from large groups of candidates.

REFERENCES

1. Van der Waaij D. Colonization resistance of the digestive tract – mechanism and clinical consequences. Nahrung 1987; **31**: 507–17.
2. Gustafsson BE. The physiological importance of the colonic microflora. Scand J Gastroenterol 1982; **77**: 117–31.
3. Meijer-Severs GJ, Van Santen E, Meijer BC. Short-chain fatty acid and organic acid concentrations in feces of healthy human volunteers and their correlation with anaerobe cultural counts during systemic ceftriaxone administration. Scand J Gastroenterol 1990; **25**: 698-704.
4. Meijer-Severs GJ, Van Santen E. Short-chain fatty acid and organic acid concentrations in feces of 10 human volunteers and their correlation with anaerobe cultural counts over a 15-month period. Scand J Gastroenterol 1989; **24**: 1276–80.
5. Welling GW, Stege GJJ, Meijer-Severs GJ. β-Aspartylpeptidase activity, a microflora associated characteristic (MAC); its presence in different strains of anaerobic bacteria. Microb Ecol Health Dis 1987; **1**: 45–9.
6. Leijonmark CE, Bergstrand O, Carlstedt-Duke B. et al. Total parenteral nutrition and the function of the intestinal microflora in Crohn's disease. Scand J Gastroenterol 1988; **23**: 59–64.
7. Moore WEC, Holdeman LV. Human faecal flora: the normal flora of 20 Japanese Hawaiians. Appl Environ Microbiol 1974; **27**: 961–79.
8. Apperloo-Renkema HZ, van der Waaij BD, van der Waaij, D. Determination of colonization resistance of the digestive tract by biotyping of Enterobacteriaceae. Epidemiol Infect 1990; **105**: 355–61.

9 Getliff JM Fry JC Using the Solitaire Plus image analyser for direct estimates of bacterial volume Binary Comp Microbiol 1990 **2** 55–7

10 Krambeck C Krambeck H Schroder D Newell SY Sizing bacterioplankton a juxtaposition of bias due to shrinkage halos subjectivity in image interpretation and asymmetric distributions Binary Comp Microbiol 1990 **2** 5–14

11 Thomas C Packer H Morphological and structural measurements on filamentous microorganisms Binary Comp Microbiol 1990 **2** 47 54

12 Wynn Williams DD The application of image analysis to natural terrestrial ecosystems Binary Comp Microbiol 1990 **2** 15 20

13 Gonzalez RC Wintz P Digital image processing Reading Massachusetts Addison Wesley 1977

14 Serra J Image analysis and mathematical morphology London Academic Press 1982

15 Meijer BC Kootstra GJ Wilkinson MHF A theoretical and practical investigation into the characterization of bacterial species by image analysis Binary Comp Microbiol 1990 **2** 21 31

16 Baquero F Fernandez Jorge A Vicente MF Alos JI Reig M Diversity analysis of the human intestinal flora a simple method based on bacterial morphotypes Microb Ecol Health Dis 1988 **1** 101–8

17 Kittler J Illingworth J Foglein J Threshold selection based on a simple image statistic Comp Vision Graph Image Proc 1985 **30** 125–47

18 Mardia KV Kent JT Bibby JM Multivariate analysis London Academic Press 1979