Metabolomics, Vol. 3, No. 4, December 2007 (© 2007)
DOI: 10.1007/s11306-007-0054-6

Comparison of human axillary odour profiles obtained by gas chromatography/mass spectrometry and skin microbial profiles obtained by denaturing gradient gel electrophoresis using multivariate pattern recognition

Yun Xu,[a] Sarah J. Dixon,[a] Richard G. Brereton,[a,*] Helena A. Soini,[b] Milos V. Novotny,[b] Karlheinz Trebesius,[c] Ingrid Bergmaier,[c] Elisabeth Oberzaucher,[d] Karl Grammer,[d] and Dustin J. Penn[e]

[a]*Centre for Chemometrics, School of Chemistry, University of Bristol, Cantocks Close, Bristol, BS8 1TS, UK*
[b]*Institute for Pheromone Research and Department of Chemistry, Indiana University, 800 E. Kirkwood Ave, Bloomington, IN, 47405, USA*
[c]*Vermicon AG, Emmy-Noether-Str. 2, 80992, Munich, Germany*
[d]*Department for Anthropology, Ludwig Boltzmann Institute for Urban Ethology, Althanstrasse 14, A-1090, Vienna, Austria*
[e]*Konrad Lorenz Institute for Ethology, Austrian Academy of Sciences, Savoyenstr. 1a, A-1160, Vienna, Austria*

Received 19 November 2006; Accepted 22 March 2007

Several studies have shown that microbial action is responsible for many compounds responsible for human odour. In this paper, we compare the pattern of microbial profiles and that of chemical profiles of human axillary odour by using multivariate pattern matching techniques. Approximately 200 subjects from Carinthia, Austria, participated in the study. The microbial profiles were represented by denaturing gradient gel electrophoresis (DGGE) analysis and the axillary odour profiles were determined in the sweat samples collected by a stir-bar sampling device and analysed by gas chromatography/mass spectrometry (GC/MS). Both qualitative and quantitative distance metrics were used to construct dissimilarity matrices between samples which were then used to represent the patterns of these two types of profiles. The distance matrices were then compared by using the Mantel test and the Procrustean test. The results show that on the overall dataset there is no strong correlation between microbial and chemical profiles. When the data are split into family groups, correlations vary according to family with a range of estimated p values from 0.00 to 0.90 that the null hypothesis (no correlation) holds. When 32 subjects who followed four basic rules of behaviour were selected, the estimated p-values are 0.00 using qualitative and < 0.01 using quantitative distance metrics, suggesting excellent evidence that there is a connection between the microbial and chemical signature.

KEY WORDS: multivariate pattern comparison; human odour profile; human microbial profile; gas chromatography/mass spectrometry; denaturing gradient gel electrophoresis.

1. Introduction

It is well known that axillary microflora play very important roles in human odour production. Previous studies have shown that the population density of certain micro-organisms has strong association with the intensity of odour (Rennie *et al*., 1990, 1991). Certain areas of the human body possess a unique odour which is produced partially by microbial action, especially the human axillary odour is a combination of secretions from eccrine, apocrine, apoeccrine and sebaceous glands together with the local microflora (Sastry *et al*., 1980; Sato *et al*., 1987). Studies have also shown that axillary odour is absent if Gram-positive bacteria are eliminated from the skin (Marples, 1969; Gower *et al*., 1985).

In this study, we compare the pattern of microbial profiles obtained by denaturing gradient gel electrophoresis (DGGE) with the pattern of chemical human axillary odour obtained by gas chromatography/mass spectrometry (GC/MS) and explore the common trends between these two types of profile. Samples were taken from approximately 200 test subjects. Most subjects were sampled twice for the microbial signature, and five times for the chemical signature, each sample taken over different fortnights. Our previous studies have shown that there is strong statistical evidence that both in GC/MS and DGGE data, reproducible signals exist for individual characterisation (Penn *et al*., 2007; Xu *et al*., in preparation). In this paper, pair-wise dissimilarity matrices (both qualitative and quantitative distance metrics) were calculated on both GC/MS and DGGE profiles to determine the patterns of these two types of data. The similarity of these two patterns was evaluated by the permuted Mantel test. We also transformed the dissimilarity matrices to sample-latent variable matrices using Principal Co-ordinates Analysis and then compared the patterns of these two matrices by using a test based on Procrustes analysis.

* To whom correspondence should be addressed.
E-mail: r.g.brereton@bris.ac.uk.

2. Experimental

2.1. Test subjects

A total number of 196 subjects participated in this study. These individuals are from an isolated population in Carinthia, Southern Austria, belonging to 17 different families, coded by a single letter (e.g. A). The samples were taken from July to August, 2005, the odour samples were collected five times for each subject, once per fortnight while microbial samples were collected twice, (on the third and fourth fortnights). In this paper, we only consider the fortnights when the microbial and odour samples were collected simultaneously. The subjects were also asked to fill in a survey when the samples were taken which contains information about their living habits, such as when was the last washing of armpit, and when was the last usage of deodorant as discussed in greater detail in Section 4. Due to experimental reasons, some subjects only provided one sample over fortnights 3 and 4 and they were excluded from the analysis, making the number of subjects reported in this paper less than the full population. A total of $P = 177$ subjects are analysed in this study, each sampled twice to give $2P$ samples, each with a corresponding GC/MS and microbial profile.

2.2. Human axillary odour samples collection and GC/MS analysis

The volatile and semivolatile compounds from armpits were collected by using a newly designed skin roller device. A stir bar is attached to the device and rolled over the skin to collect chemical compounds. The stir bar was then subjected to the GC/MS analysis. Stir bar sampling on the skin surface is likely to favour compounds which bind in the oily layer present on the human skin: in fact, the relatively long storage stability time of 20 days under cooled conditions supports the hypothesis. Compounds with high volatility and of very hydrophilic characteristics would be discriminated against using this sampling approach. During the same day, reproducibility for typical skin compounds taken from five individuals was 3–25% (RSD, relative standard deviation, three samples per individual). The long term RSD of the internal standard is 14.3%. More detailed information about the characteristics and the reproducibility of both the sampling device and the GC/MS experiments are given elsewhere (Soini *et al.*, 2006).

The GC equipment for quantitative analysis consisted of an Agilent 6890N gas chromatograph connected to a 5973i MSD mass spectrometer (Agilent Technologies, Palo Alto, CA) with a Thermal Desorption Autosampler (TDSA, Gerstel, Mülheim an der Ruhr). Positive electron ionisation (EI) mode at 70 eV was used with a scanning rate of 4.51 scans/s over the mass range of m/z 35–350. The ion source and quadrupole temperatures were set at 230 °C and 150 °C, respectively. The separation capillary was DB-5MS (20 m×0.18 mm, i.d., 0.18 μm film thickness) from Agilent Technologies (Wilmington, DE). Samples were thermally desorbed in a TDSA automated system, followed by injection into the column with a Cooled Injection System, CIS-4. The TDSA operated in a splitless mode. The temperature program for desorption was 20 °C (0.5 min), then 60 °C/min to 250 °C (3 min). The temperature of the transfer line was set at 280 °C. The CIS was cooled with liquid nitrogen to −80 °C. After desorption and cryotrapping, the CIS was heated at 12 °C/s to 280 °C with the hold time of 10 min. The CIS inlet was operated in the solvent vent mode, with a vent pressure of 14 psi, a vent flow of 50 mL/min, and a purge flow of 50 mL/min. The temperature program in the GC operation was 50 °C for 1 min, then increasing to 160 °C at the rate of 5 °C/min, followed by the second ramp at the rate of 3 °C/min to 200 °C (hold time 16 min). The carrier gas head pressure was 14 psi (flow rate, 0.7 mL/min at constant flow mode). The GC temperature program lasted for 52.33 min, with mass spectrometric detection commencing after a deadtime of 1.88 min (solvent delay). To increase throughput, two instruments of identical specifications were used to analyse the samples. The configuration of both instruments was the same, and tests had been done to ensure reference samples analysed on each instrument were of acceptable similarity. There was a slight difference in scan rates between the two instruments, with instrument 1 sampling 13,481 scans over the analysis period, and instrument 2 sampling 13,460 due to slightly different software versions. In the mass spectrometry software (Agilent ChemStation) a parameter is set, below which a scan will be recorded as having zero intensity. This is the detection limit of the instrument, and can be set by the instrument operator. In this work, it was set to 300: this number is essentially arbitrary and will depend on the specific instrument, but in the context of the current study the smallest peaks which were visible above the noise were of height around 600 units. Typical peak heights in a mass channel range from 1000 to 100,000 (varying from sample to sample). The mass resolution was reduced to unity before analysis. During the analysis, test runs using quality control samples were performed on a regular basis to ensure that the instruments were performing acceptably.

2.3. Microbial sampling and DGGE analysis

2.3.1. Fixation of samples

Microbial samples were taken in Greifenburg (Carinthia, Austria) from the armpit of different subjects. Sampling of the axillary microflora of the armpit was performed by the washing-scrub method of Williamson and Kligman (1965). A plastic cylinder open at both ends was placed on the armpit. About 1.5 mL detergent solution was filled into the cylinder. A glass stirrer was moved with constant pressure over the skin to detach the microorganisms. The solution was removed and transferred to

a sterilised reaction tube and the procedure was repeated. The obtained solutions were fixed with ethanol at a ratio 1:1. The total sample volume of 6 mL consisted of 3 mL sampling buffer containing the microbes (sample) and 3 mL of 96% ethanol.

2.3.2. DNA extraction

About 1.3 mL of the sample was centrifuged (10 min, 14,000 rpm) and the supernatant was discarded. This step was repeated once, by adding 1.3 mL sample to the pellet and an additional centrifugation step (10 min, 14,000 rpm). The obtained pellet was washed in 200 μL 1× phosphate-buffered saline (PBS). After centrifugation (10 min, 14,000 rpm) the pellet was resuspended in 100 μL 6% Chelex® 100 solution (BioRad, Munich, Germany) according to Rodrìguez-Lázaro *et al.* (2004), and incubated at 56 °C for 20 min. The sample was then thoroughly mixed and incubated further at 100 °C for 8 min. Subsequently, the sample was mixed and cooled for 5 min on ice. Next a centrifugation step (10 min, 14,000 rpm) was performed. The supernatant containing the DNA was removed.

2.3.3. PCR amplification of target DNA for DGGE

The extracted genomic DNA was amplified using the forward primer 341F-GC (Muyzer *et al.*, 1993) with a GC clamp 5′-CGC CCG CCG CGC GCG GCG GGC GGG GCG GGG GCA CGG GGG GCC TAC GGG AGG CAG CAG -3′ and the reverse primer 518R 5′-ATT ACC GCG GCT GCT GG-3′. The final 50 μL reaction mixture contained: 2 μL template DNA, 25 pmol primers each, 2.5 U of *Taq* DNA polymerase (Promega, Mannheim, Germany), 1-fold PCR buffer (Promega), 75 mM $MgCl_2$ (Promega), 10 mM dNTPs, (Promega). The PCR protocol included a 5 min initial denaturation at 94 °C, 30 cycles of 94 °C for 0.5 min, 44 °C for 1 min, 72 °C for 1.5 min followed by 10 min at 72 °C for final extension in a Primus 96 thermocycler (MWG, Ebersberg, Germany). PCR products were stored at −20°C until further use.

2.3.4. DGGE analysis

DGGE analysis was performed on DcodeTM-System (Bio-Rad). Samples were loaded onto a 8% (w/v) acrylamide gel (37.5:1 acrylamide-bisacrylamide) in 1× TAE buffer with a denaturant gradient ranging from 20% to 60% prepared in accordance with Muyzer *et al.* (1995). (100% denaturant contains 7 M urea and volume ratio of 40% formamide.)

To standardise DGGE gels, reference standards were applied to each gel. The reference standard consisted of a mixture of PCR products of 11 different bacterial species which are commonly found in human skin samples (Trebesius *et al.*, in preparation). The banding pattern resulted from PCR products obtained by the same primer pair as described above.

The electrophoresis was performed at 60 °C, initially at 25 V for 15 min following at 130 V for 4 h. The gel was silver stained based on Sanguinetti *et al.* (1994) by the following procedure. About 150 mL of fixing solution (10% ethanol, 0.5% acetic acid) was applied to the gel and shook gently for 3 min. Subsequently, the gel was incubated for 10 min at room temperature in a silver nitrate solution (0.2% $AgNO_3$, 10% ethanol, 0.5% acetic acid). After discharging the silver nitrate solution a washing step in distilled H_2O for 2 min followed. Hereafter, 150 mL of "developer solution" was applied to the gel (3% NaOH containing, 300 μL of 37% formaldehyde) for 5 min, while shaking gently. The staining procedure was stopped by incubating the gel for 5 min in 10% ethanol, 0.5% acetic acid solution.

A typical example of a DGGE gel is presented in figure 1.

2.4. Software

The GC/MS instrument was controlled by Agilent ChemStation software versions D.01.00 (system 1) and D.01.02 (system 2). The GC/MS data was exported to AIA/netCDF (network Common Data Format) format and then imported into MATLAB (The Mathworks, Inc., Natick, MA) using a freely available conversion tool (available at http:// mexcdf.sourceforge.net/). The stained gel was transferred on an overhead transparency sheet and documented on a SnapScan 1236 scanner (Agfa, Ridgefield Park, NJ). The scanning mode was transparent, 300 dpi and 24 Bit colour. The resultant pictures were converted to TIFF files for processing and imported into MATLAB by using Image Processing Toolbox. All data processing was performed using MATLAB version 7.0.4.365, Release 14, Service Pack 2.

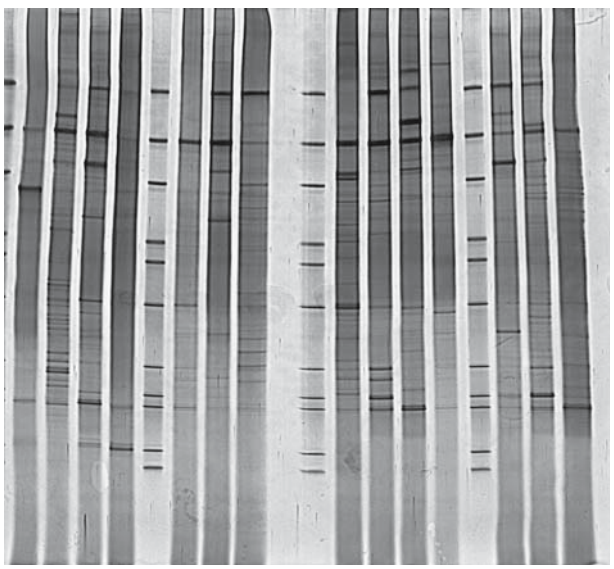

Figure 1. A typical DGGE gel.

3. Data analysis methods

3.1. Data preparation

3.1.1. DGGE data preparation

The images of scanned DGGE gels were processed by an in-house image digitisation software. For each lane (sample) on the gel, the bands were detected and their positions recorded. The position of each band was then corrected with respect to the positions of reference standards to cope with different separation behaviour between gels. The detailed description is given in a separate paper (Xu *et al.*, in preparation) and not repeated here for brevity.

The main difficulty in this study is the band assignment. DGGE is a 1-D technique hence the only useful information for the band assignment is the position information and there is no spectroscopic or other information which can help in identifying which band is which. When there were many lanes and gels, it is often hard to assign some bands unambiguously, i.e. when two bands come from different lanes/gels with slightly different positions, it is difficult to decide whether these two bands originate from the same micro-organism or two different ones. However, data analysis methods such as PCA (Wold *et al.*, 1987; Brereton, 2003) based on a sample-feature matrix require each variable to originate from the same source, otherwise the results will be influenced by the incorrect assignments of variables. To overcome this problem, we used a pair-wise dissimilarity matrix between sample profiles rather than a full sample-variable matrix since it is much easier to measure the dissimilarity between two samples as the number of ambiguous assignment of the bands is small when comparing two lanes, rather than using global band assignments. Previously we developed a fuzzy distance metric to measure this dissimilarity. The metric is weighted by a fuzzy function and the value of this function depends on the difference of the positions of two bands, so it does not require an accurate assignment of each band and tolerates slight imprecision of the positional information. We briefly summarise the method here which will be described elsewhere (Xu *et al.*, in preparation): in this paper we use the square root of the intensities rather than the raw intensities to be comparable to the GC/MS data, otherwise the method is identical.

The first step is to correct the position of bands in each lane according to a reference lane consisting of extracts from 11 microbes, the corrected position relates to how close the bands are to the microbial extracts in the nearest reference lane, for example a band that is detected half way between bands 3 and 4 in the reference lane will be given a value of 3.5. These positions relative to the reference bands are then used for subsequent analysis as follows.

The next step is to determine a dissimilarity measure between the reference band corrected profiles in each lane. Suppose lane i has N_i bands and lane j has N_j bands, and also assume $N_i \leq N_j$. A $N_i \times N_j$ position difference matrix is constructed; each row representing the positional difference of one band in lane i to all the bands in lane j. The minimum value in each row will be the nearest band. For each dissimilarity metric calculation, there can be N_i comparisons at most. However, sometimes two or more bands in lane i share the same nearest neighbour in lane j. In such case, we only consider the closest pair and the others will be discarded (i.e. consider them as unique bands in lane i). We denote the number of band pairs k that have been identified as b ($b \leq N_i$).

Both quantitative and qualitative (presence/absence criterion) fuzzy weighted distance metrics are used to measure the dissimilarity between two DGGE samples. The quantitative metric takes into account band intensities as well as uncertainties in position using the following equation:

$$d(i,j) = 1 - \frac{\sum_{k=1}^{b} \left[\sqrt{x_{ik}.x_{jk}} \cdot w_k\right]}{\sqrt{\sum_{n=1}^{N_i} x_{in} \sum_{n=1}^{N_j} x_{jn}}} \tag{1}$$

where b is the number of pairs being considered (see above) each matched pair being denoted by k, where x_{ik} and x_{jk} are the integrated intensities of the band pair k in lanes i and j. The qualitative distance metric is the square root of the fuzzy weighted Jaccard distance (Jaccard, 1908) metric defined as

$$d(i,j) = \sqrt{1 - \frac{\sum_{k=1}^{b} w_k}{\sum_{k=1}^{p} w_k + N_i + N_j - b - \sum_{k=1}^{p} w_k}} = \sqrt{1 - \frac{\sum_{k=1}^{b} w_k}{N_i + N_j - b}} \tag{2}$$

The fuzzy weight function $w_k = H(\Delta d_k)$ is determined by the absolute difference in the corrected position (Δd_k) of the band pair k ($=1$ to p) defined by

$$w_k = H(\Delta d_k) = \frac{1}{2} \cdot erfc(A \cdot \Delta d_k - S) \tag{3}$$

where function $erfc(x)$ is the complementary error function (Zwillinger, 1997), defined by $efrc(x) = \frac{2}{\pi} \cdot \int_x^{\infty} e^{-t^2} dt$ and t is an integration factor, from x to $+\infty$. S and A are two adjustable free parameters which control the sensitivity to the positional difference of the

weight function. With an appropriate setting of S and A this function has the following properties.

(a) When the difference is small enough, the output is 1 or very close to 1,
(b) when the difference is moderate, the output is between 0 and 1 and decays exponentially with increasing Δd,
(c) when the difference is large enough, the output is 0 or very close to 0.

In this paper, S is set to 5 and A to 30. If the corrected positional difference is closer than about 0.1 they are assumed to be a perfect match, if differing by more than 0.2 they are assumed not to match at all.

3.1.2. GC/MS data preparation

A peak table was constructed based on all the GC/MS total ion chromatograms which is a matrix whose rows correspond to samples and whose columns correspond to summed peak areas over all significant masses of the corresponding compounds. The method for peak identification and alignment has been described in detail in Dixon *et al.* (2006). The peaks corresponding to known background compounds such as siloxanes originating from stir bars or vial septa were identified based on their mass spectra and removed from the peak table. Next, peaks that occur in 4 or less samples in the overall dataset are removed as they are unlikely to have any diagnostic significance. A further reduction in the size of the peak table is performed to consider only those peaks that are detected in at least 4 out of 5 fortnights in at least one subject leaving $N_{GCMS} = 373$ peaks in total over all subjects: this is so as to retain peaks that are likely to relate to stable biological signatures which are detected in the majority of samples from one or more individuals. Finally peak areas were scaled as follows: the square root of the peak areas were computed, these peak areas were summed to a constant total of 1 in each chromatogram. The reason we square root peak areas is that there can be large peaks in some chromatograms that could dominate the profile, hence distorting data after normalisation; a common alternative of log scaling is not suitable in this study because many peaks are not detected in chromatograms so a large number of zero numbers would need to be replaced, square root scaling is a common alternative for reducing contrasts. More detailed discussion about these data transformation methods can be found elsewhere (Box and Cox, 1964; Huber *et al.*, 2002).

In order to compare the patterns of the GC/MS and DGGE data, we constructed two pair-wise distance matrices for each data set, one using qualitative distance metric and another using quantitative distance metric. For the quantitative distance between GC/MS i and j we use

$$d(i,j) = 1 - \frac{\sum_{k=1}^{N_{GCMS}} [(x_{ik} \cdot x_{jk})]}{\|\boldsymbol{x_i}\| \cdot \|\boldsymbol{x_j}\|} \quad (4)$$

where x_{ik} is the normalised square root intensity of peak k in chromatogram i and $\boldsymbol{x_i}$ is the vector containing all N_{GCMS} (=373) intensities of this chromatogram. For the qualitative distance we use the square root of the Jaccard distance (Jaccard, 1908)

$$d(i,j) = \sqrt{1 - \frac{a}{a+b+c}} \quad (5)$$

where a are the number of peaks that are common to both samples, and b and c the number of peaks unique to one of the two samples.

The reason we square root the distances will be explained in Section 3.2.2.

3.2. Comparison of multivariate patterns

3.2.1. Mantel test

To compare two distance matrices directly, the Mantel test (Mantel, 1967) is a commonly used and effective method. Given two distance matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with dimensions of $N \times N$, the lower (or upper) triangular parts of the matrices are unfolded into two vectors: $\boldsymbol{a} = \{a_{11}, a_{21}, \ldots a_{N1}, a_{22}, \ldots a_{N2}, \ldots, a_{NN}\}$ and $\boldsymbol{b} = \{b_{11}, b_{21}, \ldots b_{N1}, b_{22}, \ldots b_{N2}, \ldots, b_{NN}\}$ where a_{mn} and b_{mn} are the mth row and nth column of the elements in matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. The correlation coefficient r between $\boldsymbol{a}$ and $\boldsymbol{b}$ is then calculated.

The significance of the correlation is determined using a Monte Carlo simulation. The order of the elements in one vector is permuted while the other remains unchanged and the correlation coefficient between these two vectors is calculated. This process is repeated sufficiently large number of times and the correlation coefficients obtained by these permutations are used to form an empirical null distribution. The significance of the correlation coefficient r calculated above is computed to provide an indication as to whether this belongs to the null distribution or not. For an estimate of the significance of the correlation coefficient we can calculate the proportion of times the Monte Carlo simulation results in a correlation coefficient greater than the observed value r (one tailed test) or the absolute value is greater than r (two tailed test). These approximate to probabilities and are denoted by p values below. In this paper we use the two sided Mantel test.

3.2.2. Multidimensional scaling and the Procrustean test

A major drawback of the Mantel test is that it is not possible to visualise the data structure as it is based on a

single number (the correlation coefficient), between two sets of data. Multidimensional scaling (MDS) is a set of methods that can transform the pattern represented by a pair-wise distance matrix back to a matrix of samples and latent variables. MDS constructs a new sample-latent variable matrix and the relative configuration of the data points has been preserved and can be visualised, if appropriate, by using 2-D or 3-D scatter plots. We used a common MDS method called Principal Coordinate Analysis (PCO) (Gower, 1966) to transform the distance matrices to sample-feature matrices. A brief description of PCO is given below.

Given an $N \times N$ symmetric distance matrix $\boldsymbol{D}$, each element d_{ij} in $\boldsymbol{D}$ represents the distance from sample i to sample j.

(1)

$$\boldsymbol{A} = (a_{ij}) = \left(-\frac{1}{2} \cdot d_{ij}^2\right) \tag{6}$$

(2) Calculate Gower's centred matrix $\boldsymbol{G}$ using

$$\boldsymbol{G} = \left(\boldsymbol{I} - \frac{1}{N}\boldsymbol{1} \cdot \boldsymbol{1}'\right) \cdot \boldsymbol{A} \cdot \left(\boldsymbol{I} - \frac{1}{N}\boldsymbol{1} \cdot \boldsymbol{1}'\right) \tag{7}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{1}$ is a column of 1s of length N.

(3) Perform eigen-decomposition on matrix $\boldsymbol{G}$ so that

$$\boldsymbol{G} = \boldsymbol{Q} \cdot \Lambda \cdot \boldsymbol{Q}' \tag{8}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with eigenvalues on the diagonal in descending order. $\boldsymbol{Q}$ is a square matrix whose columns are corresponding orthonormal eigenvectors. It is important to point out that the eigenvalues and eigenvectors here are defined by linear algebra: each pair of eigenvalues λ and corresponding eigenvectors $\boldsymbol{q}$ of the square matrix $\boldsymbol{G}$ satisfies $\boldsymbol{G} \cdot \boldsymbol{q} = \lambda \cdot \boldsymbol{q}$. The definition of an eigenvalue often used in the chemical PCA literature (Wold *et al.*, 1987; Brereton, 2003), the sum of squares of scores, has a similar property to the eigenvalues as defined above: however unlike eigenvalues in PCA literature, linear algebra defined eigenvalues cannot be guaranteed to be non-negative for all types of matrices. A matrix $\boldsymbol{G}$ having all diagonal elements of $\boldsymbol{\Lambda}$ positive is called positive definite ; if one or a few diagonal elements of $\boldsymbol{\Lambda}$ are 0 and others are all positive $\boldsymbol{G}$ is denoted a positive semi-definite matrix (p.s.d). In context of PCO, non-negative eigenvalues can only be guaranteed when the Euclidean distance has been used for determining dissimilarities (Gower, 1966).

(4) If l dimensions are desired for modelling the data, retain the first l columns in $\boldsymbol{Q}$ and first l rows and l columns in $\boldsymbol{\Lambda}$, the PCO scores matrix $\boldsymbol{T}$ can be obtained by $\boldsymbol{T} = \boldsymbol{Q} \cdot \boldsymbol{\Lambda}^{1/2}$ with imaginary elements if there are negative eigenvalues.

It is important to pay special attention to the values of the diagonal elements of $\boldsymbol{\Lambda}$. If a non-Euclidean distance is used to construct the distance matrix, it is possible that one or a few eigenvalues in $\boldsymbol{\Lambda}$ are negative. In such case, the original distance matrix cannot be perfectly embedded into Euclidean space because some PCs exist in imaginary space. However, if the magnitude of the negative eigenvalues is relatively small compared to that of positive eigenvalues, the negative part is usually indicative of noise. The rationale behind such methodology is given by the Wielandt-Hoffman theorem (Golub and Loan, 1996): the best approximation of a matrix $\boldsymbol{G}$ whose eigenvalues $\boldsymbol{\Lambda}$ contain negative elements is by setting all negative elements in $\boldsymbol{\Lambda}$ to zero to give a matrix $\hat{\Lambda}$ and then reconstructing $\hat{\boldsymbol{G}}$ using $\boldsymbol{Q}$ and the new matrix $\hat{\Lambda}$. This method can be justified as noise reduction so long as the negative eigenvalues are small relative to the positive ones and has been extensively used in PCO and kernel learning related literature (Krzanowski and Marrior, 1994; Graepel *et al.*, 1998; Peres-Neto and Jackson, 2001; Pekalska *et al.*, 2002). In our study, the quantitative dissimilarity matrices given by equations (3) and (4) contain only a few very small negative eigenvalues but Jaccard distance matrices often result in some very large negative eigenvalues. Fortunately, as demonstrated by Gower and Legendre (1986), although the Jaccard distance itself is non-Euclidean, the square rooted Jaccard distance has Euclidean properties and hence can be perfectly embedded into Euclidean distance space. The square rooted fuzzy weighted Jaccard distance sometimes still has some negative eigenvalues but such imperfection has been largely reduced, reflected by the ratios of the sum of absolute negative eigenvalues over that of positive eigenvalues are significantly lower than before and never exceed 10%, hence we just ignore all PCs with negative eigenvalues and keep those with positive eigenvalues. Hence in equations (2) and (5) we use the square root of the Jaccard dissimilarity metric.

Once the pair-wise distance matrices are transformed to sample-latent variable matrices using PCO there are various methods available to measure the correlation between these matrices. We employ the Procrustean Test (Jackson, 1995) to measure the correlation between the two types of data inputs (GC/MS and DGGE). The most attractive aspect of this method is that it is possible to superimpose two patterns on the same scatter plot and visualise the matching. The test involves two steps.

(1) The Procrustean transformation (Rohlf and Slice, 1990) is performed on one data matrix and the other used as the target. The transformed matrix as well as the matching error when it is compared to the target matrix is calculated as follows. Given two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ to be compared, we set matrix $\boldsymbol{A}$ as the

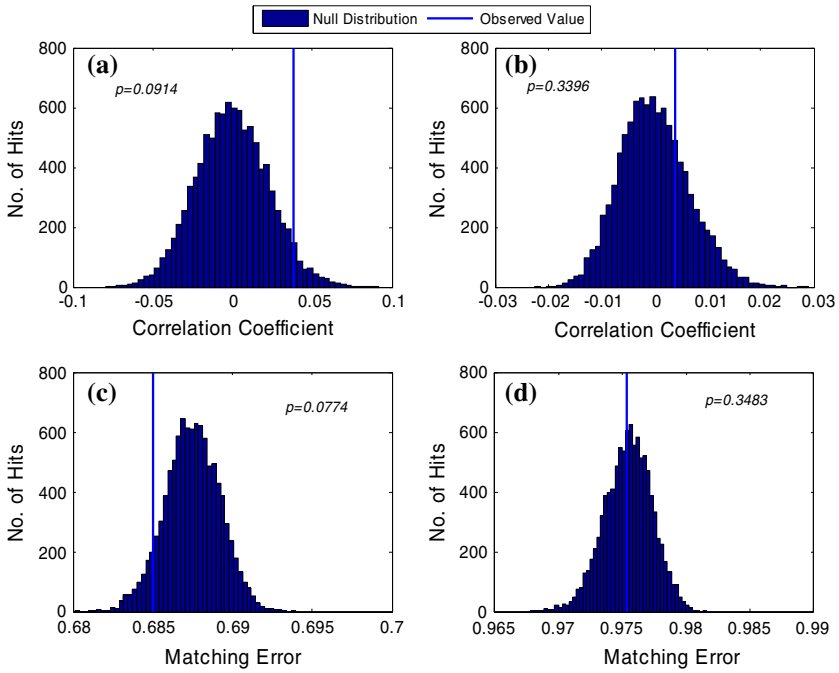

Figure 2. Mantel test and Procrustean test applied to the full data set of 354 samples: (a) Mantel test on qualitative distance matrices; (b) Mantel test on quantitative distance matrices; (c) Protest on qualitative distance matrices and (d) Protest on quantitative distance matrices. A total of 10,000 permutations are performed and for the purpose of the bar chart, data are divided into 50 equally spaced bins between the lowest and light values on the graph, the number of hits is the number of time the relevant parameter falls within a bin.

target and transform $\boldsymbol{B}$ to match $\boldsymbol{A}$ as closely as possible. The Procrustean transformation involves three basic operations:

(a) *Translation*. Column mean centre both $\boldsymbol{A}$ and $\boldsymbol{B}$ and scale both matrices so that the sum of squares of all the elements in each matrix equals 1. The mean vector and the root sum of squares of all elements in $\boldsymbol{A}$ has is computed for further use, as discussed below. We denote the matrices after the translation as $\boldsymbol{A}_{scl}$ and $\boldsymbol{B}_{scl}$.

(b) *Rotation*. The rotation matrix $\boldsymbol{R}$ is obtained by performing Singular Value Decomposition (Gentle, 1998) on the matrix $\boldsymbol{A}'_{scl}\boldsymbol{B}_{scl}$ to get $\boldsymbol{A}'_{scl}\boldsymbol{B}_{scl} = \boldsymbol{USV}'$, $\boldsymbol{R} = \boldsymbol{VU}'$ and in this application all N non-zero components are retained. Note that SVD, unlike eigen-decomposition, always results in a positive diagonal matrix $\boldsymbol{S}$.

(c) *Scaling*. The scaling factor is obtained by $s = \sum_{n=1}^{N} s_{nn} \|\boldsymbol{A}\|$, where $||\boldsymbol{A}||$ is the root sum of square for all the elements in matrix $\boldsymbol{A}$. The transformed matrix obtained from $\boldsymbol{B}$ can be calculated by $\boldsymbol{B}_{trans} = s \cdot \boldsymbol{B}_{scl} \cdot \boldsymbol{R} + \boldsymbol{1} \cdot \boldsymbol{a}$ where $\boldsymbol{a}$ is the row vector of means of matrix $\boldsymbol{A}$ and $\boldsymbol{1}$ is column vector of 1s. The matching error e is given by $e = ||\boldsymbol{B}_{trans} - \boldsymbol{A}||$.

(2) Evaluate the significance level of the matching error using a Monte Carlo simulation procedure. Like the Mantel test, a null distribution of the matching error is derived by permuting the order of samples in one data matrix (the order of the variables remains the same) while keeping the other unchanged. Procrustean transformation is performed on this new pair of matrices and the matching error computed as described above. This process is repeated a large number of times and the matching errors form the null distribution. The null distribution is then used to assess the significance level of the observed matching error and to give p values as described above for the Mantel test.

Since the matching error can never be negative, the Procrustean test is always a one-tailed test. One can also visualise the matching of two most significant components of the data matrices by superimposing the Procrustean transformed matrix with the target matrix in a 2-D or 3-D scatter plot.

It is necessary to note that the p values of both the Mantel test and the Procrustean test are estimated from empirical null distributions which are formed by using Monte Carlo simulations. If none of the randomised resampling experiments produced a higher (or lower) value than the true observed value, the estimated p value is reported as 0. Such a case suggests that the observed value is significantly different from the null distribution. However this does not necessarily mean that the probability of the observed value derived from the null distribution is actually 0, i.e. that the null hypothesis can be rejected at 100% confidence level.

4. Results

The first step was to compare the patterns of the GC/MS and DGGE profiles by using all $2P$ ($=354$ pairs of samples). The null distributions for the Mantel and Procrustean test were obtained using 10,000 permutations. Both the two-sided Mantel test and Procrustean test suggest that there is a weak correlation ($p < 0.1$) between these two data blocks when qualitative distance metrics were used while no significant correlation can be found when quantitative distance metrics were used. The results are presented in figure 2, using 50 equally spaced bins as illustrated.

We then performed such a comparison on each of the family groups separately. At first sight, the results for each family give apparently inconsistent conclusions. Some families showed very significant correlation between two data blocks while some families do not show any significant correlation at all. When the qualitative distance metrics were used, the strongest correlation between the DGGE and GC/MS signals come from family D (11 members, 22 samples), both the Mantel test and Procrustean test result in a value of $p = 0$, i.e. not a single case in 10,000 random permutations results in such high correlation and the observed value of correlation coefficient or matching error of the Procrustean transformation is also far away from the corresponding null distribution. Family A (17 members, 34 samples) also shows very strong correlation at the level of $p = 0$ but the observed values are closer to the null distributions compared to family D. The superimposed PCO scores plots of two data blocks of these two families are shown in figure 3. To interpret these graphs, it is necessary to observe how close the corresponding DGGE and GC/MS samples appear. Each family member has a unique identifying number, and the scores of the two samples from each individual are averaged after the Procrustes transformation, for simplicity. Note, for example in figure 3(b) that the two samples for individual 86 appear in the top right corner, 83 in the bottom left corner, and 91 and 88 are in the middle. Had there been no correlation we would expect these paired samples to appear more or less randomly. Families G, P and Q also show strong correlations ($p < 0.05$); families B, C, J show weak correlations ($p < 0.1$) but correlations between the two blocks for families E, H, L, M, N, O, R, S and U are consistent with the random model for the null distribution at a level of $p > 0.1$. When the quantitative distance metrics were used, only three families, G, P and Q show correlations whose p values varied from 0.01 to 0.07. The results from all 17 families are summarised in table 1.

However it is important to consider that human odour can be significantly influenced by many environmental factors, so we examined the behaviour surveys which the subjects filled in when the samples were taken. We found that there are considerable variations in human behaviour, such as the time elapsed from the last armpit washing, the time from the last use of deodorant or soap and whether the suggested T-shirt (chemical free) was worn. This variation was suspected to influence the GC/MS volatile compound profiles. We expect that such variations can impact subject's odour, as well as their microbial profiles. Hence we defined four subject screening rules to reduce possible influence from environmental factors.

(1) The time of the last armpit washing must be between 12 and 48 h.
(2) The time of the last usage of deodorant must be no less than 48 h.
(3) The suggested T-shirt must be worn before the sampling.
(4) The suggested soap must be used in the last armpit washing.

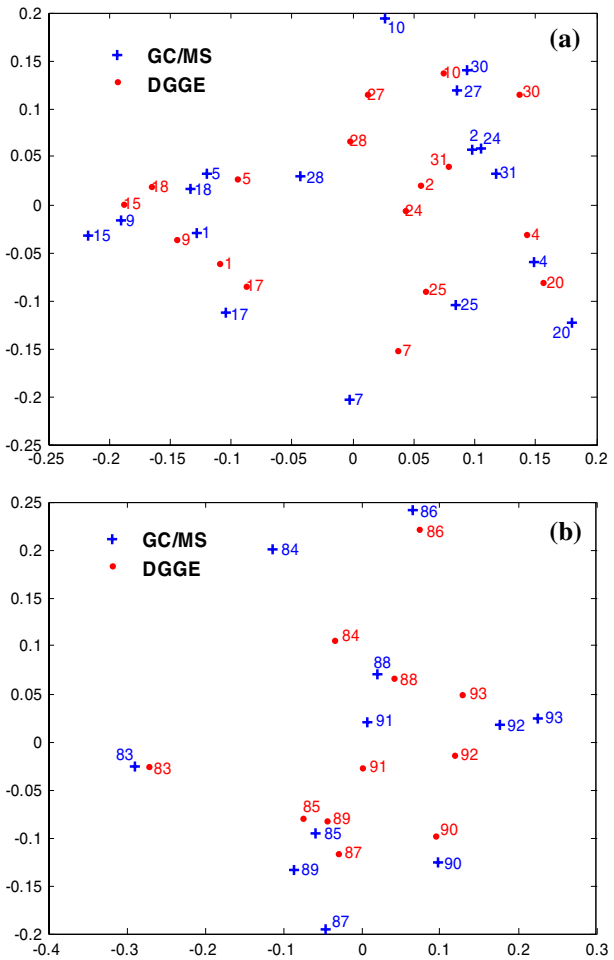

Figure 3. Superimposed scatter plots of PCO scores of the first two components of the GC/MS data and Procrustean transformed DGGE data on families A and D. (a) Family A; (b) Family D. Each number in the figure represents one test subject; the PCO scores of two repeats of one testing subject have been averaged to avoid making the plots too crowded.

Table 1
The results of Mantel test and Protest on all families

Family	Number of subjects	*p*-value of Mantel test		*p*-value of Protest	
		Qualitative distance metric	Quantitative distance metric	Qualitative distance metric	Quantitative distance metric
A	17	***0***	0.1676	***0***	0.1231
B	15	***0.0571***	0.2895	***0.0615***	0.137
C	12	***0.0641***	0.2764	***0.0607***	0.2514
D	11	***0***	0.1457	***0***	0.1512
E	8	0.4075	0.4806	0.5837	0.8977
G	9	***0.0423***	***0.067***	***0.0431***	***0.045***
H	11	0.3114	0.355	0.2451	0.3055
J	10	***0.0912***	0.3181	***0.0634***	0.2181
L	15	0.2088	0.1112	0.4714	0.4755
M	8	0.5007	0.5592	0.5884	0.5983
N	10	0.128	0.4324	0.1076	0.3478
O	5	0.1763	0.2691	0.1136	0.2919
P	9	***0.0016***	***0.0185***	***0***	***0.0112***
Q	10	***0.0095***	***0.0139***	***0.0032***	***0.0103***
R	10	0.4331	0.3271	0.4174	0.6425
S	9	0.4053	0.4513	0.4187	0.519
U	8	0.2989	0.3587	0.2985	0.3056
All	177	***0.0914***	0.3396	***0.0774***	0.3483

All *p*-values less than 0.1 are in italic and bold.

By applying these four rules, only 32 of the 177 subjects meet all the requirements. The Mantel test and Protest were applied to these subjects using a null distribution based on 64 rather than 354 samples. The results show very strong correlation between two data blocks at the level of $p = 0$ with the qualitative distance metrics and $p < 0.01$ with the quantitative distance metrics (see figure 4). For comparison purpose, we also randomly selected another 32 subjects from the remaining 145 subjects that had not obeyed these behavioural rules (excluding the 32 individuals above) and did the same experiment and repeat the computations 10 times with different random selections of subjects. The results showed that 1 experiment resulted in strong correlation at the level of $p < 0.01$ and 1 experiment resulted in weak correlation at the level of $p < 0.1$ using a qualitative distance metric; 1 experiment resulted in strong correlation at the level of $p < 0.05$ with the quantitative distance metric. The results are summarised in table 2.

We also compared the null distributions for the qualitative and quantitative distance metrics generated by Monte Carlo simulation both for the overall population and for each family separately. Although the null distribution generation procedures are exactly the same and the only difference involves using different distance metrics, the characteristics of the null distributions are very different. For the Mantel test, both null distributions of qualitative and quantitative distances are centred around 0 as expected (as a correlation coefficient of 0 implies that there is no link between two blocks of data) but the distributions of the qualitative distance metrics always appeared to have much larger standard

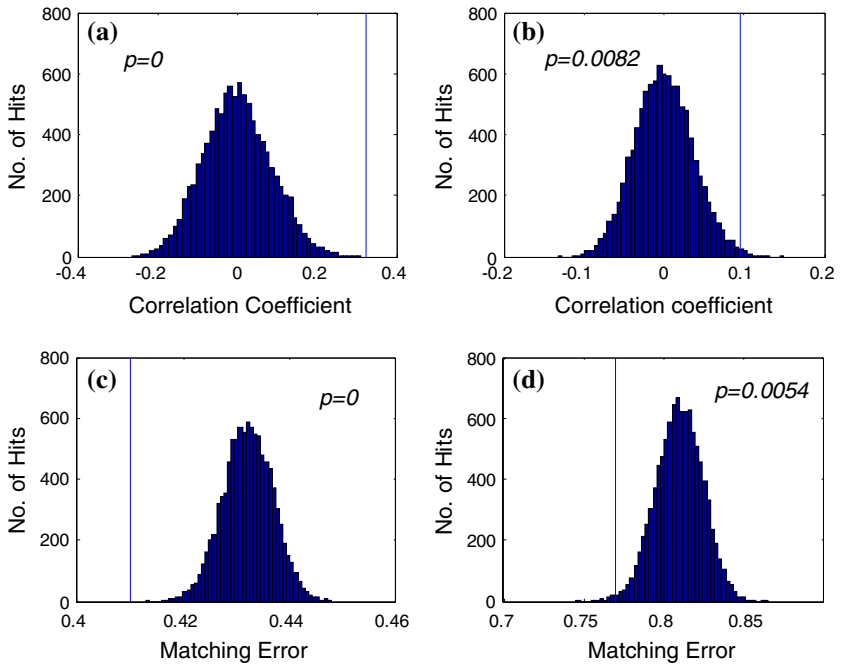

Figure 4. Mantel test and Procrustean test applied to the selected data set of 32 subjects (64 samples) that obeyed four behavioural rules over the period of sampling. (a) Mantel test on qualitative distance matrices; (b) Mantel test on quantitative distance matrices; (c) Protest on qualitative distance matrices and (d) Protest on quantitative distance matrices. A total of 10,000 permutations are performed and for the purpose of the bar chart, data are divided into 50 equally spaced bins between the lowest and light values on the graph, the number of hits are the number of time the relevant parameter falls within a bin.

Table 2
The results of Mantel test and Protest on randomly selected sets of 32 subjects chosen among the 145 subjects that had not obeyed the four basic behavioural rules

	p-value of Mantel test		*p*-value of Protest	
Run	Qualitative distance metric	Quantitative distance metric	Qualitative distance metric	Quantitative distance metric
1	0.2145	0.5412	0.1947	0.4922
2	0.3388	0.4178	0.4408	0.7998
3	***0.0092***	***0.0412***	***0.0088***	***0.0359***
4	0.3919	0.4082	0.6147	0.7912
5	0.1987	0.3024	0.1471	0.2921
6	0.3014	0.3954	0.4157	0.6955
7	0.4001	0.4413	0.7041	0.8589
8	0.1687	0.3011	0.1577	0.4998
9	***0.0821***	0.1498	0.0714	0.1501
10	0.2687	0.3541	0.2514	0.6562

All *p*-values less than 0.1 are in italic and bold.

deviation compared to those of the quantitative distances, i.e. the histogram of the p.d.f. (probability distribution function) of the null distribution for the qualitative distance metric is much broader. It may suggest that for qualitative distance matrices, the permutation process has much higher influence on the calculated correlation coefficient for each simulation. For the Procrustes test, matching errors are used to form the null distribution. The most obvious difference is that the null distribution for the qualitative distance always has much lower average (centre of the p.d.f.) compared to that of the quantitative distance. It suggests that when the qualitative distance metric is used, the matching error calculated from Procrustes analysis could be low even if there is in fact no real correlation, hence should be interpreted only in the light of Monte Carlo simulations.

5. Conclusion

Based on the results of the Mantel test and the Procrustean test, it appears that there is no strong correlation for the overall population between microbial and chemical odour profiles. However when divided into individual families, some strong correlations are revealed and it is probable that correlations are overwhelmed by environmental factors such as the personal habits and living conditions of the subjects. When we select individuals whose behaviour follows certain rules, there is indeed a significant correlation between the microbial and chemical odour profiles: the differences between the families could also be interpreted in terms of the different personal behaviour patterns in certain family groups. We also notice that in most cases, the correlation between human odour profiles and microbial profiles is usually stronger when the qualitative distance metrics were used compared to quantitative distance metrics. This is interpreted in part as the amount of sweat sampled not being precisely controlled and the method for DGGE gels is only semi-quantitative so quantitative dissimilarity metrics are more sensitive to such variations, whereas qualitative patterns based on presence/absence of microbes and chemicals are more robust.

Acknowledgements

Alexandra Katzer is thanked for her superb organisational skills. Hejun Duan of the Centre for Chemometrics is thanked for helping organise the GC/MS data, and Fan Gong for preliminary help in the microbial analysis. This work was sponsored by ARO Contract DAAD19-03-1-0215. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

References

Brereton, R.G. (2003). *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Wiley, Chichester.

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *J. R. Stat. Soc. B* **26**, 211–252.

Dixon, S.J., Brereton, R.G., Soini, H.A., Novotny, M.V. and Penn D.J. (2006). An automated method for peak detection and alignment in gas chromatography-mass spectrometry as applied to a large metabolomic dataset from human sweat. *J. Chemomet.* in press.

Gentle, J.E. (1998). *Numerical Linear Algebra for Applications in Statistics*. Springer-Verlag, Berlin.

Golub, G.H. and Loan, C.F.V. (1996). *Matrix Computations*. The Johns Hopkins University Press, London.

Gower, J.C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Gower, D.B., Bird, S., Sharma, P. and House, F.R. (1985). Axillary 5α-androst-16-en-3-one in men and women: Relationships with olfactory activity to odorous 16-androstenes. *Cell. Mol. Life Sci.* **41**, 1134–1136.

Gower, J.C. and Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *J. Classif.* **3**, 5–48.

Graepel, T., Herbrich, R., Bollmann-Sdorra, P. and Obermayer, K. (1998). Classification on pairwise proximity data in Jordan, MI, Kearns, MJ and Solla, SA (Eds), *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA, pp. 438–444.

Huber, W., von Heydebrek, A., Sültmann, H., Poustka, A. and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18**(suppl. 1), S96–S104.

Jaccard, P. (1908). *Bull. Soc. Vaud. Sci. Nat.* **44**, 223–270.

Jackson, D.A. (1995). Protest: A Procrustean randomization test of community environment concordance. *Ecoscience* **2**, 297–303.

Krzanowski, W.J. and Marrior, F.H.C. (1994). *Multivariate Analysis, Part I. Distributions, Ordination and Inference*. Arnold, London.

Mantel, N.A. (1967). The detection of disease clustering and a generalized regression approach. *Can. Res.* **27**, 209–220.

Marples, M.J. (1969). Life on the human skin. *Sci. Am.* **220**, 108–115.

Muyzer, G., de Waal, E.C. and Uitterlinden, A.G. (1993). Profiling of complex microbial populations by denaturing gradient gel

electrophoresis analysis of polymerase chain reaction-amplified genes encoding for 16S rRNA. *Appl. Environ. Microbiol.* **59**, 695–700.

Muyzer, G., Hottenträger, S., Teske, A. and Wawer, C. (1995). Denaturing gradient gel electrophoresis of PCR-amplified 16S rRNA: A new molecular approach to analyse the genetic diversity of mixed microbial communities in Akkermans, A.D., Elsas, J.D.van and Bruijin, F.J.de (Eds), *Molecular Microbial Ecology Manual*. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1–23.

Pekalska, E., Paclik, P. and Duin, R.P.W. (2002). A generalized Kernel approach to dissimilarity-based classification. *J. Mach. Learn. Res.* **2**, 175–221.

Penn, D.J., Oberzaucher, E., Grammer, K., Fischer, G., Soini, H.A., Wiesler, D., Novotny, M.V., Dixon, S.J., Xu, Y. and Brereton, R.G. (2007). Individual and gender fingerprints in body odour. *J. R. Soc.: Interface* **4**, 331–340.

Peres-Neto, P.R. and Jackson, D.A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129**, 169–178.

Rennie, P.J., Gower, D.B., Holland, K.T., Mallet, A.I. and Watkins, W.J. (1990). The skin microflora and the formation of human axillary odour. *Int. J. Cosmet. Sci.* **12**, 197–208.

Rennie, P.J., Gower, D.B. and Holland, K.T. (1991). In-vitro and in-vivo studies of human axillary odour and the cutaneous microflora. *Br. J. Dermatol.* **124**, 596–602.

Rodrìguez-Lázaro, D.A., Jofré, T., Aymerich, M., Hugas, M. and Pla, M. (2004). Rapid quantitative detection of *Listeria monocytogenes* by Real-Time PCR. *Appl. Environ. Microbiol.* **70**, 6299–6301.

Rohlf, F.J. and Slice, D.E. (1990). Extensions of the Procrustes method of the optimal superimposition of landmarks. *Syst. Zool.* **39**, 40–59.

Sanguinetti, C.J., Dias Neto, E. and Simpson, A.J.G. (1994). Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* **17**, 915–919.

Sastry, S.D., Buck, K.T., Janak, J., Dressler, M. and Preti, G. (1980). Volatiles emitted by humans in Waller, G.R. and Dermer, O.C. (Eds), *Biochemical Applications of Mass Spectrometry*. John Wiley & Sons, New York, pp. 1085–1129.

Sato, K., Leidal, R. and Sato, F. (1987). Morphology and development of an apoeccrine sweat gland in human axillae. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **252**, R166–180.

Soini, H.A., Bruce, K.E., Klouckova, I., Brereton, R.G., Penn, D.J. and Novotny, M.V. (2006). In-situ surface sampling of biological objects and preconcentration of their volatiles for chromatographic analysis. *Anal. Chem.* **78**, 7161–7168.

Wold, S., Esbensen, K. and Geladi, P. (1987). Principal component analysis. *Chemom. Intell. Lab. Syst.* **2**, 37–52.

Williamson, P. and Kligman, A.M. (1965). A new method for the quantitative investigation of cutaneous bacteria. *J. Invest. Dermatol.* **45**, 498–503.

Zwillinger, D. (1997). *Handbook of Differential Equations*. (third ed.). Academic Press, Boston.