Dominant components of the Thoroughbred metabolome characterised by 1H-NMR spectroscopy: a metabolite atlas of common biofluids

Article

Accepted Version

Dominant components of the Thoroughbred metabolome characterised by ^{1}H-NMR spectroscopy: A metabolite atlas of common biofluids.

E. Escalona*[1], J. Leng[2], A. Dona[1], C. Merrifield[2], E. Holmes[1], C. Proudman[3] and J. Swann[2]

[1]Section Computational and Systems Medicine, Division of Surgery and Cancer, Faculty of Medicine, Imperial College London, Sir Alexander Fleming Building, London SW7 2AZ, UK;

[2]Department of Food and Nutritional Sciences, University of Reading, Whiteknights, Reading, Berkshire, RG6 6AH, UK;

[3] Department of Gastroenterology/School of Veterinary Medicine, University of Liverpool, Neston CH64 7TE, UK.

*Corresponding author email: e.escalona10@imperial.ac.uk

Keywords: horse; metabonomics; metabolomics; metabolites; biofluids; nuclear magnetic resonance

Summary

Reasons for performing study: Metabonomics is emerging as a powerful tool for disease screening and investigating mammalian metabolism. This study aims to create a metabolic

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/evj.12333

framework by producing a preliminary reference guide for the normal equine metabolic milieu.

Objectives: To metabolically profile plasma, urine and faecal water from healthy racehorses using high resolution ^{1}H-NMR spectroscopy and to provide a list of dominant metabolites present in each biofluid for the benefit of future research in this area.

Study design: This study was performed using seven Thoroughbreds in race training at a single time-point. Urine and faecal samples were collected non-invasively and plasma was obtained from samples taken for routine clinical chemistry purposes.

Methods: Biofluids were analysed using ^{1}H-NMR spectroscopy. Metabolite assignment was achieved *via* a range of 1D and 2D experiments.

Results: A total of 102 metabolites were assigned across the three biological matrices. A core metabonome of 14 metabolites was ubiquitous across all biofluids. All biological matrices provided a unique window on different aspects of systematic metabolism. Urine was the most populated metabolite matrix with 65 identified metabolites, 39 of which were unique to this biological compartment. A number of these were related to gut microbial host co-metabolism. Faecal samples were the most metabolically variable between animals; acetate was responsible for the majority (28%) of this variation. Short chain fatty acids were the predominant features identified within this biofluid by ^{1}H-NMR spectroscopy.

Conclusions: Metabonomics provides a platform for investigating complex and dynamic interactions between the host and its consortium of gut microbes and has the potential to uncover markers for health and disease in a variety of biofluids. Inherent variation in faecal extracts along with the relative abundance of microbial-mammalian metabolites in urine and invasive nature of plasma sampling, infers that urine is the most appropriate biofluid for the purposes of metabonomic analysis.

46

Introduction

Metabonomics is a powerful systems biology approach that aims to simultaneously measure all the low molecular weight metabolites present in a biofluid or tissue. This approach to global untargeted characterisation of the metabolic phenotype allows the study of multidimensional biochemical responses of complex biological systems to genetic or environmental stimuli [1]. Metabolic profiling captures information from both intrinsic (genetics, protein expression) and environmental inputs (diet, gut microbiota), providing holistic information on the global system. This strategy has proven highly effective for unravelling the complex metabolic interactions between the mammalian host and its resident gut microbiota. Metabonomics is a tool of particular interest to equine researchers given the vast impact of the equine gut microbiome on the bioavailability of food, medication and energy. Metabonomics, along with other 'omic' technologies such as genomics, proteomics and transcriptomics is increasingly showing potential in clinical settings as both a screening tool and a means for mechanistic elucidation of disease pathways [2-4].

To date, there are fewer metabonomic studies exploring veterinary concerns than there are human and rodent studies. The majority of mammalian work has concentrated on laboratory animals [5-7] and animal models with high translatability to human health such as the pig [8; 9]. Less attention has been given to herbivorous hind-gut fermenters and the majority of equine metabolic work concentrates on drug detection within the racing industry [10; 11]. Equine-specific metabonomic studies include the use of the horse as a model for the metabolic response to a dextrose challenge in type-2 diabetes [12] and a laminitic plasma NMR study using an oligofructose overload model [13]. Other equine studies include metabolic analysis of biofluids in response to age [14] and osteochondrosis [15]. Pappalardo

et al. have explored metabolic variation in association with breed of horse and importantly revealed significant differences that are likely to be attributed to differing growth rates and protein utilisation [16]. These bodies of equine metabonomic work address the enormous potential for exploring normal equine physiology and pathology-based variation. However, there is still a paucity of baseline data on the metabolic phenotype of horses and this study serves as a reference tool for the Thoroughbred racehorse for clinicians wishing to use metabonomic technologies to complement their research in either a diagnostic or mechanistic capacity.

Two major analytical platforms are commonly used for metabolic profiling, nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS). These approaches simultaneously capture quantitative information from a range of low molecular weight metabolites across various different sample types. In metabonomic studies, NMR and MS are used both independently and in tandem to achieve an in-depth coverage of the metabolome. In this study we comprehensively characterise the dominant features within the urinary, plasma and faecal metabolomes of Thoroughbred racehorses using ^{1}H-NMR spectroscopy, as this provides a reliable, reproducible [17] screening tool and is non-destructive of samples. Thoroughbred racehorses represent a highly uniform equine population both genetically and environmentally. Normal clinical chemistry parameters exist for this equine subtype to assist racing industry standards and this study will build upon data already in the literature and in clinical use.

Materials and methods

Sample collection

From a large–scale study of urinary metabonomics of in-training Thoroughbred across a number of yards, 7 animals were selected at random to provide a baseline metabolic reference point for future comparison. Due to ethical constraints of invasive sampling and logistical limitations of simultaneous collection of all 3 biofluids, 7 animals were selected as representative of a wider population of animals. Early morning free-catch urine and faecal samples were collected into sterile plastic containers and snap-frozen in liquid nitrogen within 2 h. Plasma was obtained from excess clinical samples. Plasma samples were collected into heparinised tubes, spun down to obtain the plasma fraction and snap-frozen within 2 h of collection. All samples were then stored at -80°C prior to NMR analysis. Comprehensive metadata for each horse is shown in Supplementary Item 1 and samples were consistently taken before a morning concentrate feed.

Sample preparation

Plasma and urine sample preparation was carried out as described by Beckonert *et al*. 2007[18].

Faecal samples (100 mg) were combined with 1.7 mm Zirconia beads and 1 mL of distilled water and homogenised in a bead-beater for 10 min and centrifuged at 13,000 *g* for 10 min. Water was evaporated from the samples using vacuum concentrator (Speed-Vac) and then reconstituted in 700 µL phosphate buffer (pH 7.4; 100% D_2O) containing 1 mM sodium 3-trimethylsilyl-1-[2,2,3,3-2H4]propionate (TSP).

Combined sampling approach for 2D NMR Experiments

2D spectra were obtained from pooled samples for each sample type to ensure comprehensive capture of metabolites. 50 µL of urine was pooled from each horse before the addition of 500

μL of phosphate buffer (and 500 μL of the resulting volume was added to the 5 mm NMR tube). The same process was repeated for the plasma samples with the addition of 500 μL of D_2O saline solution. For the faecal samples, individual samples were processed as previously described and then pooled after reconstitution with 100 μL of phosphate buffer.

Acquisition of 1D ^{1}H-NMR spectra

Spectroscopic analysis of all samples was carried out on a 700 MHz Bruker NMR spectrometer[b] operating at 300K and equipped with a 5 mm ^{1}H(^{13}C/^{15}N) inverse cryoprobe.

Urine and faecal samples

Standard one-dimensional ^{1}H-NMR spectra were acquired for all urine and faecal samples. We employed a standard one-dimensional pulse sequence (noesypr1d) that employs the first increment of a NOE sequence to achieve suppression of the water resonance with water peak suppression using a standard pulse sequence [19]. For each sample, 8 dummy transients were followed by 256 transients and collected in 64K data point. Irradiation of the solvent (D_2O) resonance was applied during presaturation delay (2.0 s) for all spectra. The pulse sequence parameters including the 90° pulse, receiver gain and pulse powers were optimised for each sample set run. The spectral width was 20 ppm for all spectra. The free induction decay (FID) was processed with an exponential line broadening of 0.5 Hz prior to Fourier transformation.

Plasma samples

Water-suppressed Carr-Purcell-Meiboom-Gill (CPMG) spin-echo spectra were acquired for the plasma samples, Here, 8 dummy scans followed by 256 scans were acquired for each sample in 64k data points with a total spin-spin relaxation delay of 1.5 s and a total delay between pulse cycles of 4.85 s.

Acquisition of 2D ^{1}H-NMR spectroscopy was undertaken with an 800 MHz Bruker NMR spectrometer[b] operating at 300K and equipped with a triple-resonance probe (TXI). J-

resolved spectroscopy (J-res) spectra were acquired from all biofluid composite samples using 64 transients per increment with 160 increments in the second dimension. The F1 (J-coupling) domain spectral width covered 120 Hz. Prior to the double FT and magnitude calculation, the F1 data was zero-filled to 1024 points. The spectra were then tilted by 45° to provide orthogonality of the chemical shift and coupling constant axes and subsequently symmetrised about the F1 axis. ^{1}H-^{1}H Correlation Spectroscopy (COSY) was performed on all 3 types of pooled biofluid samples in order to detect correlations between protons on adjacent carbons. Transients were acquired with 4096 data points (sweep width of 7200 Hz in both axis) with 64 scans per increment and 320 increments in the F1 axis. The relaxation between successive pulse cycles was 2.3 s and were weighted using a sine bell function in T_1 and T_2 prior to fourier transformation and subsequently symmetrised about the diagonal axis. Two-dimensional echo/anti-echo ^{1}H–^{13}C heteronuclear single quantum correlation (HSQC) spectra were also obtained. 256 scans were collected (16 dummy scans) at a spectral resolution of 4k in F2 across a spectral width of 12 ppm for ^{1}H and 170 ppm for the ^{13}C axes. An acquisition time of 0.852 s and a relaxation delay of 1.2 s were used and delays were set for a 145 Hz one bond ^{1}H–^{13}C coupling constant. Spectra were zero-filled in the F2 dimension by a factor of 2 to 8k, and zero-filling and linear prediction was applied in F1 to result in a resolution of 1k.

NMR spectral data pre-processing

Data [-1.0 to 10.0 ppm] were imported into MatLab environment (7.0 The Mathsworks[c]), where they were automatically phased, baseline-corrected and referenced to TSP (δ 0.00) or glucose (δ 5.233) for plasma using scripts written in-house [20]. To reduce analytical variation between samples the residual water signal (4.67 – 4.98 ppm) was truncated from the

data set. Probabilistic-quotient normalisation was used on each biofluid class separately to account for differing sample dilutions [21].

Data analysis of biological matrices

Unsupervised multivariate analysis was undertaken to visualise clustering and differences between samples [22]. Principal component analysis (PCA) was constructed using unit-variance scaled data (UV) [23]..

Metabolite assignment of endogenous metabolites was made by compiling the following information from each peak: the chemical shift and relative integral height from 1D spectra, the multiplicity using J-res spectra, the proton coupling information from COSY spectra and carbon shifts from HSQC spectra. This information was then used to search for matching metabolites from in-house databases, online databases (http://www.hmdb.ca/) and reference to published literature data [24-26]. Statistical Total Correlation Spectroscopy (STOCSY) was also employed to aid metabolite identification [27]. This method uses statistical connectivity between data points within a spectral profile. However, unlike 2D-NMR this method will also pick up metabolites involved in the same pathways due to biological covariance [28].

Results

Assignment of dominant metabolites across easily obtainable biofluids yields a preliminary equine metabolic atlas

The majority of known NMR-detectable metabolites were assigned in each biological matrix (Fig 1), with numerical reference to Table 1. Metabolite assignments were performed using two-dimensional correlation (COSY) and J-resolved (J-res) spectroscopy to ascertain peak

multiplicity, coupling constants and to overcome peak overlap. This is demonstrated in the COSY (Fig 2A) and the J-res (Fig 2B). Heteronuclear Single Quantum Coherence (HSQC) was also performed to provide ^{13}C shifts and confirm assignments. Putative metabolite assignment was made using ^{1}H and ^{13}C chemical shifts, peak multiplicity, coupling constants and relative peak integrals. Overall, 102 metabolites were identified by ^{1}H-NMR in the 3 biofluids (Table 1). Detailed assignment information is shown in Supplementary Item 2.

Cross-compartmental analysis revealed a core metabonome, along with compartment specific metabolites

Metabolites assigned to the 3 biofluids were compared to ascertain ubiquitous metabolites and those that were specific to each biological compartment. The metabolic variation across the biological matrices is displayed in a Venn diagram (Fig 3) to easily visualise inter-compartmental overlap. A total of 14 metabolites were ubiquitous to all biofluids, which we will refer to as 'core' metabolites. These include energy-related metabolites such as glucose and lactate as well as a number of amino-acids including alanine, arginine, glycine, glutamine, taurine, threonine and valine. The microbial related metabolites acetate, formate and *p*-hydroxyphenylacetate were also conserved across all biofluids studied.

PCA revealed that inter-animal metabolic variation was lower than the variation between the different biofluids as visualised *via* clustering in the PCA scores plot (Fig 4A). As would be expected, the PCA samples cluster based on biofluid type. However, importantly the degree of clustering is different amongst biological matrices. Faecal samples demonstrate that they are inherently variable compared to either matched urine or plasma samples. The faecal metabonome displayed relatively higher concentrations of SCFA (butyrate, acetate and propionate) compared to plasma and urine. Urine samples had relatively higher level of gut-

214 microbial co-metabolites such as hippurate, phenylacetylglycine (PAG), *p*-cresyl sulfate and
215 trimethylamine-*N*-oxide (TMAO). In contrast, energy-related metabolites such as glucose and
216 pyruvate were observed in relatively higher concentrations in plasma compared to urine.
217 Variance plots shown in Fig 4A and B display the mean spectra of faeces and urine
218 respectively coloured by the variance, represented as a percentage of the total variance. Here,
219 creatinine can be seen to represent over 10% of the total variance seen within the urinary
220 profiles (Fig 4B) and acetate accounted for over 28% of the total variance observed in the
221 faecal profile (Fig 4A) compared to 1.8% in the plasma profile (not shown).
222
223 *Plasma*
224 Equine plasma contained relatively few metabolites compared to urine and faeces when
225 analysed by NMR spectroscopy, partly as a result of the overlap of signals from
226 macromolecular components such as lipoproteins and low molecular weight chemicals. A
227 total of 38 metabolites were identified (Table 1), and of these 12 were unique to this biofluid,
228 including the amino-acids tyrosine and phenylalanine and ketone bodies (α- and β-
229 hydroxybutyrate). Plasma was observed to contain the highest levels of glucose compared to
230 other biofluids within the data matrix.
231 *Urine*
232 The equine urinary metabolic profile was the most metabolically abundant of all biofluids
233 measured (Table 1). In total 65 metabolites were identified. Of these, 11 were unique to urine
234 and plasma, 3 were present in both urine and faeces and 39 metabolites were specific to
235 urine. Urinary-specific metabolites included a number of aromatic compounds that arise from
236 microbial-host co-metabolism. Hippurate, PAG, *p*-cresyl glucuronide and sulphate were

notably prominent in the aromatic region of the horse urine spectra (Fig 1B). Other urine specific metabolites of note include dietary compounds such as proline betaine.

Faeces

The faecal metabotype contained 43 metabolites, including 21 specific to faeces (Fig 3). These metabolites include microbial fermentation products, SCFA (butyrate and propionate), and a number of dietary metabolites (caprylate and arabinose, maltose, glycerol and xylose). Acetate was present in significantly greater concentrations than other metabolites.

Discussion

This is the first systematic description of the dominant metabolites of the healthy Thoroughbred racehorse and is important in establishing a metabolic reference from which to compare pathology-related variation. Characterisation of the equine metabolome highlights the diversity of information available in different biological matrices and as such provides a useful guide for researchers. Importantly, as has been shown in other species, urine provides a biological window into host-microbial metabolic interactions in the horse [8]. Although faeces may be considered a more direct representation of microbial metabolism, Fig 4A and B highlights the relatively greater variation in the faecal metabolic profiles compared to plasma and urine. Additionally, microbial-derived compounds absorbed from the gut are commonly not well represented in the faecal signature rendering the biological usefulness of this biofluid in gastrointestinal disease and mammalian-microbial co-metabolsim questionable.

A ubiquitous metabonome is present amongst mammals. The core equine metabonome comprises metabolites present in all 3 biofluids studied. 14 metabolites were identified including amino-acids, gut microbial metabolites and energy metabolites. Many of these

metabolites are ubiquitous, with varied functions and located in numerous tissue types. Similar findings were reported by Merrifield *et al*. with 22 metabolites shared across 4 biological matrices in pigs [8].

Marked ^{1}H-NMR metabolic similarities between horses and humans are suggested by observations of a number of shared metabolites. In total, there were 32 plasma, 23 urine and 27 faecal NMR detectable compounds common to both horses and humans [29-31]. Metabolic consistency was strongest across the plasma profiles. This validates the concept of conservation of mammalian physiology across species and that homeostatic metabolic control is tight in both systems. A lower percentage of the equine faecal and urine metabolic profiles were shared with humans but this was to be expected due to different digestive systems and metabolic pathways. In contrast, ruminants share a number of gut microbial co-metabolites such as hippurate and PAG [32-34]. These herbivores have similar digestive strategies to obtain nutrients from a cellulose-rich diet.

Excretory biofluids (urine and faeces) are the most metabolite-rich. The environmentally-determined nature of these biofluids renders them under less tight homeostatic control than that of the plasma. However, urine was also remarkably tightly controlled, given the relatively homogeneous genetic and environmental backgrounds of Thoroughbred racehorses. Differing creatinine excretion in one animal (due to the sample being collected after exercise) was responsible for the increased inter-animal variation observed (Fig 4A). In urine, 39 metabolites were identified as being unique whereas 21 metabolites were faeces-specific and 12 metabolites were plasma-specific. However, there were a small number of low-concentration unassigned metabolites that have not been included in these counts. All 3 biofluids are likely to contain thousands of metabolites which are too dilute to produce a significant NMR signal or were not detected by NMR in this study, but we feel we were able

285 to capture a good representation of the normal equine metabonome and this has been verified
286 by our ongoing work on a larger cohort of animals (data not shown).
287 Inter-compartmental variation is greater than inter-animal variation and is consistent with
288 other mammalian studies [8; 35]. Faecal samples varied from other biological matrices due to
289 the presence of a number of SCFAs, formate and isovalerate. SCFAs are the product of gut-
290 microbial fermentation of dietary fibre and contribute up to 70% of a horse's energy
291 requirements [36; 37]. Acetate level variability is likely to be due to differing bacterial
292 communities and consequently SCFA production. Collection time in relation to feeding time
293 is known to exert an effect on SCFA levels in equine faeces [38]. Samples were consistently
294 taken before a morning feed. However, this variability could be due to the difference in
295 individual intestinal transit times [39; 40]. Other SCFAs were not seen to vary to the same
296 extent as acetate. Butyrate is the main energy source for colonocytes and hence is likely to
297 have been utilised rather than excreted [41]. Butyrate and propionate are extensively
298 metabolised by first-pass metabolism and therefore absent in NMR detectable quantities in
299 urine and plasma samples. Creatinine was the greatest source of variation amongst urine
300 samples. Creatinine is a waste product of muscle metabolism formed from creatine in order to
301 maintain ATP levels during exercise [42]. The concentration of urine and thus metabolites
302 can change dramatically. Normalisation of the data prior to analysis helps to minimise
303 spectral anomalies caused by differences in urinary dilution. Creatinine levels can vary
304 according to factors such as muscle mass, physical exercise, diet, age and muscle damage
305 from previous strenuous exercise [43].
306 The urine metabolome provides a metabolic window into gut microbial co-metabolism.
307 Urine was found to differ from the other biological matrices due to the presence of a number
308 of gut-microbial co-metabolites (hippurate, PAG and *p*-cresyl sulphate and TMAO). These

309 compounds originate from exogenous sources (microbial and dietary) and are incorporated
310 into the host circulation after absorption. They subsequently undergo enzymatic conjugation
311 in the liver and gut mucosa to increase their polarity and enable renal excretion. In mammals,
312 enterohepatic recycling means further metabolism can occur at the gut-level [44]. Hippurate
313 is an aromatic compound predominantly formed from glycine conjugation of dietary or
314 microbial benzoate and PAG is a glycine-conjugated microbial metabolite of phenylalanine
315 metabolism (glutamine in higher apes and humans) [45]. *P*-cresol is formed from bacterial
316 degradation of tyrosine and is subsequently sulfated or glucuronidated in the liver or gut [46].
317 Both forms of conversion occur in rodents and horses, whereas humans predominately form
318 sulphate conjugates and pigs predominately glucuronidate the cresol molecule.
319 Trimethylamine (TMA) is a microbial degradation product of dietary choline; this metabolite
320 is absorbed from the gut and subsequently oxidised in the liver to produce TMAO [47]. A
321 number of anaerobic bacterial populations are known to produce these metabolites including
322 clostridia [48-50]. This taxonomic classification of bacteria has been associated with
323 intestinal disease in horses as well as being part of the normal microflora [51-54].
324 ^{1}H NMR spectroscopy is a robust method for assessing the inter-animal variation in
325 Thoroughbred racehorses. It is important to address potential sources of variation when
326 assessing metabonomic studies. The sensitive nature of such investigations necessitates the
327 collection of metadata to help explain possible variation between the samples. This
328 information should include details from sample collection, sample storage and run order, as
329 well as information relating to sample subject (health status, age, sex for example). The effect
330 of these is widely reported in the literature [35; 55; 56]. Although the 7 horses were taken
331 from 2 different yards, the samples clustered tightly, highlighting the metabolic uniformity of
332 these Thoroughbred racehorses despite differing age ranges, location and stages of fitness

(see Supplementary Item 1). This study was intended to create a reference tool for research into Thoroughbred racehorse metabolism and microbial co-metabolism. Since the samples from this first pilot clustered closely for all 3 biofluids indicating that the dominant metabolites visible by NMR were conserved across animals, it was deemed unnecessary to collect samples from further animals to minimise unnecessary sampling. Although, the sample size in this study is small and only one metabolic snap shot was taken for each horse, the fact that inter-animal variation in metabolic profiles was low, suggests that 'healthy' Thoroughbreds share a similar metabolic phenotype, which we explored using a range of NMR-based structural elucidation tools including 2-D pulse sequences to elicit carbon-proton correlations and statistical spectroscopy methods. Our future work includes a larger study investigating normal variation amongst different racehorse populations over time.

This work comprehensively assigns dominant features of the ^{1}H NMR spectra of the equine metabonome from plasma, urine and faeces and for the first time provides baseline information for future studies in equine health and disease. Urine and faecal profiles provide an insight into host-microbial metabolic interactions, whereas plasma profiles are more likely to represent host physiological processes. The purpose of this study is to provide an analytical template to researchers thinking about adding metabonomic analysis to their experiments and to indicate which biofluids may be of use to them. We have showed that faecal samples are more variable that either urine or plasma. Plasma is under tight homeostatic control and thus might be expected to show relatively less variation. However, urine, other than one identified outlier, exhibited less variation than the other biofluids under investigation and contributed the greatest number of identifiable metabolites. Further studies are warranted using metabonomic and metagenomic technology to explore the role of gut microbes on equine physiology and metabolism.

Authors' declaration of interests

No competing interests have been declared.

Source of funding

Horserace Betting Levy Board Research Scholarship at Imperial College London.

Ethical Animal Research

Sampling and metadata collection were carried out under University of Liverpool ethics approval RETH000363, with the informed consent of the trainer.

Acknowledgements

We are grateful to the racehorse trainers and owners who allowed us to sample their horses, to Elizabeth Metcalfe from the Animal Health Trust, Newmarket for technical assistance and Aisling Glennie at Boehringer Ingelheim for help editing the manuscript.

Authorship

E. Escalona contributed to all sections. J. Leng contributed to study execution, and data analysis and interpretation. J. Swann and A. Dona contributed to study design, study execution, and data analysis and interpretation. C. Merrifield contributed to the preparation of the manuscript. All authors gave their final approval of the manuscript.

Figure legends:

Fig 1: (A) 700 MHz 1D ^{1}H-NMR spectrum of urine, (B) CPMG ^{1}H-NMR spectrum of plasma, (C)1D ^{1}H-NMR spectrum of faecal water. All spectra partially labelled according to the assignments made in Table 1 and Supplementary Item 2.

Fig 2: (A) 800 MHz ^{1}H COSY NMR spectrum of the aromatic region of urine (B) J-res NMR spectrum of plasma highlighting the aliphatic region between 3-4.5 ppm. Key to metabolite identification is provided in Table 1 and Supplementary Item 2.

Fig 3: (A) Venn diagram highlighting the degree of metabolite overlap between biological compartments. The central section represents the number of core metabolites visible across all biological matrices (14 ubiquitous metabolites). The outer circle with numbered slices represents biological matrix specific metabolites that are ordered according to origins/function, and numbered according to metabolites in Table 1. Compartments are colour coded - plasma (red), urine (yellow) and faeces (green). Coloured dots represent metabolic functions, protein and amino-acid metabolism (purple), energy metabolism (orange) and fat metabolism (blue). Asterisk denotes metabolite can be mammalian in origin and a 'd' denotes metabolite can be from dietary origin.

Fig 4: (A) PCA scores plot demonstrating increased faecal variability relative to other biological matrices. Plasma (red), urine (yellow) and faecal (green) samples. Principal component 1 (PC1) accounts for 51%, PC2 for 15%, and PC3 for 5% of the total variation. Variance colour plots indicating the percentage of the total variance for each metabolite in (B) faeces and (C) urine (Mean spectrum is plotted, coloured by variance expressed as a percentage of the total variance).

Table 1: Metabolites identified using ^{1}H-NMR of equine plasma, urine and faeces. Metabolite numbers correspond with annotated ^{1}H-NMR spectra in Figs 1, 2 and 3. Metabolites are assigned to biofluids denoted by coloured dots; urine (yellow), plasma (red), faeces (green). The 14 core metabolites are highlighted in purple.

Manufacturers' addresses

[a]Bertin Technologies, Montigny-le-Bretonneux, France

[b]Bruker, Massachusetts, USA

[c]MATLAB, Mathwork, Massachusetts, USA

Additional Supplementary Items may be found in the online version of this article at the publisher's website:

Supplementary Item 1: The table highlights sample metadata and possible sources of variation within the data set. Yard, gender, age and training schedule are included. Gender is denoted as G=gelding, C=colt.

Supplementary Item 2: Table showing metabolites found in plasma (red circle), urine (yellow circle) and faeces (green circle) using ^{1}H-NMR spectroscopy. Peak multiplicities and chemical shifts are shown and structural information is also provided. † indicates tentative assignment. Details on each metabolite's origin and function are highlighted as well as a link to the metabolite's page in the hmdb database.

Supplementary Item 3: Table denoting feeding regimes on the 2 yards sampled. Top section highlights concentrate feeds and bottom highlights roughage types offered.

Supplementary Item 4: 800 MHz ^{1}H COSY NMR spectrum of the aliphatic region of faeces, between 3-4 ppm. Key to metabolite identification is provided in Table 1 and Supplementary Item 2**.**

Supplementary Item 5: 800 MHz ^{1}H ^{13}C HSCQ NMR spectrum of the aromatic region of urine, between 8-7 ppm. Key to metabolite identification is provided in Table 1 and Supplementary Item 2.

References

[1] Nicholson, J.K., Lindon, J.C. and Holmes, E. (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **29**, 1181-1189.

[2] Zhang, S., Nagana Gowda, G.A., Asiago, V., Shanaiah, N., Barbas, C. and Raftery, D. (2008) Correlative and quantitative 1H NMR-based metabolomics reveals specific metabolic pathway disturbances in diabetic rats. *Anal*. *Biochem*. **383**, 76-84.

[3] Poli, D., Carbognani, P., Corradi, M., Goldoni, M., Acampa, O., Balbi, B., Bianchi, L., Rusca, M. and Mutti, A. (2005) Exhaled volatile organic compounds in patients with non-small cell lung cancer: cross sectional and nested short-term follow-up study. *Respir*. *Res*. **6**, 71.

[4] Gowda, G.A., Zhang, S., Gu, H., Asiago, V., Shanaiah, N. and Raftery, D. (2008) Metabolomics-based methods for early disease diagnostics. *Expert Rev*. *Mol*. *Diagn*. **8**, 617-633.

[5] Peng, J.-B., Jia, H.-M., Xu, T., Liu, Y.-T., Zhang, H.-W., Yu, L.-L., Cai, D.-Y. and Zou, Z.-M. (2011) A (1)H NMR based metabonomics approach to progression of coronary atherosclerosis in a rabbit model. *Process Biochemistry* **46**, 2240-2247.

[6] Hooper, L.V., Bry, L., Falk, P.G. and Gordon, J.I. (1998) Host-microbial symbiosis in the mammalian intestine: exploring an internal ecosystem. *Bioessays* **20**, 336-343.

[7] Saric, J., Wang, Y., Li, J., Coen, M., Utzinger, J., Marchesi, J.R., Keiser, J., Veselkov, K., Lindon, J.C., Nicholson, J.K. and Holmes, E. (2008) Species variation in the fecal metabolome gives insight into differential gastrointestinal function. *Journal of Proteome Research* **7**, 352-360.

[8] Merrifield, C.A., Lewis, M., Claus, S.P., Beckonert, O.P., Dumas, M.-E., Duncker, S., Kochhar, S., Rezzi, S., Lindon, J.C., Bailey, M., Holmes, E. and Nicholson, J.K. (2011) A metabolic system-wide characterisation of the pig: a model for human physiology. *Molecular Biosystems* 7, 2577-2588.

[9] He, Q., Ren, P., Kong, X., Wu, Y., Wu, G., Li, P., Hao, F., Tang, H., Blachier, F. and Yin, Y. (2012) Comparison of serum metabolite compositions between obese and lean growing pigs using an NMR-based metabonomic approach. *Journal of Nutritional Biochemistry* **23**, 133-139.

[10] Scarth, J.P., Spencer, H.A., Timbers, S.E., Hudson, S.C. and Hillyer, L.L. (2010) The use of in vitro technologies coupled with high resolution accurate mass LC-MS for studying drug metabolism in equine drug surveillance. *Drug Testing and Analysis* **2**, 1-10.

[11] Liu, Y., Uboh, C.E., Soma, L.R., Li, X., Guan, F., You, Y. and Chen, J.-W. (2011) Efficient Use of Retention Time for the Analysis of 302 Drugs in Equine Plasma by Liquid Chromatography-MS/MS with Scheduled Multiple Reaction Monitoring and Instant Library Searching for Doping Control. *Analytical Chemistry* **83**, 6834-6841.

[12] Hodavance, M.S., Ralston, S.L. and Pelczer, I. (2007) Beyond blood sugar: the potential of NMR-based metabonomics for type 2 human diabetes, and the horse as a possible model. *Analytical and Bioanalytical Chemistry* **387**, 533-537.

[13] Keller, M.D., Pollitt, C.C. and Marx, U.C. (2011) Nuclear magnetic resonance-based metabonomic study of early time point laminitis in an oligofructose-overload model. *Equine Vet. J.* **43**, 737-743.

[14] Ralston, S.L., Pappalardo, L. and Pelczer, I. (2011) Breed and age effects on metabolic profiles of young horses using NMR-based Metabonomic analyses of serum. *Journal of Equine Veterinary Science* **31**, 304-305.

[15] Ralston, S.L., Pelczer, I., Shatynski, C. and Spears, P.F. (2009) NMR-based Metabonomic Analysis of Serum From Standardbred Yearlings With or Without Hock Osteochondrosis Dessicans Lesions: Preliminary Study. *Journal of Equine Veterinary Science* **29**, 385-386.

[16] Pappalardo, L., Pelczer, I. and Ralston, S.L. (2013) Metabolic Differences between Draft-cross and Mustang Horses Detected by Metabonomic Analyses. *Journal of Equine Veterinary Science* **33**, 1044-1049.

[17] Dumas, M.E., Maibaum, E.C., Teague, C., Ueshima, H., Zhou, B.F., Lindon, J.C., Nicholson, J.K., Stamler, J., Elliott, P., Chan, Q. and Holmes, E. (2006) Assessment of analytical reproducibility of H-1 NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. *Analytical Chemistry* **78**, 2199-2208.

[18] Beckonert, O., Keun, H.C., Ebbels, T.M.D., Bundy, J.G., Holmes, E., Lindon, J.C. and Nicholson, J.K. (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols* **2**, 2692-2703.

[19] Wright, B., Greatbanks, D., Taberner, J. and Wilson, I.D. (1995) Comparison of Three Methods for Water Suppression in Biofluid NMR: Advantages of NOESYPRESAT. *Pharmacy and Pharmacology Communications* **1**, 197-199.

[20] Pearce, J.T.M., Athersuch, T.J., Ebbels, T.M.D., Lindon, J.C., Nicholson, J.K. and Keun, H.C. (2008) Robust algorithms for automated chemical shift calibration of 1D H-1 NMR spectra of blood serum. *Analytical Chemistry* **80**, 7158-7162.

[21] Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H. (2006) Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabonomics. *Analytical Chemistry* **78**, 4281-4290.

[22] Trygg, J., Holmes, E. and Lundstedt, T. (2007) Chemometrics in metabonomics. *Journal of Proteome Research* **6**, 469-479.

[23] Craig, A., Cloareo, O., Holmes, E., Nicholson, J.K. and Lindon, J.C. (2006) Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Analytical Chemistry* **78**, 2262-2267.

[24] Fan, T.W.M. (1996) Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Progress in Nuclear Magnetic Resonance Spectroscopy* **28**, 161-219.

[25] Nicholson, J.K. and Wilson, I.D. (1989) High resolution proton magnetic resonance spectroscopy of biological fluids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **21**, 449-501.

[26] Bollard, M.E., Keun, H.C., Beckonert, O., Ebbels, T.M.D., Antti, H., Nicholls, A.W., Shockcor, J.P., Cantor, G.H., Stevens, G., Lindon, J.C., Holmes, E. and Nicholson, J.K. (2005) Comparative metabonomics of differential hydrazine toxicity in the rat and mouse. *Toxicology and Applied Pharmacology* **204**, 135-151.

[27] Cloarec, O., Dumas, M.E., Craig, A., Barton, R.H., Trygg, J., Hudson, J., Blancher, C., Gauguier, D., Lindon, J.C., Holmes, E. and Nicholson, J. (2005) Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Analytical Chemistry* **77**, 1282-1289.

[28] Kirwan, G.M., Coffey, V.G., Niere, J.O., Hawley, J.A. and Adams, M.J. (2009) Spectroscopic correlation analysis of NMR-based metabonomics in exercise science. *Analytica Chimica Acta* **652**, 173-179.

[29] Nicholson, J.K., Foxall, P.J., Spraul, M., Farrant, R.D. and Lindon, J.C. (1995) 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. *Anal*. *Chem*. **67**, 793-811.

[30] Holmes, E., Foxall, P.J., Spraul, M., Farrant, R.D., Nicholson, J.K. and Lindon, J.C. (1997) 750 MHz 1H NMR spectroscopy characterisation of the complex metabolic pattern of urine from patients with inborn errors of metabolism: 2-hydroxyglutaric aciduria and maple syrup urine disease. *J*. *Pharm*. *Biomed*. *Anal*. **15**, 1647-1659.

578 [31] Le Gall, G., Noor, S.O., Ridgway, K., Scovell, L., Jamieson, C., Johnson, I.T.,
579 Colquhoun, I.J., Kemsley, E.K. and Narbad, A. (2011) Metabolomics of Fecal
580 Extracts Detects Altered Metabolic Activity of Gut Microbiota in Ulcerative Colitis
581 and Irritable Bowel Syndrome. *Journal of Proteome Research* **10**, 4208-4218.
582
583 [32] Ametaj, B.N., Zebeli, Q., Saleem, F., Psychogios, N., Lewis, M.J., Dunn, S.M., Xia,
584 J.G. and Wishart, D.S. (2010) Metabolomics reveals unhealthy alterations in rumen
585 metabolism with increased proportion of cereal grain in the diet of dairy cows.
586 *Metabolomics* **6**, 583-594.
587
588 [33] Nyberg, N.T., Nielsen, M.O. and Jaroszewski, J.W. (2010) Metabolic trajectories
589 based on H-1 NMR spectra of urines from sheep exposed to nutritional challenges
590 during prenatal and early postnatal life. *Metabolomics* **6**, 489-496.
591
592 [34] Li, J., Wijffels, G., Yu, Y., Nielsen, L.K., Niemeyer, D.O., Fisher, A.D., Ferguson,
593 D.M. and Schirra, H.J. (2011) Altered fatty acid metabolism in long duration road
594 transport: An NMR-based metabonomics study in sheep. *J. Proteome Res*. **10**, 1073-
595 1087.
596
597 [35] Bollard, M.E., Stanley, E.G., Lindon, J.C., Nicholson, J.K. and Holmes, E. (2005)
598 NMR-based metabonomic approaches for evaluating physiological influences on
599 biofluid composition. *Nmr in Biomedicine* **18**, 143-162.
600
601 [36] Argenzio, R.A. (1975) Functions of equine large-intestine and their interrelationship
602 in disease. *Cornell Veterinarian* **65**, 303-330.
603
604 [37] Cummings, J.H. and Macfarlane, G.T. (1997) Role of intestinal bacteria in nutrient
605 metabolism. *Clinical Nutrition* **16**, 3-11.
606
607 [38] Desrousseaux, G., Santos, A.S., Pellikaan, W.F., Van der Poel, A.F.B., Cone, J.W.,
608 Guedes, C.M.V., Ferreira, L.M.M. and Rodrigues, M.A.M. (2012) Effect of collection
609 time on the fermentative activity of microbes in equine faeces. *Animal Feed Science*
610 *and Technology* **178**, 183-189.
611
612 [39] Van Weyenberg, S., Sales, J. and Janssens, G.P.J. (2006) Passage rate of digesta
613 through the equine gastrointestinal tract: A review. *Livestock Science* **99**, 3-12.
614
615 [40] Hudson, N.P.H. and Merritt, A.M. (2008) Equine gastrointestinal motility research:
616 Where we are and where we need to go. *Equine Vet. J*. **40**, 422-428.
617
618 [41] Bugaut, M. and Bentejac, M. (1993) Biological effects of short-chain fatty acids in
619 nonruminant mammals. *Annu. Rev. Nutr*. **13**, 217-241.
620
621 [42] Wyss, M. and Kaddurah-Daouk, R. (2000) Creatine and creatinine metabolism.
622 *Physiol. Rev*. **80**, 1107-1213.
623

[43] Baxmann, A.C., Ahmed, M.S., Marques, N.C., Menon, V.B., Pereira, A.B., Kirsztajn, G.M. and Heilberg, I.P. (2008) Influence of muscle mass and physical activity on serum and urinary creatinine and serum cystatin C. *Clinical Journal of the American Society of Nephrology* **3**, 348-354.

[44] Nicholson, J.K., Holmes, E. and Wilson, I.D. (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nature Reviews Microbiology* **3**, 431-438.

[45] Lees, H.J., Swann, J.R., Wilson, I.D., Nicholson, J.K. and Holmes, E. (2013) Hippurate: The Natural History of a Mammalian-Microbial Cometabolite. *J. Proteome Res*. **5**, 1527-1546.

[46] Clayton, T.A., Baker, D., Lindon, J.C., Everett, J.R. and Nicholson, J.K. (2009) Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc. Natl. Acad. Sci. USA* **106**, 14728-14733.

[47] Bennett, B.J., de Aguiar Vallim, T.Q., Wang, Z., Shih, D.M., Meng, Y., Gregory, J., Allayee, H., Lee, R., Graham, M., Crooke, R., Edwards, P.A., Hazen, S.L. and Lusis, A.J. (2013) Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab*. **17**, 49-60.

[48] Waldram, A., Holmes, E., Wang, Y., Rantalainen, M., Wilson, I.D., Tuohy, K.M., McCartney, A.L., Gibson, G.R. and Nicholson, J.K. (2009) Top-Down Systems Biology Modeling of Host Metabotype-Microbiome Associations in Obese Rodents. *Journal of Proteome Research* **8**, 2361-2375.

[49] Peppercorn, M.A. and Goldman, P. (1971) Caffeic acid metabolism by bacteria of the human gastrointestinal tract. *J. Bacteriol*. **108**, 996-1000.

[50] Li, J.V., Ashrafian, H., Bueter, M., Kinross, J., Sands, C., le Roux, C.W., Bloom, S.R., Darzi, A., Athanasiou, T., Marchesi, J.R., Nicholson, J.K. and Holmes, E. (2011) Metabolic surgery profoundly influences gut microbial-host metabolic cross-talk. *Gut* **60**, 1214-1223.

[51] Diab, S.S., Rodriguez-Bertos, A. and Uzal, F.A. (2013) Pathology and diagnostic criteria of Clostridium difficile enteric infection in horses. *Vet. Pathol*. **50**, 1028-1036.

[52] Diab, S.S., Songer, G. and Uzal, F.A. (2013) Clostridium difficile infection in horses: A review. *Veterinary Microbiology* **167**, 42-49.

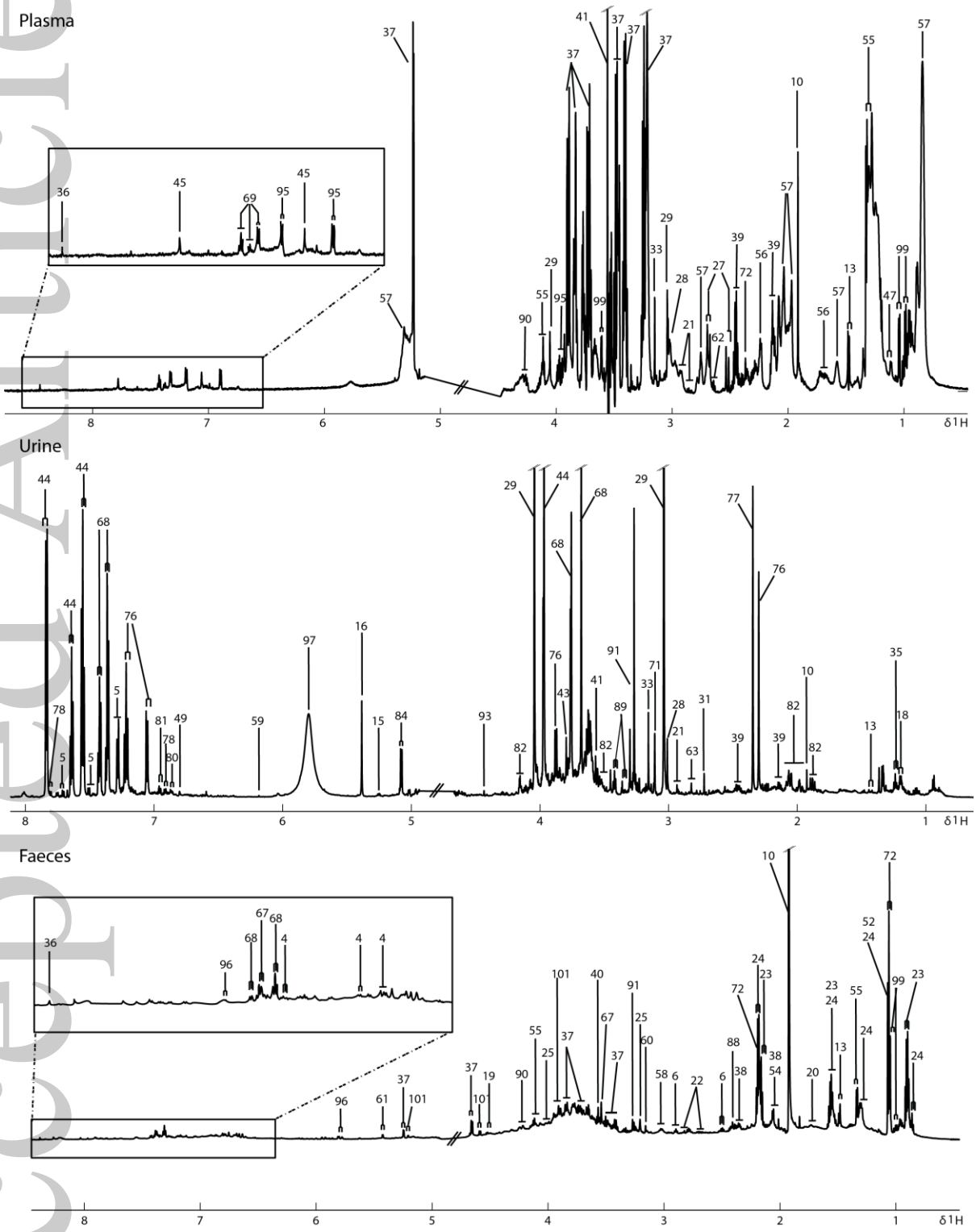
[53] Costa, M.C., Arroyo, L.G., Allen-Vercoe, E., Stampfli, H.R., Kim, P.T., Sturgeon, A. and Weese, J.S. (2012) Comparison of the fecal microbiota of healthy horses and horses with colitis by high throughput sequencing of the V3-V5 region of the 16S rRNA gene. *PLoS One* **7**, e41484.

[54] MM, O.D., Harris, H.M., Jeffery, I.B., Claesson, M.J., Younge, B., PW, O.T. and Ross, R.P. (2013) The core faecal bacterial microbiome of Irish Thoroughbred racehorses. *Lett. Appl. Microbiol.* **57**, 492-501.

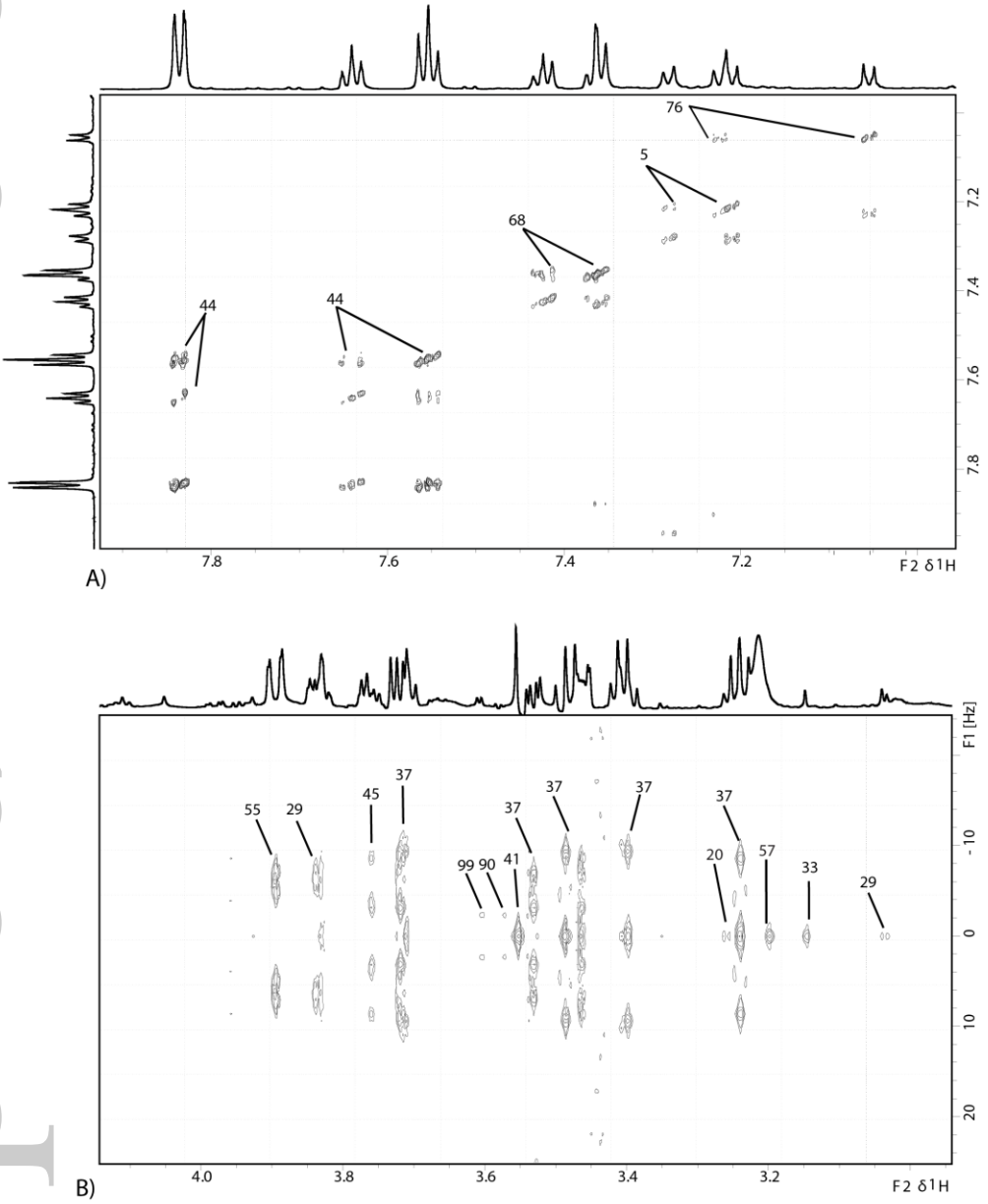
[55] Bollard, M.E., Holmes, E., Lindon, J.C., Mitchell, S.C., Branstetter, D., Zhang, W. and Nicholson, J.K. (2001) Investigations into biochemical changes due to diurnal variation and estrus cycle in female rats using high-resolution H-1 NMR spectroscopy of urine and pattern recognition. *Analytical Biochemistry* **295**, 194-202.

[56] Saude, E.J. and Sykes, B.D. (2007) Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics* **3**, 19-27.

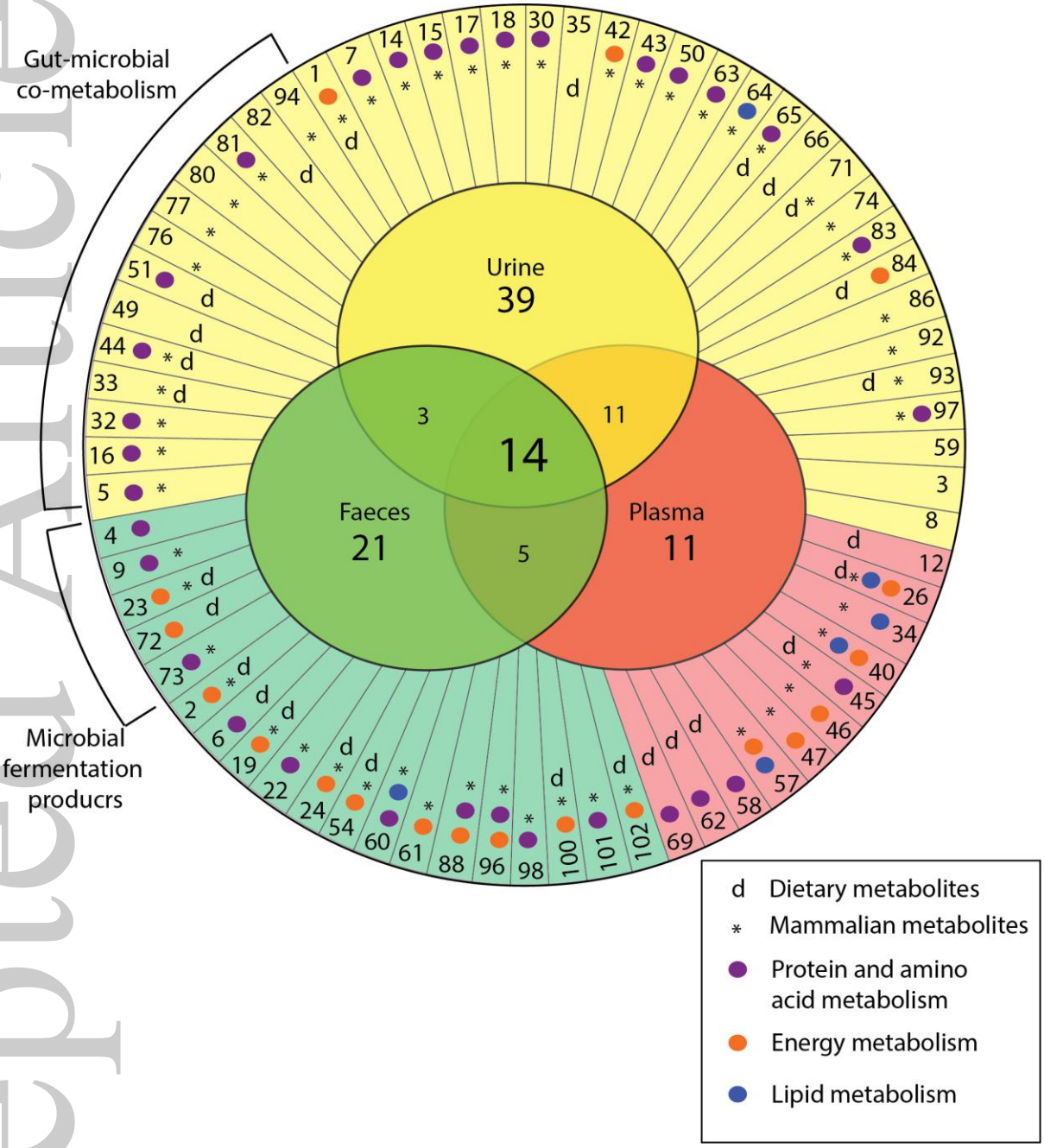
686



687
688 evj_12333_f1
689

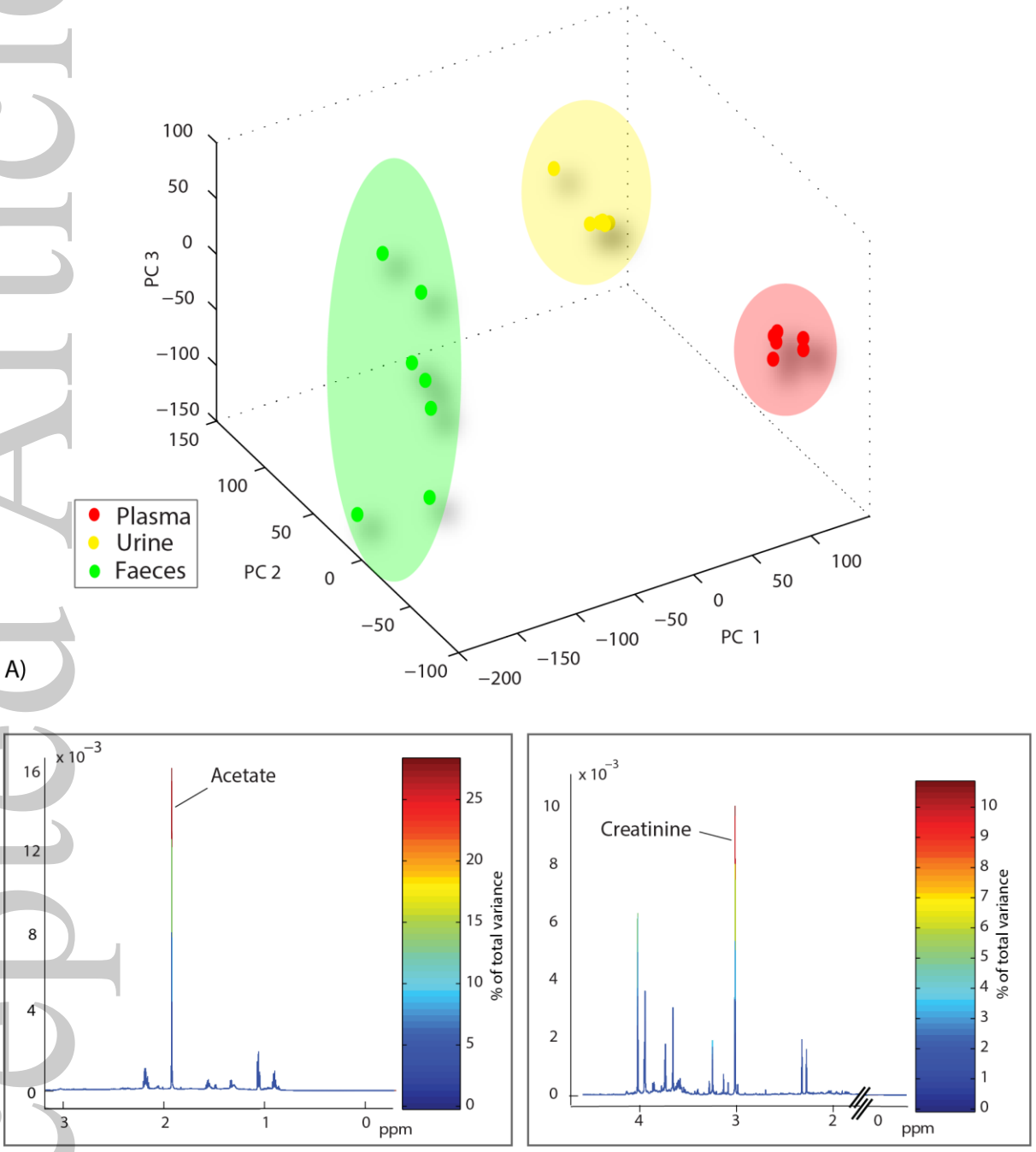


690
691
evj_12333_f2
692



693
694
695

evj_12333_f3



696
697
698

evj_12333_f4

N°	Compound
1	2-Hydroxy-3-methylbutyric acid
2	2-Methylbutyrate
3	3-hydroxy-4-methoxymandelic acid
4	3-Hydroxyphenylacetate
5	3-Indoxyl sulphate
6	3-phenylpropionoate
7	3-Ureidopropionic acid
8	4-hydroxy-3-methoxymandelic acid
9	5-Aminovalerate
10	Acetate
11	Acetoacetate
12	Adipate
13	Alanine
14	β Alanine
15	Allantoate
16	Allantoin
17	α-Aminoadipate
18	Beta-Aminoisobutyrate
19	Arabinose
20	Arginine
21	Asparagine
22	Aspartate
23	Butyrate
24	Caprylate
25	Choline
26	Cholestrol
27	Citrate
28	Creatine
29	Creatinine
30	Dihydrothymine
31	Dimethylamine (DMA)
32	Dimethylglycine (DMG)
33	Dimethyl sulphone
34	Ethanolamine
35	Ethyl glucoside
36	Formate
37	Glucose
38	Glutamate
39	Glutamine
40	Glycerol
41	Glycine
42	Glycogen
43	Guanidoacetate
44	Hippurate
45	Histidines
46	α Hydroxybutyrate
47	β Hydroxybutyrate
48	Alpha- Hydroxyisobutyrate
49	Hydroquinone
50	Hypotaurine
51	Indole-3-acetate

N°	Compound
52	Isobutyrate
53	Isoleucine
54	Isovalerate
55	Lactate
56	Leucine
57	Lipids
58	Lysine
59	Maleic anhydride
60	Malonate
61	Maltose
62	Methionine
63	Methylguanidine
64	Nicotinurate
65	Orotate
66	Pantothenate
67	Phenylacetate
68	Phenylacetylglycine (PAG)
67	Phenylalanine
69	Proline betaine
70	Phosphocholine
71	Proline betaine
72	Propionate
73	Putrescine
74	Pyridoxine
75	Pyruvate
76	*p*-cresol glucuronide
77	*p*-cresol sulphate
78	*p*-hydroxybenzoate
79	*p*-hydroxyphenylacetate
80	*p*-hydroxyphenyllactate
81	*p*-hydroxyphenylpyruvate
82	Quinate
83	Quinone
84	Raffinose
85	Scyllo-inositol
86	Sebacate
87	Serine
88	Succinate
89	Taurine
90	Threonine
91	Trimethylamine-N-oxide (TMAO)
92	Trans-aconitate
93	Trigonelline
94	Trimethylamine (TMA)
95	Tyrosine
96	Uracil
97	Urea
98	Uridine
99	Valine
100	Valerate
101	Xanthine
102	Xylose

699
700 evj_12333_t1
701