# Metabolite fingerprinting of urine suggests breed-specific dietary metabolism differences in domestic dogs

Manfred Beckmann[1]*, David P. Enot[1,2], David P. Overy[1,3], Ian M. Scott[1], Paul G. Jones[4], David Allaway[4] and John Draper[1]

[1]*Institute of Biological Environmental and Rural Sciences, Aberystwyth University, Penglais Campus, Aberystwyth SY23 3DA, UK*
[2]*BIOCRATES Life Sciences AG, Innrain 66, A-6020 Innsbruck, Austria*
[3]*University of Prince Edward Island, Duffy Research Center (NRC-INH), 550 University Avenue, Charlottetown, PEI, Canada C1A 4P3*
[4]*WALTHAM Centre for Pet Nutrition, Freeby Lane, Waltham-on-the-Wolds, Melton Mowbray, Leicestershire LE14 4RT, UK*

Selective breeding of dogs has culminated in a large number of modern breeds distinctive in terms of size, shape and behaviour. Inadvertently, a range of breed-specific genetic disorders have become fixed in some pure-bred populations. Several inherited conditions confer chronic metabolic defects that are influenced strongly by diet, but it is likely that many less obvious breed-specific differences in physiology exist. Using Labrador retrievers and miniature Schnauzers maintained in a simulated domestic setting on a controlled diet, an experimental design was validated in relation to husbandry, sampling and sample processing for metabolomics. Metabolite fingerprints were generated from 'spot' urine samples using flow injection electrospray MS (FIE-MS). With class based on breed, urine chemical fingerprints were modelled using Random Forest (a supervised data classification technique), and metabolite features (*m/z*) explanatory of breed-specific differences were putatively annotated using the ARMeC database (http://www.armec.org). GC-MS profiling to confirm FIE-MS predictions indicated major breed-specific differences centred on the metabolism of diet-related polyphenols. Metabolism of further diet components, including potentially prebiotic oligosaccharides, animal-derived fats and glycerol, appeared significantly different between the two breeds. Analysis of the urinary metabolome of young male dogs representative of a wider range of breeds from animals maintained under domestic conditions on unknown diets provided preliminary evidence that many breeds may indeed have distinctive metabolic differences, with significant differences particularly apparent in comparisons between large and smaller breeds.

**Dog breeds: Metabolomics: Random Forest: Urine**

The domestic dog (*Canis familiaris*) is the most phenotypically diverse mammalian species, with an enormous range of sizes, shapes and behaviours[1]. Over the last 150 years, pedigree breeding barriers and standards have resulted in levels of genetic polymorphism that can be greater among domestic breeds than between wild canid species[2]. Alongside selection of desired morphological and behavioural traits, a range of less obvious (including metabolic) genetic disorders have also been inadvertently selected and enriched within a number of breeds[3–6]. In consequence, pure-bred dogs are considered useful models of human diseases and lifestyle-related disorders[7]. Although some advances have been made in the last three decades, however, many of the nutritional requirements of companion animals remain unknown[7]. With the popularity of complete dog foods manufactured from a wide range of ingredients, it may be possible to tailor diet formulations for dogs with differing physiological needs and metabolic characteristics[7–9].

A recent appreciation of the importance of interactions between genetics, behaviour, environment and nutrition on human health has resulted in the emergence of the field of nutrigenomics[10,11]. The interactions between diet and genotype in domestic dogs are exceedingly complex. Certain breeds are specifically prone to chronic ailments associated with nutrition, for example, copper toxicosis in Bedlington Terriers[3]. Certain breeds are also prone to common complaints such as kidney/bladder stones and sensitive bowel problems, in which nutrition can play a role in development and treatment[4,5,12,13]. However, the underlying physiological/ metabolic mechanisms and the contribution of genotype– diet interactions are often poorly understood.

Nutritional studies using commercially available pet food are hampered by the fact that most food raw materials are in the form of a chemically complex matrix. To ensure nutritional requirements stipulated by local legislative bodies are met, manufacturers may supplement diets to produce a complete and balanced pet food. Beyond this, different brands/ manufacturers may add other defined constituents such as vitamins, minerals, plant oils and prebiotics with functional nutritional claims[14,15]. However, the value of such additives at the level of individual breeds may not have been ascertained.

In addition to factors related to breed genotype and diet composition, the diverse microbial flora in the mammalian gastrointestinal tract is capable of a wide range of enzyme-catalysed chemical transformations of metabolites derived from the diet[16]. The metabolite profiles of accessible biofluids such as blood, urine and saliva are also affected by behavioural factors such as water intake, meal times, exercise and stress. For example, it is estimated that urine can contain in excess of 2000 different metabolites in omnivorous mammals, with many chemicals displaying huge dynamic variation[17,18]. Urine is a useful analytical biofluid because it is a concentrated collection of metabolites that are a result of mechanisms to maintain the homoeostasis of the individual. It is likely, therefore, that different physiological stresses may manifest themselves in the urine metabolite fingerprint. Urine composition is also likely to contain information on how the individual responds to diet. Furthermore, the non-invasive nature of urine collection makes it a practical biofluid to collect[19,20].

Recent improvements in both analytical chemistry and data mining have resulted in the establishment of metabolomics technology platforms capable of analysing objectively the global composition of complex biofluids such as urine[19–24]. In the absence of reference chemicals for all expected metabolites, methods for 'metabolite fingerprinting'[25] have proven valuable for exploring genotypic and phenotypic differences among organisms. For example, Fourier transform IR[26] and NMR[19,27] spectroscopies can generate global chemical fingerprints without specialised sample preparation. However, although reproducible and intrinsically quantitative, NMR is relatively insensitive compared with other techniques and, for the vast majority of metabolites, a further level of directed analysis is generally required to link differences in chemical shifts to specific chemistry and thus develop new biomarkers[19,21,28–31]. Fingerprinting techniques based on MS such as flow injection electrospray MS (FIE-MS) are more sensitive, and offer the advantage that the measured 'variables' (mass-to-charge ($m/z$) ratios) can be linked more directly to specific metabolites by the additional indication of molecular mass, particularly in data derived from ultra-high mass accuracy instruments[25,32–36]. In FIE-MS, a chromatographic step is not utilised and a fingerprint is developed following 'soft' ionisation of the sample during injection over 1–2 min. Such fingerprints can be regarded as simplified images of total sample composition, in that each $m/z$ variable may integrate the levels of several isobaric metabolites. During soft ionisation, the main products are charged versions of the parent molecule and salt adducts; fragmentation products are relatively rare and thus the identity of molecules producing signal at specific $m/z$ can be investigated directly based on predicted mass of the metabolite[37].

The present study set out to explore the utility of metabolite fingerprinting of urine, based on MS, for nutritional studies in domestic dogs. Recent NMR investigations have validated the utility of urine metabolomics for breed classification in dogs[19]. Additionally, urine analysis of Labrador retrievers suggested that there are urinary markers consistent with age and caloric regimen[28]. The objectives were to determine whether a urine metabolic fingerprint based on MS was a useful tool to investigate breed-specific metabolic differences and whether other physiological or environmental factors have more of an impact upon the urine metabolome. One experimental population consisted of male and female Labrador retriever and miniature Schnauzer dogs housed in a single, purpose-built, environmentally enriched facility and maintained on the same diet. Another population consisted of client-owned male dogs of ten breeds fed unspecified diets in private domestic environments.

## Experimental methods

### Animal maintenance and urine collection

Labrador retriever and miniature Schnauzer dogs in the controlled-diet studies were housed in small groups in the complex of purpose-built, environmentally enriched facilities at the WALTHAM Centre for Pet Nutrition, in accordance with the Centre's research ethics and UK Home Office Regulations. They had free access to water and were given commercially available, complete dog food throughout the study at energy levels to maintain adult body weight. After an overnight fast ($>20$ h), dogs were given individual morning walks, during which urine was obtained by a 'free catch' in an individually designated Uripet collection vessel. Urine samples were at ambient temperature for up to 30 min before being placed on ice, and then frozen within 1 h of sampling. Samples were kept at $-80°C$ until shipment on dry ice to Aberystwyth University, where they were stored at $-80°C$.

In the technical validation study, four male Labrador retrievers (aged 6–8 years) were maintained on Pedigree Complete Adult dog food (Mars Petcare UK Ltd, Melton Mowbray, UK). In the main controlled-diet study, Labrador retrievers (five male and six female) and miniature Schnauzers (seven male and five female) were fed Pedigree Complete Adult Small Breed dog food, for at least 4 weeks before urine collection. For each animal, up to eight urine samples were collected on different days within a 2-week longitudinal study.

The dogs from private homes were healthy males recruited in the US at several veterinary clinics before a client-elected surgical castration. The breeds represented in the present study were Labrador retriever, Golden retriever, Rottweiler, German shepherd, Doberman, Beagle, Cocker spaniel, Poodle, Yorkshire terrier and Shih tzu. Urine was collected by catheter during routine veterinary surgery (neutering), and stored immediately at 80°C, before shipment and further storage as above.

### Metabolite fingerprinting by flow injection electrospray MS

FIE-MS was carried out as previously described with modification to accommodate urine samples[25]. Aliquots of thawed urine (50 μl) were diluted in sterile distilled water (100 μl) and then added to MeOH (350 μl). The samples were vortexed and centrifuged for 6 min at 14 000 $g$. For FIE-MS analysis, supernatant (60 μl) was transferred to HPLC crimp cap glass vials containing a 200 μl micro glass insert. The vial caps were crimped and then stored at 4°C for same day analysis. Vials were randomised before injection using an autosampler (tray temperature: 15°C). The diluted samples (20 μl) were injected into a flow of 60 μl/min water–methanol (50:50 v/v) using a Surveyor liquid chromatography system (Thermo Scientific, Waltham, MA, USA) as previously described[25].

For each sample, data were acquired in positive and negative ionisation modes in two scan ranges (low: 15–200 $m/z$; and high: 110–2000 $m/z$) on a LTQ linear ion trap (Thermo Electron Corporation, San Jose, CA, US). Acquisition time was 5 min. Raw data of the whole infusion profiles were exported and mass intensities of each scan electronically binned to a nominal mass of 1 amu (between $-0.3$ and 0.7 amu). The resulting mass spectrum for each analysis was calculated as the mean of eleven scans about the apex of the infusion profile. Mass spectra of all analytical runs per sample and ionisation mode were combined in a single intensity matrix (runs × $m/z$ values). Data were $\log_{10}$ transformed and normalised to total ion current before analysis.

### Flow injection electrospray MS signal interpretation

The initial feature selection by Random Forest (RF) produced a list of $m/z$ signals ranked by importance scores or $P$ value for each classification task. Using ARMeC, a bespoke database developed for FIE-MS signal annotation[37], the top ranked $m/z$ values (generally top 40) were examined for groups of potentially related signals that could represent salt adducts (e.g. Na = M + 23; K = M + 39; Cl = M + 35), common neutral losses (e.g. water = M − 18; formate = M − 46) or isotopes (M + 1 or +2) of individual metabolites. As several overlapping solutions predicting the presence of different metabolites were often possible, the most likely combination of ions identifying a specific metabolite was confirmed by further examining signal relationships in a correlation analysis using just $m/z$ with an appropriate low $P$ value. A comprehensive list of metabolites present in the mammalian metabolome was compiled from exhaustive literature and database searches. Further investigation of predicted metabolome differences between samples centred on the targeted analysis of GC-time-of-flight MS (GC-TOF-MS) data, as the majority of metabolite classes predicted to be explanatory of differences among breeds were suitable for GC-MS analysis.

### Metabolite profiling using GC-time-of-flight MS analysis

Neat urine (50 µl) was dried *in vacuo* and GC-TOF-MS analysis performed as previously described[34]. Methoximation of carbonyl moieties in 100 µl of a 20 mg/ml solution of methoxyamine hydrochloride in pyridine (Fluka, St Gallen, Switzerland) was carried out at 30°C for 90 min. Acidic protons were subsequently derivatised with 100 µl N-methyl-N-(trimethylsilyl)-trifluoroacetamide (Machery-Nagel GmbH, Düren, Germany) at 37°C for 30 min. One microlitre of the resulting solution was injected (injector temperature 250°C; 1:2 split ratio) into a Pegasus III GC-TOF-MS system (Leco Inc., St Joseph, MI, USA) and chromatographed on a 20 m DB5-MS column (20 m × 0.25 µm internal diameter × 0.25 µm film) using a temperature gradient (80–330°C over 17 min). Mass spectra were recorded after a solvent delay of 112 s over an $m/z$ range of 54–500 at an acquisition rate of 20 spectra/s and ion source temperature of 230°C.

Peak finding and deconvolution were performed using Leco ChromaTof software. Mass spectra of all detected compounds were compared with in-house standards and spectra in the National Institute of Standards and Technology library (http://www.nist.gov/srd/nist1.htm), and other publicly available databases. All data pre-treatment procedures, such as baseline correction, chromatogram alignment, data compression and curve resolution, were performed using custom scripts in Matlab v.6.5.1 (The Math Works Inc., Natick, MA, USA). Targeted peak lists were generated, and the peak apex intensity of a characteristic mass in a retention time window for each GC-MS dataset was saved in the form of an intensity matrix (run × metabolite) using Matlab.

### Data analysis

All calculations were performed in the R v.2.3 + (http://cran.r-project.org) and Matlab environments as previously described[35,38]. Principal component analysis and eigenvalue decomposition were performed on the covariance matrix using mean-centred data. Linear discriminant analysis (LDA) was implemented in Matlab according to Thomaz[39]. RF analysis was done with the R package *randomForest* (v 4.5 + ) with default settings except that the number of trees was set to 1000. The statistical significance of the RF importance score and model margins were assessed by 1000 random permutations: $P$ was defined as the fraction of scores larger than or equal to the original importance score of the unpermutated data. Model margins obtained under permutation were used to estimate distribution quartiles[40]. Receiver operating characteristic curves were used as an alternative measure of model-predictive abilities[40,41] and summarised as the area under the curve statistic. Correlation analysis of individual signals was performed using variables that fell under a specified threshold in at least one of the comparisons. The absolute value of the correlation coefficient between variables was used as the similarity measure for performing hierarchical cluster analysis with complete linkage as the clustering method (*hclust* function in R). The Wilcoxon rank sum test in Matlab was used as a non-parametric univariate test to assess whether two samples came from the same distribution.

## Results

### Validation of sampling and flow injection electrospray MS fingerprinting for urine sample classification in male Labrador retrievers

To validate the technique, urine from four male Labrador retrievers was subjected to FIE-MS fingerprinting and multivariate data analysis. Principal component analysis revealed that samples from all four dogs were well dispersed in a single large grouping in a scores plot of principle components (PC1 × PC2; Fig. 1(a)). LDA was performed to determine whether urine samples from individual dogs could be discriminated (Fig. 1(b)). Each point represents a chemical fingerprint (high range, positive ion data: 50–2000 $m/z$) of urine from a single dog on eight different days during a 2-week period. In all cases, samples from individual dogs are found in discrete associations of similar dimension, but model statistics suggested that inter-dog differences were minimal (i.e. Tw or eigenvalues < 2 in both DF1 and DF2[38,40]). These data indicated that the metabolome of each dog was sufficiently consistent and the sampling method was deemed adequate for further studies.
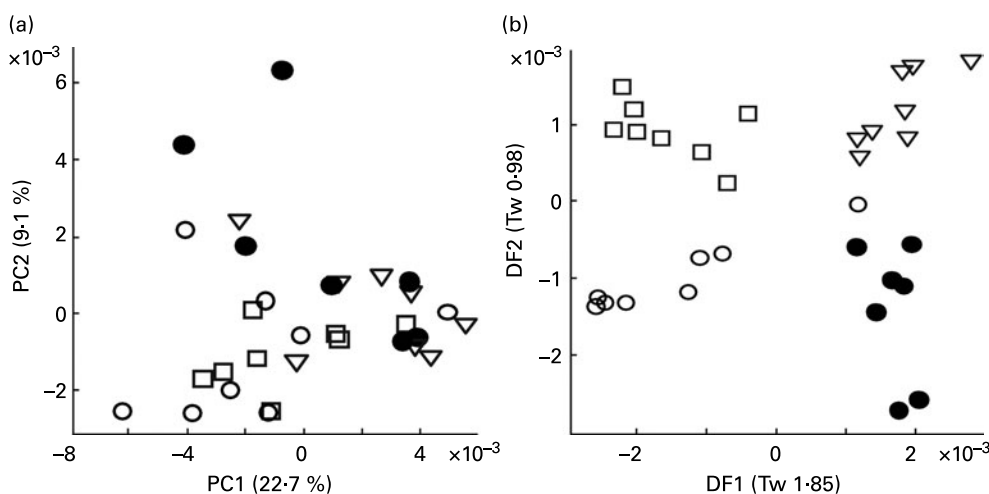
**Fig. 1.** Reproducibility of urine sampling from male Labrador retrievers assessed by multivariate analysis of flow injection electrospray MS fingerprints. Symbols denote individual dogs, and points represent fingerprints (positive ions; *m/z* 50–1100) of daily urine samples taken over 2 weeks. (a) Scores plot of the first two principal components (PC1 and PC2). Numbers in brackets indicate percentage of variance explained by each PC. (b) Linear discriminant analysis (LDA) on same samples. Numbers in brackets indicate eigenvalues (Tw) for each discriminant function (DF).

*British Journal of Nutrition*

*Flow injection electrospray MS fingerprinting to distinguish breed and sex in Labrador retrievers and miniature Schnauzers*

With sampling and analytical chemistry methods established, fuller experiments were conducted to determine the relative importance of sex and breed in discriminating between the urine FIE-MS fingerprints. Urine was collected (*n* 143 individual samples) from five male (*n* 36) and six female (*n* 33) Labrador retrievers, and seven male (*n* 48) and five female (*n* 26) miniature Schnauzers. FIEMS data are highly dimensional and, in the case of urine, were expected to display considerable biological variability.

Such data require the use of robust multivariate data-mining tools. In preliminary experiments, sample classification accuracy was compared using a range of supervised data analysis methods that we have previously used successfully with complex metabolome data[38,40]. LDA and RF were applied to all four FIE-MS datasets (positive and negative ion; high and low *m/z* range) to assess the difficulty of determining both breed and sex in all possible binary comparisons. For two of the datasets, Table 1 summarises LDA and RF model statistics, in the forms of percentage classification accuracies, and of area under the curve for receiver operating characteristic curves. The latter have a value of 0·5 if samples from either class are uniformly distributed across the decision boundary, and a value of 1 if the decision boundary completely discriminates the classes[38,40].

Similar trends were observed in relation to each task with both data-mining methods; breed classification was always much more accurate and generated more robust models than sex assignment. These major features of sample class discrimination are clearly evident in LDA projections (Fig. 2) of the first two discriminant functions.

*Random Forest data mining for metabolome features discriminating dog breeds*

We have previously shown that although LDA and RF will often classify samples with comparable accuracies, there are

advantages in using RF for the subsequent identification of explanatory variables[35]. One output of RF is an importance score for each variable for any particular classification problem. In previous experiments, importance scores ranked in descending order generally fell dramatically and then rapidly levelled out within the range 0·0030–0·0015, below which variables often lack significant explanatory power at $P = 0.005$[35,36,38,40]. Importance scores derived from RF models comparing FIE-MS fingerprints of Labrador retrievers and miniature Schnauzers exhibited similar trends (Fig. 3). In low molecular weight range data, approximately 15–20 *m/z* signals in both ionisation modes potentially have significant explanatory power to discriminate MS from LR dogs (Fig. 3(a)). A similar number of potentially significant signals were evident in the high-range positive ion data with a smaller

**Table 1.** Model statistics for classification of dogs by breed or sex using flow injection electrospray MS fingerprints

| | Classification accuracy (%) | | AUC | |
|---|---|---|---|---|
| Classes compared | RF | LDA | RF | LDA |
| Positive ions, high *m/z* range (50–1100) | | | | |
| LF–LM | 81·5 | 84·9 | 0·88 | 0·89 |
| LF–MF | 92·6 | 97·5 | 0·96 | 0·99 |
| LF–MM | 93·2 | 96·6 | 0·97 | 0·99 |
| LM–MF | 94·5 | 99·6 | 0·99 | 1·00 |
| LM–MM | 90·5 | 96·9 | 0·96 | 0·98 |
| MF–MM | 84·4 | 91·1 | 0·92 | 0·96 |
| Negative ions, low *m/z* range (5–200) | | | | |
| LF–LM | 75·7 | 82·9 | 0·84 | 0·91 |
| LF–MF | 92·7 | 96·3 | 0·97 | 0·99 |
| LF–MM | 92·3 | 93·3 | 0·97 | 0·98 |
| LM–MF | 92·7 | 97·7 | 0·98 | 0·99 |
| LM–MM | 95·0 | 93·5 | 0·98 | 0·97 |
| MF–MM | 77·2 | 77·0 | 0·86 | 0·86 |

AUC, area under the (receiver operating characteristic) curve; RF, Random Forest; LDA, linear discriminant analysis; LF, Labrador retriever female; LM, Labrador retriever male; MF, miniature Schnauzer female; MM, miniature Schnauzer male.
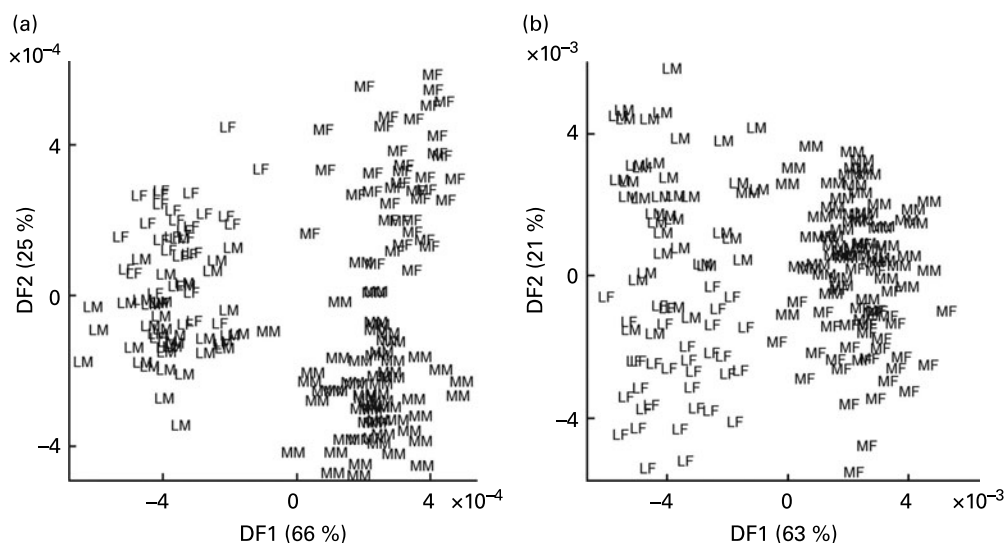
**Fig. 2.** Classification of dogs by breed and sex using urine metabolite fingerprints. Samples from neutered male and female Labrador retrievers and miniature Schnauzers were analysed by flow injection electrospray MS in both ionisation modes, and the spectra subjected to linear discriminant analysis. (a) Positive ion mode; (b) negative ion mode. Percentage of variance accounted for by first two discriminant functions (DF1 and DF2) is in brackets. LF, Labrador retriever female; LM, Labrador retriever male; MF, miniature Schnauzer female; MM, miniature Schnauzer male.

number of ions (8–15) having explanatory power in high-range negative ion data (Fig. 3(b)).

All potentially explanatory $m/z$ signals ($P \leq 0.005$) are shown in RF importance score rank order within all four data-sets (Fig. 4). Putative structural assignments from ARMeC[37] are organised in columns by metabolite class with individual-predicted metabolites numbered as indicated in the figure legend. The annotation of metabolites was guided by previous analyses of human urine in the Human Metabolome Project database[18]. The predictions suggest that more than 50 % of the significant signals in both ionisation modes represent ionisation products originating from a wide range of phenolic molecules. Alternative predictions to phenolics centred largely on hydroxylated organic acids structurally related to butanoate, such as citramalate, 2-hydroxy-3-methylbutyrate or 2-hydroxybutanoate. Signals possibly representing fatty

acids, particularly carnitine derivatives, were also predicted to discriminate between FIE-MS fingerprints of the two breeds. Particularly noteworthy are 240 and 224 $m/z$ (positive ion mode), which were highly correlated and could represent octanoylglycine. Two strongly correlated positive ion signals (527 and 543 $m/z$) commonly associated with trisaccharides such as raffinose and kestose in FIE-MS data[34] were also highly ranked. Signals possibly derived from pantothenate (vitamin $B_5$), glycerol, urea and several polyamines could also have significant explanatory value.

As there were many overlapping potential structural assignments, the phenolic signals were examined in more detail by mapping predicted metabolites onto known metabolic pathways and assessing the degree of correlated signal behaviour in the datasets. A likely solution centred on the metabolism of dietary cinnamates (including chlorogenate) and
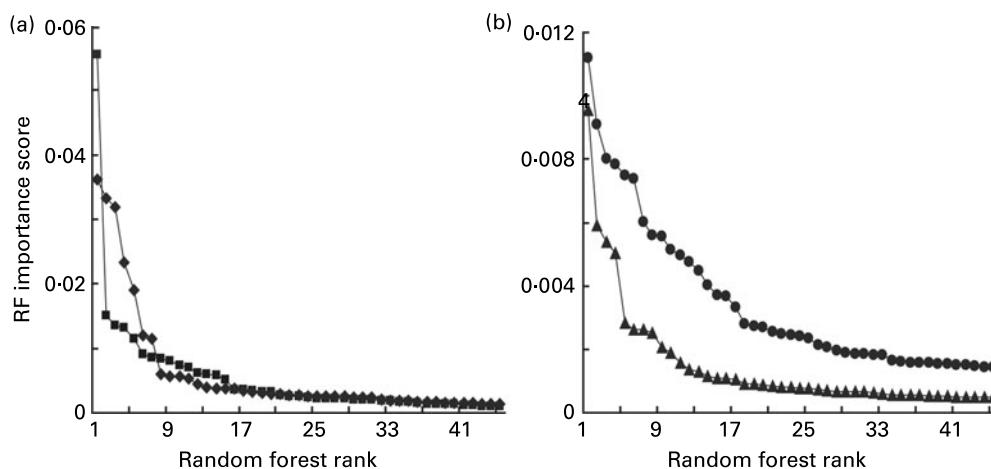


**Fig. 3.** Relative significance of explanatory signals discriminating dog breeds in flow injection electrospray MS (FIE-MS) fingerprints. Importance scores of the top fifty $m/z$ signals (in rank order) from Random Forest (RF) models comparing FIE-MS fingerprints of Labrador retrievers (LR) and miniature Schnauzers (MS) using (a) low $m/z$ range (15–200) or (b) high $m/z$ range (110–1100) data. (a) ■, NegL; ◆, PosL. (b) ▲, NegH; ●, PosH.

**PosL**

| m/z | Phenolics | Organic acids | Fatty acids + choline | Amino acids | Polyamines | Other |
|---|---|---|---|---|---|---|
| 88 | | 1 | | | | |
| 84 | | | 4,5 | | | |
| 157 | 1,23,32 | 3,4,5 | | | | |
| 143 | 2,31 | 6 | | | | 5 |
| 130 | 3 | | 4,5 | 1 | | |
| 108 | 4 | | | | | |
| 56 | | | | | | |
| 115 | 5 | | | | 2 | |
| 43 | | | | | | |
| 148 | 3 | | 4,6 | | | |
| 140 | | | | 1 | | |
| 46 | | | | | | |
| 103 | 2,6,7 | 7 | | | 3 | |
| 183 | 8,33 | | 10 | | | |
| 174 | | | | | | 1 |
| 125 | 3 | 7 | | | 3 | |
| 171 | 5,7,13 | | | | | |
| 131 | 3 | | 4,5 | | | |
| 123 | | | | | | |
| 120 | 3,9,10 | | | | | |
| 55 | | 10 | | | | |
| 200 | 11 | | | | | |
| 199 | 11 | 8 | 6 | | | |
| 89 | | | | | | |

**PosH**

| m/z | Phenolics | Organic acids | Fatty acids + choline | Amino acids | Polyamines | Other |
|---|---|---|---|---|---|---|
| 335 | 7 | | | | | |
| 405 | | | | | | |
| 157 | 1,23,32 | 3,4,5 | | | | |
| 543 | | | | | | 2 |
| 350 | 12,13 | | | | | |
| 143 | 2,31 | 6 | | | | 5 |
| 265 | 5,30 | | | 1 | 2 | |
| 183 | 8,33 | | 10 | | | |
| 258 | | | | | | 1 |
| 527 | | | | | | 2 |
| 240 | | | | 2 | | |
| 226 | | | | 3 | | |
| 442 | | | | | | |
| 375 | 3,14 | | | | | |
| 130 | 3 | | 4,5 | 1 | | |
| 354 | | | | 4 | | |
| 406 | | | | | | |
| 204 | | | | 3 | | |
| 140 | | | | 1 | | |
| 260 | | | | 5 | | |
| 256 | 21 | | | | | |
| 439 | | | | | | 1 |
| 351 | 12,13 | | | | | |
| 319 | 7,15 | | | | | |
| 223 | | | | 6 | | |
| 220 | 16 | | | | 1 | 1 |
| 376 | | | | | | 1 |
| 689 | | | | | | |
| 404 | | | | | | |
| 366 | | | | | | |
| 148 | 3 | | | 4,6 | | |
| 356 | 17 | | | | | |
| 472 | 18 | | | | | |
| 254 | 19,20 | | | | | |
| 224 | 21 | | | 2 | | |
| 120 | | | | | | |

**NegL**

| m/z | Phenolics | Organic acids | Fatty acids + choline | Amino acids | Polyamines | Other |
|---|---|---|---|---|---|---|
| 139 | 1,3,4,7 | 7 | | | 3 | |
| 109 | | 11 | | | | |
| 120 | | 3 | | | | |
| 127 | 22,32 | 10 | | | 3 | |
| 96 | | | | | | |
| 153 | | 11 | | | | |
| 108 | 23 | | | | | |
| 107 | 23 | | | | | |
| 169 | 5,7,24 | 2,5,9 | | | 2 | |
| 121 | 9,10,25 | | | | | |
| 141 | 2,22 | 1 | 7,8 | | | |
| 178 | 21 | | | | | |
| 88 | | | 1,2,3 | | | |
| 150 | 26 | | | | | |
| 176 | 6,11 | | | | | |
| 146 | | | | 4 | 1 | |
| 103 | 7,24 | 1,6,12 | 7 | | | |
| 175 | 6,11 | | | | | |
| 195 | | | | | | |
| 160 | | | 9 | 7 | | |
| 82 | | | | | | |
| 163 | | 27 | | | | |
| 119 | | 2,27,28 | | | | |
| 99 | | | 10,11 | | | |

**NegH**

| m/z | Phenolics | Organic acids | Fatty acids + choline | Amino acids | Polyamines | Other |
|---|---|---|---|---|---|---|
| 139 | 1,4,7 | 7 | | | 3 | |
| 530 | | | | | | |
| 178 | 21 | | | | | |
| 121 | 9,10,25 | | | | | |
| 308 | 19,20,29 | | | | | |
| 380 | | | | | | |
| 124 | | | | | | |
| 175 | 6,11 | | | | | |
| 426 | | | | | | |
| 119 | 2,27,28 | | | | | |
| 335 | 14 | | | | | |
| 522 | | | | | | |
| 395 | 21 | | | | | |
| 333 | 7,24,32 | | | | | |
| 514 | | | | | | |
| 150 | 26 | | | | | |
| 297 | 6,32 | | | | | |
| 146 | | | | | | |
| 291 | | | | | | |
| 475 | | | | | | |
| 621 | 30 | | | | | |

Fig. 4. Metabolite signals discriminating Labrador retrievers from miniature Schnauzers in urine flow injection electrospray MS fingerprints. The left hand side of each of the four panels lists potentially explanatory m/z signals (P≤0·005) in RF importance score rank order within all four datasets (PosL, positive ion low m/z range; PosH, positive ion high m/z range; NegL, negative ion low m/z range; NegH, negative ion high m/z range). In each dataset (ionisation mode × mass range), tentative structural assignments from the signal annotation tool ARMeC (http://www.armec.org) are organised in columns by metabolite class. Individual-predicted metabolites are numbered as follows. *Phenolics*: 1, quinaldic acid; 2, phenylacetaldehyde; 3, phenylanaline; 4, 3-hydroxybenzylalchohol; 5, cinnamaldehyde; 6, 3- or 4-hydroxybenzoate; 7, *trans*-cinnamate; 8, 3,4- or 2,3-dihydroxyphenylacetic acid; 9, 3-hydroxybenzaldehyde; 10, benzoate; 11, 3,4-dihydroxybenzoate; 12, phenylanalineglucoside; 13, *trans*-cinnamate glucoside; 14, 3,4-dihydroxyphenylacetic acid; 15, chlorogenate; 16, tyrosine; 17, 3-hydroxyphenylpropionic acid; 18, cynadin-3-glucoside; 19, naringenin chalcone; 20, pelargodinin; 21, hippuric acid; 22, benzaldehyde; 23, 3-hydroxyphenylacetic acid; 24, *m*-coumaraldehyde; 25, 3-hydroxy-3-phenylpropanoate; 26, hydroxyhippurate; 27, *m*-coumarate; 28, phenylpyruvate; 29, cyanadin; 30, 3- or 4- hydroxybenzoic acid glucoside; 31, *cis-cis*-muconic acid; 31, pre-phenylacetate. *Organic acids*: 1, 3-hydroxyisobutyrate; 2, citrate; 3, 2-hydroxy-3-methylbutyrate; 4, succinate; 5, malate; 6, 2- or 3-hydroxybutyrate; 7, succinate semialdehyde; 8, pyruvate; 9, 2-hydroxy-3-methylbutyrate; 10, 2,3-dihydroxybutyrate; 11, citramalate; 12, hydroxypyruvate; 13, pantoate. *Fatty acids*: 1, tetradecanoic (myristic) acid; 2, octanoyl glycine; 3, acetyl carnitine; 4, decanoyl carnitine; 5, hexanoyl carnitine; 6, dodecenoic (lauric) acid; 7, choline; 8, 2-octanoic acid; 9, carnitine; 10, octanoate; 11, hexenol. *Amino acids*: 1, β-alanine; 2, L-alanine; 3, *n*-methylglycine; 4, glutamate; 5, pyrogluta-mate; 6, taurine; 7, methyl-L-glutamate. *Polyamines*: 1, cittruline; 2, ornithine; 3, cadaverine. *Other*: 1, pantothenate; 2, raffinose; 3, glycerol; 4, glycerate; 5, urea.

flavonoids initially by the gut microflora to generate benzoate or coumarate derivatives destined for excretion (Fig. 5). An alternative pathway could represent the degradation of phenylalanine following deamination to 3-hydroxyphenylace-tate via phenylacetaldehyde.

## GC-time-of-flight MS analysis to confirm the identity of discriminatory metabolome signals

With a clear indication that urine from Labrador retrievers and miniature Schnauzers can be discriminated by metabolite fingerprinting, replicate samples were further subjected to GC-TOF-MS analysis[34,36]. As urine was derivatised directly after drying down without solvent extraction or partitioning, it was expected that GC-TOF-MS should allow the analysis of a wide range of metabolite classes. Preliminary experiments revealed that the urine matrix was complex, highly variable and contained a wide range of structurally unknown chemistry. Under such circumstances, it is extremely difficult to use instrument manufacturers' or third-party software tools to align chromatograms, identify metabolite peaks, deconvolve spectra and quantify peaks. As an alternative approach, we focused on the identification of retention time windows in which total ion currents exhibited breed-specific differences as detected by the Wilcoxon rank sum test following rough alignment of chromatograms using in-house software. Peaks within each of these regions were then manually deconvolved, aligned and their relative intensity calculated. Peaks differing significantly in intensity between breeds in a Wilcoxon test (P≤0·00001) are shown in Table 2. In line with the metabolite fingerprinting, predictions more than 50% of the structurally characterised metabolites were phenolic in nature. Specifically, the fact that 3-hydroxybenzoic acid, 3-hydroxyphenylpropanoic acid, *m*-coumaric acid and hippuric acid were all strongly discriminatory suggested that the major compositional differences in urine of the two dog breeds centred on metabolism of dietary polyphenols (Fig. 5; compounds in boxes), rather than endogenous metabolism of phenylalanine via phenylacetaldehyde. ANOVA confirmed that these metabolites displayed significantly different mean-relative intensities in the two breeds, with the levels of phenolics in urine being much greater in Labrador retrievers (Fig. 6). As anticipated from metabolite fingerprinting results, raffinose, kestose and glycerol levels
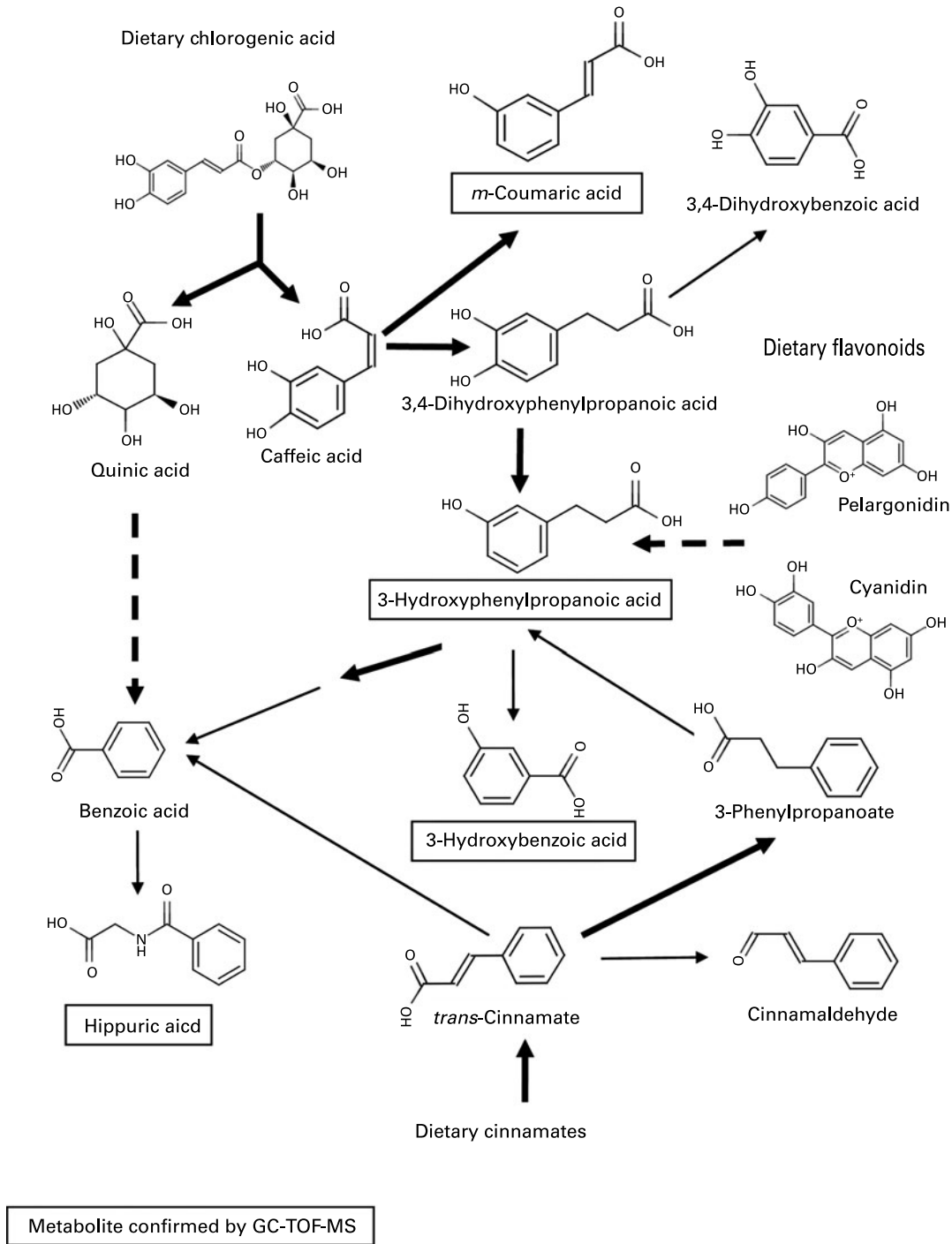
Fig. 5. Predicted metabolism of dietary polyphenols to explain urine differences between Labrador retrievers and miniature Schnauzers. The metabolic pathway shown is a 'best-fit' model, which accounts for a large number of the explanatory *m/z* signals in flow injection electrospray MS fingerprints tentatively annotated as derived from phenolic compounds. Metabolites in text boxes were confirmed by GC-MS. – – ►, multiple steps in a metabolic pathway; → , represent one-step biotransformations; ——►, represent steps likely to be carried out by colonic microflora; → represent endogenous biotransformations.

differed significantly between the two dog breeds (Table 2; Fig. 6). Additionally, citric acid was also highlighted as a discriminatory metabolite. In this case, ANOVA revealed that, although citrate levels generally appeared higher in miniature Schnauzers, the class means were rather similar and the upper quartile range was much greater, which was indicative of significant inter-dog variability (data not shown).

As hydroxylated butanoate derivatives were predicted as highly ranked alternatives to phenolics, the GC-TOF-MS chromatograms were re-examined manually for evidence of peak differences between dog classes. The level of 2-hydroxybutanoate and 3-aminoisobutyric acid was discovered to be very similar between breeds but demonstrated sex-specific concentration differences (Fig. 6).

**Table 2.** Metabolite peaks differing significantly between dog breeds in GC time-of-flight MS analysis of urine

| Peak identifier | Assignment by NIST database |
| --- | --- |
| 1 | Glycerol |
| 2 | Unknown |
| 3 | Unknown organic acid |
| 4 | 3-Hydroxybenzoic acid |
| 5 | Unknown |
| 6 | Unknown carbohydrate |
| 7 | 3-Hydroxyphenylpropanoic acid |
| 8 | Citric acid |
| 9 | *m*-Coumaric acid |
| 10 | Hippuric acid |
| 11 | Unknown carbohydrate |
| 12 | Unknown aromatic |
| 13 | Unknown |
| 14 | Unknown |
| 15 | Kestose |
| 16 | Raffinose |

NIST, National Institute of Standards and Technology.

### Breed discrimination by analysis of urine metabolome of pet dogs fed their normal diet

The study above indicated stable, breed-specific differences in the urine metabolome of dogs in a highly controlled experimental design. Further work was undertaken to assess the utility of metabolomics for animals from domestic homes, whose diet and prior behavioural activity were not specified. Ten young male dogs of each of ten breeds had urine sampled after catheterisation just before client-elected castration at a range of veterinary clinics in the US. Following FIE-MS analysis, all breeds were compared in a single LDA projection (Fig. 7(a)). Some clustering of breed samples is evident, but with the exception of Beagles in DF1, breed separation is not complete. Thus, breed-specific discrimination was weaker in this less experimentally controlled study.

We have shown previously that pairwise comparison of FIE-MS fingerprints by RF is a powerful way to classify different genotype classes in plant populations[35]. Both classification accuracy and the model margin were thus calculated for all possible pairwise combinations of breeds (Fig. 7(b)). A stratified cross-validation method was adapted to test modelling accuracies. Several pairwise comparisons have $> 70\%$ correct predictions but have differing margin estimates, which suggested that over-optimistic models were being generated. However, some pairwise comparisons have RF margins close to the threshold of an adequate model $(0.2)^{[38]}$ suggesting that breed-specific differences in urine metabolome do exist.

### Discussion

Obtaining biofluid samples such as urine or blood from domesticated animals demands that strict, ethically acceptable, procedures are followed and often require habituation of animals and intensive carer resource. Such stringent requirements naturally limit the number of sample replicates that can be obtained, which can be a major problem for 'omic' level studies[42,43]. The present work demonstrates that, despite such limitations on replication (in addition to the normal variability of urine as a homoeostatic fluid), the metabolite composition of fasted ($> 20$ h) 'spot' urine samples from two domestic dog breeds is consistent enough to allow successful classification based on metabolome fingerprint data. Using Labrador retrievers maintained in a simulated domestic setting, an experimental design in relation to husbandry, sampling and metabolite analysis was developed, which minimised the effect of external factors on the reproducibility of the metabolite fingerprint. With adequate procedures in place, comparison of the urinary chemical fingerprints of Labrador retrievers and miniature Schnauzers fed an identical diet suggested distinct, major differences centred largely on the metabolism of diet-related polyphenols. These data confirm the findings of a NMR study using identical samples[19]. Although breed discrimination was relatively accurate, it proved more difficult to correctly predict sex.

Preliminary examination of the urinary metabolome of young male dogs representative of a wider range of breeds from animals maintained under domestic conditions on unknown diets provided evidence that some breeds may have distinctive metabolic differences. For example, the present limited data revealed Beagle urine to be significantly different in composition from that of other breeds investigated, suggesting potentially distinctive interactions with diet. Furthermore, when breeds are designated as either small, medium or large in stature (see key in Fig. 7(a)), it is apparent that large breeds can be separated out from breeds of other sizes (Fig. 7(b)). Urine-specific gravity is higher, but frequency of urination and daily average urine output volumes are lower in miniature Schnauzers compared to Labrador retrievers[44], which may account for the fact that urea signals in FIE-MS were highly significant discriminatory signals between the two breeds. When considering a urine collection strategy, it may be beneficial to take into account, particularly any behavioural differences between breeds. Small breeds of dog generally have more concentrated urine and many are more prone to Ca stone formation in comparison to large breeds[45]. Urinary citric acid levels have long been viewed as a marker of renal metabolism[46] and have been correlated to net absorption of alkali in the genitourinary tract, which is reduced in animals at risk of forming Ca stones[45,47].

The demonstration that non-targeted metabolite profiling identified expected differences in citrate levels among breeds is an indication of the comprehensiveness of this 'first-pass' analysis method. However, this last observation does raise issues of data interpretability when dealing with semi-quantitative measurements. The data are presented as relative intensities (TIC ratio) and not as absolute amounts. Furthermore, FIE-MS data may be affected by ion suppression or enrichment within the urine matrix. So while citrate may be a significant peak in discriminating between the two breeds, it is not possible to conclude that there is a higher concentration of citrate in miniature Schnauzer urine in comparison to Labrador retriever without further quantitative analyses. The intensity value shown also provides no evidence of a biologically significant effect on maintaining genitourinary tract health. As such, while the observation is useful for hypothesis generation, data should be interpreted with caution.

From a practical perspective, the first-pass analysis of urine by FIE-MS fingerprinting was simple and rapid, producing
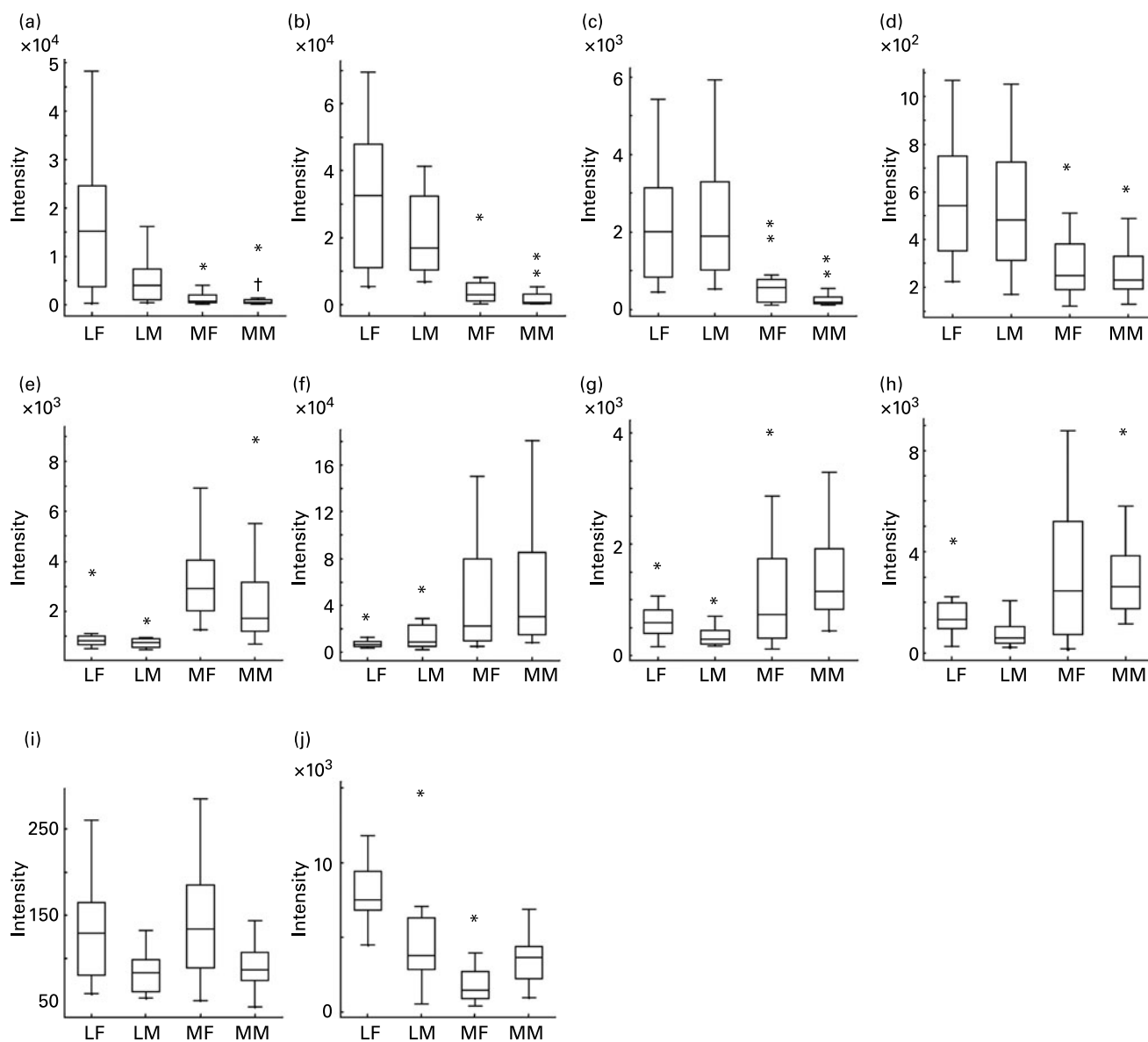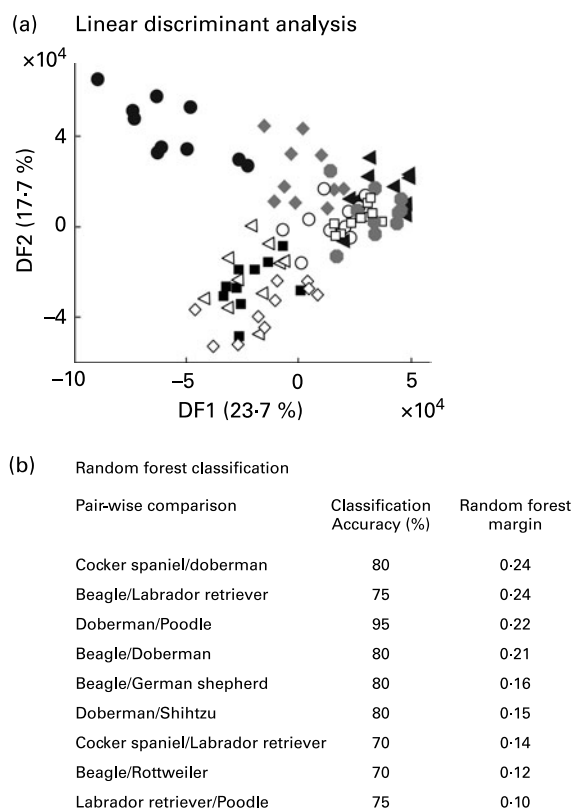
**Fig. 6.** Relative levels of metabolites discriminating urines of Labrador retrievers and miniature Schnauzers. The figure shows for each breed × sex combination the relative signal intensity ratios for ten specific metabolites as box plots derived from analysis of GC-MS data (Wilcoxon text; $P \leq 0.00001$). LF, Labrador retriever female; LM, Labrador retriever male; MF, miniature Schnauzer female; MM, miniature Schnauzer male. (a) Hippuric acid; (b) coumaric acid; (c) 3-hydroxyphenyl propanoic acid; (d) 3-hydroxybenzoic acid; (e) glycerol; (f) citric acid; (g) raffinose; (h) kestose; (i) 2-hydroxybutyric acid; (j) β-amino-isobutyric acid. *Outliers.

information-rich data with high potential for further interpretation[25,35,37]. A parallel study on the same urine samples using NMR fingerprinting revealed ten metabolites that differed in concentration between Labrador retrievers and miniature Schnauzers[19], but of these signals only citrate and hippurate were structurally assigned. The discriminatory power of these metabolites was indeed confirmed in the present study, along with a much larger number of 'explanatory' signals representing metabolites associated with several areas of metabolism. It is concluded that an approach based on MS complements metabolic fingerprinting by NMR, but additionally, by its increased sensitivity and opportunity for direct metabolite putative annotations based on $m/z$ value[37] offers increased scope for data modelling and interpretation before any deeper analysis by more targeted analytical methods.

Large differences in the levels of phenolic metabolites in urine were a major discriminatory feature in the specific comparison of Labrador retrievers and miniature Schnauzers. The most intense discriminatory phenolic metabolite peak in urine was associated with $m$-coumarate, which is likely to have its origin in the microbial breakdown of dietary chlorogenic acid, an abundant hydroxycinnamic acid found in a large variety of fruits and vegetables[48]. Similarly, the presence of elevated levels of 3-hydroxyphenylpropanoic acid in mammalian urine is commonly associated with degradation of dietary flavonoids (such as the anthocyanins, pelargonidin and cyanidin) and condensed tannins by the gut microflora[48,49]. 3-Hydroxybenzoic acid in the urine is likely the product of further endogenous β-oxidation of 3-hydroxyphenylpropanoic acid[50]. Benzoic acid, derived from both endogenous and

## (a) Linear discriminant analysis



## (b) Random forest classification

| Pair-wise comparison | Classification Accuracy (%) | Random forest margin |
|---|---|---|
| Cocker spaniel/doberman | 80 | 0·24 |
| Beagle/Labrador retriever | 75 | 0·24 |
| Doberman/Poodle | 95 | 0·22 |
| Beagle/Doberman | 80 | 0·21 |
| Beagle/German shepherd | 80 | 0·16 |
| Doberman/Shihtzu | 80 | 0·15 |
| Cocker spaniel/Labrador retriever | 70 | 0·14 |
| Beagle/Rottweiler | 70 | 0·12 |
| Labrador retriever/Poodle | 75 | 0·10 |

**Fig. 7.** Discrimination of male dogs by breed using FIE-MS fingerprinting. (a) Linear discriminant analysis scores plot of nine dog breeds (DF1 × DF2). (b) Summary of Random Forest performance (classification – accuracy and model margin) in pairwise comparison of urines from selected dog breeds using flow injection electrospray MS fingerprint data. Key: Breed (size): ◄, Doberman (L); ○, Rottweiler (L); ●, Labrador retriver (L); □, German sheperd (L); ■, Poodle (M); ◁, Shih tzu (S); ●, Beagle (M); ◇, Cocker spaniel (M); ◆, Golden retriver (L); where, S, small; M, medium; L, large.

gut microflora metabolism of dietary phenolics, is rapidly conjugated with glycine to produce hippuric acid[50].

There has been considerable literature advocating the potential beneficial effects on human health of dietary polyphenols with antioxidant properties, such as flavonoids[51]. Bioavailability of phenolics with desirable antioxidant properties may depend on the balance between absorption by intestinal epithelial mono-carboxylic acid transporters and metabolism by the gut microflora[50,52]. As illustrated in Fig. 5, the first steps in the degradation of dietary phenolics in the present study are likely to be the result of colonic microbial activity, which suggests that precursor molecules must be entering the large intestine in both breeds. Thus, the differences in concentration of phenolic breakdown products in urine could result from a more efficient absorption of intact dietary flavonoids and cinnamates into the blood stream in the upper gastrointestinal tract in miniature Schnauzers, as reported previously for chlorogenate in rats and human subjects[53]. Alternatively, the presence of much higher levels of phenolics in urine from Labrador retrievers suggests either a more efficient catabolism of diet-derived flavonoids and cinnamates by the colonic microflora or differential absorption in the colon. Whole grain wheat and maize, rice and sugarbeet pulp, together with bran derived from both maize and soya, comprise a large proportion of the raw ingredients used in the manufacture of the commercial dog foods used in the present study. As whole grains cereals and bran products contain high levels of polyphenols and are rich in cinnamates, coumarates, chlorogenate, caffeic acid and ferulate[52,54], then it is possible that the differential utilisation of these potentially 'bioactive' plant secondary metabolites might have implications for animal health[16] and thus commercial pet foods optimised for individual breeds, in some instances, may prove beneficial[7].

The major contribution of phenolic signals discriminating miniature Schnauzers and Labrador retrievers in classification models based on FIE-MS data was confirmed independently by GC-MS analysis. In addition to animal carcass meal, a significant component of commercial pet foods is derived from complex fat-based by-products, including liquid poultry fat, beef tallow and vegetable oils[55] together, with pure compounds such as glycerol that are used as moisture stabilisers[56]. Although differences in the levels of organic acids derived from butyric acid were more associated with sex, it is possible that the non-targeted profiling GC-MS methodology in the present study did not adequately cover all classes of organic acids and fatty acids. Indeed, a previous NMR study suggested that metabolites related to 2-hydroxybutyrate may differ in urines from these two breeds[19], and the present GC-MS data highlighted at least one further unknown organic acid as discriminatory. Thus, further differences in fatty acid metabolism as tentatively suggested by FIE-MS fingerprinting, particularly relating to butyrate derivatives or carnitine derivatives of larger fatty acids, may be significant between breeds. Related to this possibility, GC-MS analysis also revealed much higher levels of glycerol in urine derived from miniature Schnauzers. As diet-derived glycerol is a rich energy source, it would be surprising if this compound was excreted unless supply is greatly in excess of need, and so this observation, together with possible differences in fatty acid excretion, may be worthy of further investigation. The levels of the trisaccharides raffinose and kestose followed the same trend as glycerol, being present at substantially higher levels in miniature Schnauzers. Raffinose and kestose represent part of the indigestible 'soluble fibre' components of animal diets that are only degraded by microbial activity in the lower gastrointestinal tract. Thus, excretion of these trisaccharides may again reflect stable and substantial differences in the microbial populations within the gastrointestinal tract of these two breeds[57–59]. This hypothesis is supported by research characterising the canine gut microbiome (C. Wallis, unpublished results). In that study, bacterial populations in faecal samples from different breeds have been shown to differ in the proportion of the major phyla identified. As inulin-type metabolite prebiotics are now part of specialist commercial animal feeds available on the market[60,61], then knowledge relating to such potential breed-specific differences may prove important to optimise diets and improve health.

It is concluded that substantial potential exists to investigate differences in nutrition-related metabolism by urine metabolite fingerprinting, especially if the animals can be maintained on the same diet before sampling, similar to human studies[31]. Such information could be valuable for developing diets designed to suit classes of dog breeds[7,19,59].

## Acknowledgements

## References

1. Sutter NB & Ostrander EA (2004) Dog star rising: the canine genetic system. *Nat Rev Genet* **5**, 900–910.
2. Lindblad-Toh K, Wade CM, Mikkelsen TS, *et al.* (2005) Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**, 803–819.
3. Yuzbasiyan-Gurkan V, Blanton SH, Cao YY, *et al.* (1997) Linkage of a microsatellite marker to the canine copper toxicosis locus in Bedlington Terriers. *Am J Vet Res* **58**, 23–27.
4. Safra N, Ling GV, Schaible RH, *et al.* (2005) Exclusion of urate oxidase as a candidate gene for hyperuricosuria in the Dalmatian dog using an interbreed backcross. *J Hered* **96**, 750–754.
5. Ko KS, Backus RC, Berg JR, *et al.* (2007) Differences in taurine synthesis rate among dogs relate to differences in their maintenance energy requirement. *J Nutr* **137**, 1171–1175.
6. Cruz F, Vila C & Webster MT (2008) The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol Biol Evol* **25**, 2331–2336.
7. Swanson KS, Schook LB & Fahey GC (2003) Nutritional genomics: implications for companion animals. *J Nutr* **133**, 3033–3040.
8. Weber MP, Stambouli F, Martin LJ, *et al.* (2002) Influence of age and body size on gastrointestinal transit time of radiopaque markers in healthy dogs. *Am J Vet Res* **63**, 677–682.
9. Brown WY, Vanselow BA & Walkden-Brown SW (2003) One dog's meat is another dog's poison – nutrition in the Dalmatian dog. *Recent Adv Animal Nutr Aust* **14**, 123–131.
10. German JB, Roberts MA & Watkins SM (2003) Personal metabolomics as a next generation nutritional assessment. *J Nutr* **133**, 4260–4266.
11. Kaput J, Ordovas JM, Ferguson L, *et al.* (2005) The case for strategic international alliances to harness nutritional genomics for public and personal health. *Br J Nutr* **94**, 623–632.
12. Meyer H, Zentek J, Habernoll H, *et al.* (1999) Digestibility and compatibility of mixed diets and faecal consistency in different breeds of dog. *J Vet Med A* **46**, 155–165.
13. Stevenson AE, Robertson WG & Markwell P (2003) Risk factor analysis and relative supersaturation as tools for identifying calcium oxalate stone-forming dogs. *J Small Anim Pract* **44**, 491–496.
14. Benyacoub J, Czarnecki- Maulden GL, Cavadini C, *et al.* (2003) Supplementation of food with *Enterococcus faecium* (SF68) stimulates immune functions in young dogs. *J Nutr* **133**, 1158–1162.
15. Mata-Bilbao MD, Andres-Lacueva C, Roura E, *et al.* (2007) Absorption and pharmacokinetics of grapefruit flavanones in Beagles. *Br J Nutr* **98**, 86–92.
16. Blaut M & Clavel T (2007) Metabolic diversity of the intestinal microbiota: implications for health and disease. *J Nutr* **137**, 751S–755S.
17. Wishart DS, Tzur D, Knox C, *et al.* (2007) HMDB: The Human Metabolome Database. *Nucleic Acids Res* **35**, D521–D526.
18. Human Metabolome Database (2009) http://hmdb.ca (accessed April 2009)
19. Viant MR, Ludwig C, Rhodes S, *et al.* (2007) Validation of a urine metabolome fingerprint in dog for phenotypic classification. *Metabolomics* **3**, 453–463.
20. Saude EJ & Sykes BD (2007) Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics* **3**, 19–27.
21. Stella C, Beckwith-Hall B, Cloarec O, *et al.* (2006) Susceptibility of human metabolic phenotypes to dietary modulation. *J Proteome Res* **5**, 2780–2788.
22. Wilson ID, Plumb R, Granger J, *et al.* (2005) HPLC-MS-based methods for the study of metabonomics. *J Chromatogr B* **817**, 67–76.
23. Gika HG, Theodoridis GA, Wingate JE, *et al.* (2007) Within-day reproducibility of an HPLC-MS-based method for metabonomic analysis: application to human urine. *J Proteome Res* **6**, 3291–3303.
24. Zhang Q, Wang G, Du Y, *et al.* (2007) GC/MS analysis of the rat urine for metabonomic research. *J Chromatogr B* **854**, 20–25.
25. Beckmann M, Parker D, Enot DP, *et al.* (2008) High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nat Protoc* **3**, 486–504.
26. Goodacre R, Shann B, Gilbert RJ, *et al.* (2000) Detection of the dipicolinic acid biomarker in Bacillus spores using Curie-point pyrolysis mass spectrometry and Fourier transform infrared spectroscopy. *Anal Chem* **72**, 119–127.
27. Griffin JL (2003) Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* **7**, 648–654.
28. Wang Y, Lawler D, Larson B, *et al.* (2007) Metabonomic investigations of aging and caloric restriction in a life-long dog study. *J Proteome Res* **6**, 1846–1854.
29. Cloarec O, Dumas M-E, Craig A, *et al.* (2005) Statistical total correlation spectroscopy: an exploratory approach for latent biomarker identification from metabolic $^1$H NMR data sets. *Anal Chem* **77**, 1282–1289.
30. Walsh MC, Brennan L, Malthouse JPG, *et al.* (2006) Effect of acute dietary standardization on the urinary, plasma, and salivary metabolomic profiles of healthy humans. *Am J Clin Nutr* **84**, 531–539.
31. Walsh MC, Brennan L, Pujos-Guillot E, *et al.* (2007) Influence of acute phytochemical intake on human urinary metabolomic profiles. *Am J Clin Nutr* **86**, 1687–1693.
32. Allen J, Davey HM, Broadhurst D, *et al.* (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* **21**, 692–696.
33. Scholz M, Gatzek S, Sterling A, *et al.* (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics* **20**, 2447–2454.

34. Catchpole GS, Beckmann M, Enot DP, *et al.* (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc Natl Acad Sci USA* **102**, 14458–14462.

35. Enot DP, Beckmann M, Overy D, *et al.* (2006) Predicting interpretability of metabolome models based on behavior, putative identity, and biological relevance of explanatory signals. *Proc Natl Acad Sci USA* **103**, 14865–14870.

36. Beckmann M, Enot DP, Overy DP, *et al.* (2007) Representation, comparison, and interpretation of metabolome fingerprint data for total composition analysis and quality trait investigation in potato cultivars. *J Agric Food Chem* **55**, 3444–3451.

37. Overy DP, Enot DP, Tailliart K, *et al.* (2008) Explanatory signal interpretation and metabolite identification strategies for nominal mass FIE-MS metabolite fingerprints. *Nat Protoc* **3**, 471–485.

38. Enot DP, Lin W, Beckmann M, *et al.* (2008) Preprocessing, classification modeling and feature selection using flow injection electrospray mass spectrometry metabolite fingerprint data. *Nat Protoc* **3**, 446–470.

39. Thomaz CE, Boardman JP, Hill DLG, *et al.* (2004) Using a maximum uncertainty LDA-based approach to classify and analyse MR brain images. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pp. 291–300. Berlin: Springer.

40. Enot DP & Draper J (2007) Statistical measures for validating plant genotype similarity assessments following multivariate analysis of metabolome fingerprint data. *Metabolomics* **3**, 349–355.

41. Sing T, Sander O, Beerenwinkel N, *et al.* (2005) ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941.

42. Somorjai RL, Dolenko B & Baumgartner R (2003) Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. *Bioinformatics* **19**, 1484–1491.

43. Berrar D, Bradbury I & Dubitzky W (2006) Avoiding model selection bias in small-sample genomic datasets. *Bioinformatics* **22**, 1245–1250.

44. Stevenson AE & Markwell PJ (2001) Comparison of urine composition of healthy Labrador retrievers and miniature Schnauzers. *Am J Vet Res* **62**, 1782–1786.

45. Stevenson AE, Hynds WK & Markwell PJ (2003) Effect of dietary moisture and sodium content on urine composition and calcium oxalate relative supersaturation in healthy miniature schnauzers and Labrador retrievers. *Res Vet Sci* **74**, 145–151.

46. Simpson DP (1983) Citrate excretion – a window on renal metabolism. *Am J Physiol* **244**, F223–F234.

47. Hess B, Michel R, Takkinen R, *et al.* (1994) Risk-factors for low urinary citrate in calcium nephrolithiasis – low vegetable fiber intake and low urine volume to be added to the list. *Nephrol Dial Transplant* **9**, 642–649.

48. Gonthier MP, Cheynier V, Donovan JL, *et al.* (2003) Microbial aromatic acid metabolites formed in the gut account for a major fraction of the polyphenols excreted in urine of rats fed red wine polyphenols. *J Nutr* **133**, 461–467.

49. Deprez S, Brezillon C, Rabot S, *et al.* (2000) Polymeric proanthocyanidins are catabolized by human colonic microflora into low-molecular-weight phenolic acids. *J Nutr* **130**, 2733–2738.

50. Gonthier MP, Verny MA, Besson C, *et al.* (2003) Chlorogenic acid bioavailability largely depends on its metabolism by the gut microflora in rats. *J Nutr* **133**, 1853–1859.

51. Scalbert A, Manach C, Morand C, *et al.* (2005) Dietary polyphenols and the prevention of diseases. *Crit Rev Food Sci Nutr* **45**, 287–306.

52. Fardet A, Llorach R, Orsoni A, *et al.* (2008) Metabolomics provide new insight on the metabolism of dietary phytochemicals in rats. *J Nutr* **138**, 1282–1287.

53. Lafay S, Gil-Izquierdo A, Manach C, *et al.* (2006) Chlorogenic acid is absorbed in its intact form in the stomach of rats. *J Nutr* **136**, 1192–1197.

54. Kim K-H, Tsao R, Yang R, *et al.* (2006) Phenolic acid profiles and antioxidant activities of wheat bran extracts and the effect of hydrolysis conditions. *Food Chem* **95**, 466–473.

55. Nguyen P, Dumon H, Biourge V, *et al.* (1998) Glycemic and insulinemic responses after ingestion of commercial foods in healthy dogs: influence of food composition. *J Nutr* **128**, 2654S–2658S.

56. Warmbier HC, Schnickels RA & Labuza TP (1976) Effect of glycerol on nonenzymatic browning in a solid intermediate moisture model food system. *J Food Sci* **41**, 528–531.

57. Rastall RA (2004) Bacteria in the gut: friends and foes and how to alter the balance. *J Nutr* **134**, 2022S–2026S.

58. Simpson JM, Martineau B, Jones WE, *et al.* (2002) Characterization of fecal bacterial populations in canines: effects of age, breed and dietary fiber. *Microb Ecol* **44**, 186–197.

59. Suchodolski JS, Ruaux CG, Steiner JM, *et al.* (2004) Application of molecular fingerprinting for qualitative assessment of small-intestinal bacterial diversity in dogs. *J Clin Microbiol* **42**, 4702–4708.

60. Verdonk J, Shim SB, van Leeuwen P, *et al.* (2005) Application of inulin-type fructans in animal feed and pet food. *Br J Nutr* **93**, S125–S138.

61. Strickling JA, Harmon DL, Dawson KA, *et al.* (2000) Evaluation of oligosaccharide addition to dog diets: influences on nutrient digestion and microbial populations. *Anim Feed Sci Technol* **86**, 205–219.