

**UCC Library and UCC researchers have made this item openly available.  
Please [let us know](#) how this has helped you. Thanks!**

<b>Title</b>	The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level
<b>Author(s)</b>	Barton, Wiley; Penney, Nicholas C.; Cronin, Owen; Garcia-Perez, Isabel; Molloy, Michael G.; Holmes, Elaine; Shanahan, Fergus; Cotter, Paul D.; O'Sullivan, Orla
<b>Publication date</b>	2018
<b>Original citation</b>	Barton, W., Penney, N. C., Cronin, O., Garcia-Perez, I., Molloy, M. G., Holmes, E., Shanahan, F., Cotter, P. D. and O'Sullivan, O. (2018) 'The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level', Gut, 67(4), pp. 625-633. doi: 10.1136/gutjnl-2016-313627
<b>Type of publication</b>	Article (peer-reviewed)
<b>Link to publisher's version</b>	<a href="http://gut.bmj.com/content/67/4/625">http://gut.bmj.com/content/67/4/625</a> <a href="http://dx.doi.org/10.1136/gutjnl-2016-313627">http://dx.doi.org/10.1136/gutjnl-2016-313627</a> Access to the full text of the published version may require a subscription.
<b>Rights</b>	© 2017, BMJ Publishing Group. All rights reserved. This article has been accepted for publication in Gut following peer review. The definitive copyedited, typeset version is available online at: <a href="https://dx.doi.org/10.1136/gutjnl-2016-313627">https://dx.doi.org/10.1136/gutjnl-2016-313627</a> <a href="http://www.bmj.com/company/products-services/rights-and-licensing/">http://www.bmj.com/company/products-services/rights-and-licensing/</a>
<b>Item downloaded from</b>	<a href="http://hdl.handle.net/10468/6514">http://hdl.handle.net/10468/6514</a>

Downloaded on 2022-06-26T13:24:20Z

1 The microbiome of professional athletes differs from that of sedentary subjects not only in  
2 composition but particularly at the functional metabolic level

3 Wiley Barton<sup>1,2,3</sup>, Nicholas C. Penney<sup>4, 5</sup>, Owen Cronin<sup>1,3</sup>, Isabel Garcia Perez<sup>4</sup>, Michael G.  
4 Molloy<sup>1,3</sup>, Elaine Holmes<sup>4</sup>, Fergus Shanahan<sup>1,3\*</sup>, Paul D. Cotter<sup>1,2</sup>, Orla O'Sullivan<sup>1,2</sup>

5  
6 <sup>1</sup>Alimentary Pharmabiotic Centre Microbiome Institute, University College Cork, National  
7 University of Ireland, Cork, Republic of Ireland

8 <sup>2</sup>Teagasc Food Research Centre, Moorepark, Fermoy, Cork, Republic of Ireland

9 <sup>3</sup>Department of Medicine, University College Cork, National University of Ireland, Cork,  
10 Republic of Ireland

11 <sup>4</sup>Section of Biomolecular Medicine, Division of Computational Systems Medicine,  
12 Department of Surgery and Cancer, Imperial College London, UK

13 <sup>5</sup>Division of Surgery, Department of Surgery and Cancer, Imperial College London, UK

14 **Corresponding author:**

15 Prof. Fergus Shanahan

16 APC Microbiome Institute,

17 University College Cork,

18 National University of Ireland, Cork,

19 Republic of Ireland

20 T: +353 21 490 1226

21 E: [f.shanahan@ucc.ie](mailto:f.shanahan@ucc.ie)

22

23 **Key words:** microbiome, exercise, metagenome, metabolome, fitness.

24 **Abbreviations:** Short chain fatty acid (SCFA), Body mass index (BMI), Ribosomal  
25 ribonucleic acid (rRNA), Vitamin Biosynthesis (VB), Lipid Biosynthesis (LB), Amino Acid  
26 Biosynthesis (AAB), Carbohydrate Biosynthesis (CB), Cofactor Biosynthesis (CfB), Energy  
27 Metabolism (EM), Creatine kinase (CK), trimethylamine-N-oxide (TMAO),

28 phenylacetylglutamine (PAG), trimethylamine (TMA), 3-Carboxy-4-methyl-5-propyl-2-  
29 furanpropionic acid (CMPF), Cardio-vascular disease (CVD).  
30

31 **Word count:** 3518/4000

32

## 33 **ABSTRACT**

### 34 **Objective**

35 It is evident that the gut microbiota and factors that influence its composition and activity  
36 influence human metabolic, immunological, and developmental processes. We previously  
37 reported that extreme physical activity with associated dietary adaptations, such as that  
38 pursued by professional athletes, is associated with changes in faecal microbial diversity and  
39 composition relative to that of individuals with a more sedentary lifestyle. Here we address  
40 the impact of these factors on the functionality/metabolic activity of the microbiota which  
41 reveals even greater separation between exercise and a more sedentary state.

### 42 **Design**

43 Metabolic phenotyping and functional metagenomic analysis of the gut microbiome of  
44 professional international rugby union players (n = 40) and controls (n = 46) was carried out  
45 and results were correlated with lifestyle parameters and clinical measurements (e.g. dietary  
46 habit and serum creatine kinase, respectively)

### 47 **Results**

48 Athletes had relative increases in pathways (e.g. amino acid and antibiotic biosynthesis and  
49 carbohydrate metabolism) and faecal metabolites (e.g. microbial produced short chain fatty  
50 acids [SCFAs] acetate, propionate, and butyrate) associated with enhanced muscle turnover  
51 (fitness) and overall health when compared to control groups.

52 **Conclusion**

53 Differences in faecal microbiota between athletes and sedentary controls show even greater  
54 separation at the metagenomic and metabolomics than at compositional levels and provide  
55 added insight into the diet-exercise-gut microbiota paradigm.

56 **SUMMARY**

57 **What is already known about this subject?**

58 Taxonomic and functional compositions of the gut microbiome are emerging as biomarkers  
59 of human health and disease.

60 Physical exercise and associated dietary adaptation are linked with changes in the  
61 composition of the gut microbiome.

62 Metabolites such as short chain fatty acids (SCFAs) have an impact on a range of health  
63 parameters including immunity, colonic epithelial cell integrity, and brain function.

64 **What are the new findings?**

65 Our original observation of differences in gut microbiota composition in elite athletes is  
66 confirmed and the separation between athletes and those with a sedentary lifestyle is even  
67 more evident at the functional or metabolic level. Microbial derived SCFAs are enhanced  
68 within the athletes.

69 **How might it impact on clinical practice in the foreseeable future?**

70 The findings provide new evidence supporting the link between exercise and metabolic  
71 health. The findings provide a platform for the rational design of diets for those engaged in  
72 vigorous exercise. The identification of specific alterations in the metabolic profile of

73 subjects engaged in high levels of exercise provides insight necessary for future efforts  
74 towards targeted manipulation of the microbiome.

## 75 **INTRODUCTION**

76 Regular exercise challenges systemic homeostasis resulting in a breadth of multi-organ  
77 molecular and physiological responses, including many that centre on immunity, metabolism  
78 and the microbiome-gut-brain axis.[1-5] Exercise exhibits systemic and end-organ anti-  
79 inflammatory effects, as well as contributing to more efficient carbohydrate metabolism, in  
80 addition to trophic effects at the level of the central nervous system.[6, 7] In fact, increasing  
81 physical activity offers an effective treatment and preventative strategy for many chronic  
82 conditions in which the gut microbiome has been implicated.[8-10] Conversely, a sedentary  
83 lifestyle is a major contributing factor to morbidity in developed Western society and is  
84 associated with heightened risk of numerous *diseases of affluence*, such as obesity, diabetes,  
85 asthma, and cardiovascular disease.[11-14] Recent evidence supports an influential role for  
86 the gut microbiome in these diseases.[15-23]

87         The concept that regular exercise and sustained levels of increased physical activity  
88 foster or assist the maintenance of a preferential intestinal microbiome has recently gained  
89 momentum and interest.[24-29] Previously, using 16S rRNA amplicon sequencing, we  
90 demonstrated taxonomic differences in gut microbiota between an elite athlete cohort of  
91 international-level rugby players and a group of age-matched high (>28 kg/m<sup>2</sup>) and low (<25  
92 kg/m<sup>2</sup>) BMI controls.[26] This analysis illustrated a significantly greater intestinal microbial  
93 diversity amongst the athletes compared to both control groups. This taxonomic diversity  
94 significantly correlated with exercise and dietary protein consumption. However, the  
95 possibility existed that these differences did not equate to differences at a functional level.  
96 Here, we re-examine the microbiome in these participants by whole metagenome shotgun

97 sequencing to provide deeper insight into taxonomic composition and metabolic potential and  
98 by complementary metabolic phenotyping analyses of host- and microbial-derived (urine and  
99 faecal respectively) metabolic profiles. This analysis shows that the differences in the gut  
100 microbiota between athletes and controls is even more pronounced at the functional  
101 metabolic level than at the compositional level as previously reported and provides further  
102 rationale for prospective controlled studies to unravel the relationship between diet, exercise  
103 and the gut microbiome.

## 104 **RESULTS**

105 The study groups were comprised of professional male athletes (n = 40) and healthy controls  
106 (n = 46).[26] To better represent the variability of BMI in the athletes, controls were  
107 classified as either low BMI (n = 22, BMI  $\leq$  25.2) or high BMI (n = 24, BMI  $\geq$  26.5).  
108 Participants made no report of gastrointestinal (GI) distress or alterations of GI transit time  
109 throughout the course of the initial study.

### 110 **Functional structure of the enteric microbiome correlates with athletic state**

111 Functional metagenomic analysis of faecal samples allowed for the prediction of the  
112 operational potential of each individual's microbiota. In total, 19,300 taxonomically linked  
113 metabolic pathways were identified in at least one individual. Comparison of phylogenetic  
114 constructions derived from the 16S rRNA amplicon data of our previous study and the  
115 functional data of this present report reveals a greater level of identification at higher levels  
116 of taxonomy (e.g. phylum) for 16S sequences,[26] while the metagenomic data had greater  
117 fidelity and superior resolution of lower levels of taxonomy (e.g. species) (Fig. 1). Consistent  
118 with previous results, the microbiota of the athletes were significantly more diverse than that  
119 of both the low and high BMI control groups at the functional level (Fig. 2A). Furthermore,  
120 our previous findings of an enrichment of *Akkermansia* in athletes was corroborated by the

121 presence of significantly higher proportions of metabolic pathways associated with this genus  
122 in athletes when compared to high BMI controls ( $p < 0.001$ ). Correlation analysis revealed  
123 that, of the total 19,300 pathways, 98 were significantly altered between the three cohorts ( $p$   
124  $< 0.05$ ) (Supplementary table 1). Subsequently, large-scale functional dissimilarity between  
125 athletes and controls was determined and distinct patterns of pathway composition between  
126 groups were revealed (Supplementary Fig. 1A). This functional distinction remained true  
127 whether applied to total pathway data or to the statistically significant subset of pathways  
128 (Supplementary Fig. 1B). Correlation of pathways present in at least one member from both  
129 cohorts further exemplified the uniformity of the athletes and the division between the  
130 athletes and control groups (Supplementary Fig. 1C). Separation according to group  
131 membership was further illustrated through Principal Coordinate Analysis (PCoA), with  
132 statistical support of the significant separation between the athletes and both control groups  
133 ( $p < 0.05$ ) (Fig. 2B). This was also the case for the statistically significant subset of pathways  
134 (Supplementary Fig. 1D). Principal Component Analysis (PCA) supplemented with a  
135 Correspondence Analysis (CA) and k-Nearest Neighbor (k-NN) semi-supervised learning  
136 approach cast further light (i.e. visualization of robustly defined class associations of specific  
137 individuals within the groups) on the clustering of participants within and between cohorts  
138 (Supplementary Fig. 1E).

139 Pathways exhibiting statistically significant variation between the athletes and both control  
140 groups were organised according to MetaCyc metabolic pathway hierarchy classification (34  
141 metabolic categories), highlighting a number of differences (Fig. 3A, Supplementary table 2).  
142 Distinct clustering patterns were observed within each cohort, with the high BMI control  
143 group having the lowest average abundance scores across 31 metabolic pathway categories  
144 (the exceptions being Vitamin Biosynthesis (VB), Lipid Biosynthesis (LB), and Amino Acid  
145 Biosynthesis (AAB) categories). The athlete group had the highest mean abundance across 29

146 of the 34 metabolic categories (e.g. Carbohydrate Biosynthesis [CB], Cofactor Biosynthesis  
147 [CfB], and Energy Metabolism [EM]) (Supplementary table 2).

148 Numerous statistically significant ( $p < 0.05$ ) associations were identified between pathway  
149 abundances and serum Creatine Kinase—an enzymatic marker of muscle activity (CK, IU/L),  
150 total bilirubin (IU/L) and dietary macronutrient intake of protein (g/day), fibre (g/day),  
151 carbohydrates (g/day), sugars (g/day), starch (g/day), fat (g/day), and total energy (KJ/day)  
152 (Fig. 3B). Each group was represented by distinct association profiles of the correlation  
153 between clinical measurements and metagenomic pathways. Dietary factors, sugars and other  
154 carbohydrates, as well as energy intake, provide the majority of the correlation for the control  
155 groups whereas the athlete group was predominantly correlated with CK, total bilirubin, and  
156 total energy intake. Of the total number of metabolic pathways with associations to the  
157 clinical data from all three groups (10,760; data not shown), relevant pathways related to the  
158 production of secondary metabolites, co-factors, and SCFAs were identified (e.g. biotin  
159 biosynthesis and pyruvate fermentation to butanoate).

#### 160 **Distinct differences between host and microbial metabolites in athletes and controls**

161 A combination of multi-platform metabolic phenotyping and multivariate analysis based on  
162 Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) was used to compare  
163 urinary and faecal samples from athletes and controls. The cross-validated (CV) OPLS-DA  
164 models show strong differences between athletes and controls in urine samples by proton  
165 nuclear magnetic resonance ( $^1\text{H-NMR}$ ) analysis ( $R^2Y=0.86$ ,  $Q^2Y=0.60$ , Fig. 2C), hydrophilic  
166 interaction ultra-performance liquid chromatography mass spectroscopy (HILIC UPLC-MS)  
167 positive mode analysis ( $R^2Y=0.85$ ,  $Q^2Y=0.74$ , Supplementary Fig. 2A) and reverse phase  
168 ultra-performance liquid chromatography mass spectroscopy (RP UPLC-MS) in both positive  
169 and negative mode analysis ( $R^2Y=0.83$ ,  $Q^2Y=0.73$ , and  $R^2Y=0.83$ ,  $Q^2Y=0.67$ ,



170 Supplementary Fig. 2B and 2C respectively). Likewise, the CV-OPLS-DA models comparing  
171 faecal samples, although weaker than the urine models, reveal significant differences between  
172 athletes and controls by <sup>1</sup>H-NMR analysis ( $R^2Y=0.86$ ,  $Q^2Y=0.52$ , Fig. 2D) and HILIC  
173 UPLC-MS positive mode analysis ( $R^2Y=0.65$ ,  $Q^2Y=0.34$ , Supplementary Fig 2D).

174 The loadings of the pairwise OPLS-DA models were used to identify metabolites  
175 discriminating between the two classes. Athletes' <sup>1</sup>H-NMR metabolic phenotypes were  
176 characterised by higher levels of trimethylamine-*N*-oxide (TMAO), L-carnitine,  
177 dimethylglycine, O-acetyl carnitine, proline betaine, creatinine, acetoacetate, 3-hydroxy-  
178 isovaleric acid, acetone, *N*-methylnicotinate, *N*-methylnicotinamide, phenylacetylglutamine  
179 (PAG) and 3-methylhistidine in urine samples and higher levels of propionate, acetate,  
180 butyrate, trimethylamine (TMA), lysine, and methylamine in faecal samples, relative to  
181 controls. Beta-alanine betaine was higher in both faecal and urine samples of athletes.  
182 Athletes were further characterised by lower levels of glycerate, allantoin and succinate and  
183 lower levels of glycine and tyrosine relative to controls in urine and faecal samples,  
184 respectively (Supplementary Table 3).

185 While numerous metabolites discriminated significantly between athletes and controls with  
186 RP UPLC-MS positive (490) and negative (434) modes for urine, as well as with HILIC  
187 UPLC-MS positive mode for urine (196) and faecal water (3), key metabolites were  
188 structurally identified using the strategy described below. UPLC-MS analyses revealed  
189 higher urinary excretion of *N*-formylanthranilic acid, hydantoin-5-propionic acid, 3-Carboxy-  
190 4-methyl-5-propyl-2-furanpropionic acid (CMPF), CMPF glucuronide, trimetaphosphoric  
191 acid, acetylcarnitine - C2, propionylcarnitine - C3, isobutyrylcarnitine - C4, 2-  
192 Methylbutyrylcarnitine - C5, Hexanoylcarnitine - C6, C9:1-carnitine, L-valine, nicotinuric  
193 acid, 4-pyridoxic acid and creatinine in athletes relative to controls. Levels of glutamine, 7-  
194 methylxanthine, imidazoleacetic acid, isoquinoline / quinolone were lower in athletes'

195 urinary samples relative to controls. Additionally, 16 unknown glucuronides were lower in  
196 the athlete samples (Supplementary Table 4).

197 SCFA levels in faeces measured by targeted GC-MS showed significantly higher levels of  
198 acetate ( $p < 0.001$ ), propionate ( $p < 0.001$ ), butyrate ( $p < 0.001$ ) and valerate ( $p = 0.011$ ) in  
199 athletes relative to controls. Isobutyrate and isovalerate did not differ significantly between  
200 the groups (Fig. 4B, Supplementary Table 5). Furthermore, concentrations of propionate  
201 strongly correlated to protein intake while butyrate was shown to have a strong association  
202 with intake of dietary fibre (Supplementary Table 6).

203

#### 204 **Correlating metabonomic and metagenomic results**

205 Correlation analysis between targeted measurements of SCFAs and taxonomic data from 16S  
206 rRNA sequencing revealed a number of correlations that remained significant following  
207 correction; *Roseburia* was positively correlated with acetate ( $p = 0.004$ ) and butyrate ( $p =$   
208  $0.018$ ) while *Family XIII Incertae Sedis* was positively correlated with isobutyrate ( $p <$   
209  $0.001$ ), isovaleric acid ( $p < 0.001$ ) and valeric acid ( $p = 0.008$ ) (Fig. 4A, Supplementary  
210 Table 7).

211

212 SCFAs were also correlated with pathway relative abundances, with all SCFAs associating  
213 with considerably more pathways in the athletes versus the controls (Fig. 4C). Multiple  
214 statistically significant (7,948) ( $p < 0.05$ ) correlations between the metabolic pathways and  
215 SCFAs were identified (Supplementary Table 8). Two distinct blocks of proportionately  
216 discriminant correlations were observed with isobutyric and isovaleric acids, which were  
217 more abundant in the athletes while acetic and butyric acids were more proportionately  
218 abundant in controls. Correlations of the SCFA concentrations to pathways related to  
219 fermentation, biosynthesis, or modification of fatty acids were identified among the

220 numerous other associations (see supplementary table 8 for complete list). Additional  
221 correlations of metabolic pathways against well-identified metabolites detected from both  
222 faecal water (Fig. 5A and 5C) and urine (Fig. 5B and 5D) presented numerous significant  
223 associations (6,186 and 13,412, respectively; data not shown) ( $p < 0.05$ ). It was also observed  
224 that 16 genera correlated with 12 metabolites (Supplementary Table 9).

## 225 **DISCUSSION**

226 The results confirm enhancement of microbial diversity in athletes compared with controls.  
227 Supporting previous insights into the beneficial influence of physical exercise and associated  
228 diet on the compositional structure of the gut microbiota,[25, 26, 30] this study has extended  
229 the paradigm to include links between physical fitness and the functional potential of the gut  
230 microbiota and its metabolites. It must be conceded that some athletes, although fit may not  
231 necessarily be more healthy.[31]

232 Athletes have an increased abundance of pathways that—giving an equivalent amount of  
233 expression activity—could be exploited by the host for potential health benefit, including  
234 biosynthesis of organic cofactors and antibiotics, as well as carbohydrate degradation and  
235 secondary metabolite metabolism compared to both control groups.[32] Furthermore, athletes  
236 have an enriched profile of SCFAs, previously associated with numerous health benefits and  
237 a lean phenotype.[33-35] While interpretation of SCFA data can be difficult as levels  
238 represent a combination of SCFA production and host-absorption rates, it is notable that , as  
239 previously presented, the athletes' diet maintained significantly higher quantities of fibre  
240 intake [29]. This along with an increased number of detected SCFA pathways in the athletes  
241 would be conducive to an enhanced rate of SCFA production[36]

242 It was noted that athletes excreted proportionately higher levels of the metabolite TMAO, an  
243 end product metabolite of dietary protein degradation. Elevated TMAO has been observed in

244 patients with cardiovascular disease and atherosclerosis, highlighting a potential downside to  
245 increased protein intake.[15-17, 22, 37] However, TMAO is also found in high levels in the  
246 urine of Japanese populations,[38] who do not have high risk for CVD. Similarly to these  
247 populations, the athletes' diet contained a significantly greater proportion of fish. Our current  
248 understanding of the implications of this result remains limited and requires elaboration in  
249 future studies. Furthermore, pathway abundance in a metagenome merely reflects functional  
250 potential and not necessarily increased expression in situ.

251 Variance of metagenomic composition between athletes and controls was exemplified with  
252 unique pathway-pathway correlations between the two groups. Analysis of categorically  
253 arranged pathway abundances within the separate cohorts provided additional insight into the  
254 previously described dichotomy between the microbiota of athletes and high BMI controls.  
255 The two groups displayed distinct structures of functional capacity, separately oriented to  
256 operate under the different physiological milieu of the two groups. Notably, from a functional  
257 perspective, the microbiota of the low BMI group was more similar to the athletes. The low  
258 BMI controls were generally engaged in a modestly active-lifestyle, reflected by their  
259 leanness and increased levels of CK. It is speculative but not implausible, that moderate  
260 improvements in physical activity, for overweight and obese individuals may confer the  
261 beneficial metabolic functions observed within the athlete microbiome.

262 Dietary contributions to the functional composition of the enteric microbial system are also  
263 evident in our study. The relative abundances of pathways related to fundamental metabolic  
264 function—amino acid biosynthesis, vitamin biosynthesis, and lipid biosynthesis—were  
265 higher on average within the high BMI control group when compared to the athlete group.  
266 The mechanisms behind these differences are unclear and might reflect chronic adaptation of  
267 the athlete gut microbiome; possibly due to a reduced reliance on the corresponding  
268 biosynthetic capacities of their gut microbiota. On the contrary, the athlete microbiome

269 presents a functional capacity that is primed for tissue repair and to harness energy from the  
270 diet with increased capacity for carbohydrate, cell structure and nucleotide biosynthesis,  
271 reflecting the significant energy demands and high cell-turnover evident in elite sport.

272 Remarkably, our examination of pathway correlation to dietary macronutrients and  
273 plasma CK, as a biomarker of exercise,[39] is suggestive of an impact of physical activity  
274 upon the utilization of dietary nutrients by the microbiota of the gut. Comparing athletes to  
275 both high and low BMI controls, a greater number of pathways correlating to specific  
276 macronutrients with the controls suggests a shift in the dynamics of these varied metabolic  
277 functions. The impact of the athletes' increased protein intake compared to both control  
278 groups was evident in the metabolomic phenotyping results. By-products of dietary protein  
279 metabolism (mostly by microbes) including TMAO, carnitines, trimethylamine, 3-Carboxy-  
280 4-methyl-5-propyl-2-furanpropionic acid, and 3-hydroxy-isovaleric acid are all elevated in  
281 the athlete cohort. Of particular interest is 3-hydroxy-isovaleric acid (potentially from egg  
282 consumption) which has been demonstrated to have efficacy for inhibiting muscle wasting  
283 when used in conjunction of physical exercise.[40] The compound is also commonly used as  
284 a supplement by athletes to increase exercise-induced gains in muscle size, muscle strength,  
285 and lean body mass, reduce exercise-induced muscle damage, and speed recovery from high-  
286 intensity exercise.[41] Numerous metabolites associated with muscle turnover—creatine, 3-  
287 methylhistidine, and L-valine—and host metabolism—carnitine—are elevated in the athlete  
288 groups. Metabolites derived from vitamins and recovery supplements common in  
289 professional sports, including glutamine, lysine, 4-pyridoxic acid, and nicotinamide, are also  
290 raised in the athlete group. It is notable that PAG—a microbial conversion product of  
291 phenylalanine—has been associated with a lean phenotype, and is increased in the  
292 athletes.[42] Furthermore, PAG positively correlates with the genus *Erysipelotrichaceae*  
293 *Incertae Sedis*, which we have previously noted to be present in relatively higher proportions

294 in the athlete group compared to both control groups. PAG is the strongest biomarker post  
295 bariatric surgery, where it is associated with an increase in the relative proportions of  
296 Proteobacteria as observed here in the athlete group. Within the SCFAs, two distinct clusters  
297 were observed; acetic acid, propionic acid and butyric acid correlate with dietary contributors  
298 (fibre and protein), while isobutyric acid, isovaleric acid and valeric acid correlate with  
299 microbial diversity. The same clusters are observed when correlating with individual taxa, in  
300 support of previously observed links between SCFAs and numerous metabolic benefits and a  
301 lean phenotype.[33-35]

302 Our on-going work in this area with non-athletes engaging in a structured exercise regime  
303 looks to further explore components of the exercise and diet-microbiome paradigm, which  
304 along with the present study may inform the design of exercise and fitness programs,  
305 including diet design in the context of optimizing microbiota functionality for both athletes  
306 and the general population.

307

## 308 **MATERIALS AND METHODS**

### 309 **Study population**

310 Elite professional male athletes (n = 40) and healthy controls (n = 46) matched for age and  
311 gender were enrolled in 2011 as previously described in the study.[26] Due to the range of  
312 physiques within a rugby team (player position dictates need for a variety of physical  
313 constitutions, i.e. forward players tend to have larger BMI values than backs, often in the  
314 overweight/obese range) the recruited control cohort was subdivided into two groups. In  
315 order to more completely include control participants, the BMI parameter for group inclusion  
316 was adjusted to  $BMI \leq 25.2$  and  $BMI \geq 26.5$  for the low BMI and high BMI groups

317 respectively. Approval for this study was granted by the Cork Clinical Research Ethics  
318 Committee.

### 319 **Acquisition of clinical, exercise and dietary data**

320 Self-reported dietary intake information was accommodated by a research nutritionist within  
321 the parameters of a food frequency questionnaire (FFQ) in conjunction with a photographic  
322 food atlas as per the initial investigation.[26] Fasting blood samples were collected and  
323 analysed at the Mercy University Hospital clinical laboratories, Cork. As the athletes were  
324 involved in a rigorous training camp we needed to assess the physical activity levels of both  
325 control groups. To determine this we used an adapted version of the EPIC-Norfolk  
326 questionnaire.[43] Creatine kinase levels were used as a proxy for level of physical activity  
327 across all groups.

### 328 **Preparation of Metagenomic libraries**

329 DNA derived from faecal samples was extracted and purified using the QIAmp DNA Stool  
330 Mini Kit (cat. no. 51504) prior to storage at -80°C. DNA libraries were prepared with the  
331 Nextera XT DNA Library Kit (cat. no. FC-131-1096) prior to processing on the Illumina  
332 HiSeq 2500 sequencing platform (see supplementary methods for further detail).

### 333 **Metagenomic statistical and bioinformatic analysis**

334 Delivered raw FASTQ sequence files were quality checked as follows: contaminating  
335 sequences of human origin were first removed through the NCBI Best Match Tagger  
336 (BMTagger). Poor quality and duplicate read removal, as well as trimming was implemented  
337 using a combination of SAM and Picard tools. Processing of raw sequence data produced a  
338 total of 2,803,449,392 filtered reads with a mean read count of 32,598,248.74 ( $\pm$  10,639,447  
339 SD) per each of the 86 samples. These refined reads were then subjected to functional

340 profiling by the most recent iteration of the Human Microbiome Project (HMP) Unified  
341 Metabolic Analysis Network (HUMAN2 v. 0.5.0) pipeline.[44] The functional profiling  
342 performed by HUMAN2 composed tabulated files of microbial metabolic pathway  
343 abundance and coverage derived from the Metacyc database.[45] Microbial pathway data was  
344 statistically analysed in the R software environment (v. 3.2.2) (for further details see  
345 supplementary methods).[46] All presented p values were corrected for multiple comparisons  
346 using the Benjamini-Hochberg False Discovery Rate (pFDR) method.[47]

347

### 348 **Metabolic profiling**

349 Urine and faecal samples were prepared for metabonomic analysis as previously  
350 described.[48, 49] Utilising established methods, urine samples underwent <sup>1</sup>H-NMR,  
351 reversed-phased (RP) and hydrophilic interaction chromatography (HILIC) profiling  
352 experiments. Faecal samples underwent <sup>1</sup>H-NMR, hydrophilic interaction chromatography  
353 (HILIC) and bile acid UPLC-MS profiling experiments and GC-MS targeted SCFA  
354 analysis.[49-51]

355

356 After data pre-processing,[52] the resulting <sup>1</sup>H-NMR and LC-MS data sets were imported  
357 into SIMCA 14.1 (Umetrics) to conduct multivariate statistical analysis. Principal Component  
358 Analysis (PCA), followed by Orthogonal Partial Least Squares Discriminant Analysis  
359 (OPLS-DA) was performed to examine the data sets and to observe clustering in the results  
360 according to the predefined classes. The OPLS-DA models in the current study were  
361 established based on one PLS component and one orthogonal component. Unit variance  
362 scaling was applied to <sup>1</sup>H-NMR data, Pareto scaling was applied to MS data. The fit and  
363 predictability of the models obtained was determined by the R<sup>2</sup>Y and Q<sup>2</sup>Y values,  
364 respectively. Significant metabolites were obtained from LC-MS OPLS-DA models through



365 division of the regression coefficients by the jack-knife interval standard error to give an  
366 estimate of the t-statistic. Variables with a t-statistic  $\geq 1.96$  (z-score, corresponding to the  
367 97.5 percentile) were considered significant. Significant metabolites were obtained from  $^1\text{H}$ -  
368 NMR OPLS-DA models after investigating correlations with correlation coefficients values  
369 higher than 0.4. Univariate statistical analysis (Mann-Whitney U test) was used to examine  
370 the SCFA data set. P-values were adjusted for multiple testing using the Benjamini-Hochberg  
371 False Discovery Rate (pFDR) method.

372

373 Confirmation of metabolite identities in the NMR data was obtained using 1D  $^1\text{H}$  NMR and  
374 2D  $^1\text{H}$ - $^1\text{H}$  NMR and  $^1\text{H}$ - $^{13}\text{C}$  NMR experiments. In addition, statistical tools such as SubseT  
375 Optimization by Reference Matching (STORM) and Statistical Total Correlation  
376 Spectroscopy (STOCSY) were also applied.[53, 54] Confirmation of metabolites identities in  
377 the LC-MS data was obtained using Tandem MS (MS/MS) on selected target ions.  
378 Metabolite identification was characterized by a level of assignment (LoA) score that  
379 describes how the identification was made.[55] The levels used were as follows: LoA 1:  
380 Identified compound, confirmed by comparison to an authentic chemical reference. LoA 2:  
381 MS/MS precursor and product ions or 1D+2D NMR chemical shifts and multiplicity match to  
382 a reference database or literature to putatively annotate compound. LoA 3: Chemical shift ( $\delta$ )  
383 and multiplicity matches a reference database to tentatively assign the compound. (For  
384 further details see supplementary methods).

385

## 386 **Acknowledgements**

387 The authors express gratitude to all participants for the donation of time and samples, Fiona  
388 Fouhy for insight provided into the library preparation of metagenomic sequencing. This  
389 research was funded by Science Foundation Ireland in the form of a centre grant (APC

390 Microbiome Institute Grant Number SFI/12/RC/2273). Research in the Cotter laboratory is  
391 funded by SFI through the PI award, “Obesibiotics” (11/PI/1137). Orla O’Sullivan and Wiley  
392 Barton are funded by Science Foundation Ireland through a Starting Investigator Research  
393 Grant award (13/SIRG/2160). Nicholas Penney is funded by the Diabetes Research and  
394 Wellness Foundation through the Sutherland-Earl Clinical Research Fellowship 2015. The  
395 authors thank the Imperial-National Institute for Health Research (NIHR) Clinical Phenome  
396 Centre for support. The centre is supported by the NIHR Imperial Biomedical Research  
397 Centre based at Imperial College Healthcare National Health Service (NHS) Trust and  
398 Imperial College London. The views expressed are those of the author(s) and not necessarily  
399 those of the NHS, the NIHR, or the Department of Health.

#### 400 **Conflicts of interest**

401 FS is a founder shareholder in Atlantia Food Clinical Trials, Tucana Health Ltd and  
402 Alimentary Health Ltd. He is director of the *APC Microbiome Institute*, a research centre  
403 funded in part by Science Foundation Ireland (APC/SFI/12/RC/2273) and which is/has  
404 recently been in receipt of research grants from Abbvie, Alimentary Health, Cremo, Danone,  
405 Janssen, Friesland Campina, General Mills, Kerry, MeadJohnson, Nutricia, 4D pharma and  
406 Second Genome, Sigmoid pharma.

#### 407 **Author Contributions**

408 W.B. prepared DNA samples for metagenomic sequencing.

409 O.O. and W.B. processed and analysed the metagenomic data.

410 E.H., I.G., and N.C.P. performed metabolomic processing and statistical analysis thereof.

411 F.S., P.D.C., O.O., and W.B. devised experimental design and approach.

412 F.S., P.D.C., O.C., O.O., M.M, E.H, N.C.P and W.B. wrote manuscript.

413 Results discussed by all authors.

#### 414 **Ethics approval**

415 Cork Clinical Research Ethics Committee.

#### 416 **Provenance and peer review**

#### 417 **Reprints and permissions information**

418 To request permissions go to: <http://group.bmj.com/group/rights-licensing/permissions>

419 To order reprints go to: <http://journals.bmj.com/cgi/reprintform>

420 To subscribe to BMJ go to: <http://group.bmj.com/subscribe/>

421 **Data Access:** In conformation of data accessibility protocol, metagenomic raw sequence data

422 from this study are deposited in EMBL Nucleotide Sequence Database (ENA)

423 (<http://www.ebi.ac.uk/ena/data/>), accession number **PRJEB15388**

#### 424 **REFERENCES**

- 425 1 Harkin A. Muscling in on depression. *The New England journal of medicine* 2014;371(24):2333-4  
426 doi:10.1056/NEJMcibr1411568
- 427 2 Benatti FB, Pedersen BK. Exercise as an anti-inflammatory therapy for rheumatic diseases-myokine  
428 regulation. *Nature reviews. Rheumatology* 2015;11(2):86-97 doi:10.1038/nrrheum.2014.193  
429 [2014/11/26].
- 430 3 Hawley JA, Krook A. Metabolism: One step forward for exercise. *Nature reviews. Endocrinology*  
431 2016;12(1):7-8 doi:10.1038/nrendo.2015.201
- 432 4 Hoffman-Goetz L, Pervaiz N, Packer N, Guan J. Freewheel training decreases pro- and increases  
433 anti-inflammatory cytokine expression in mouse intestinal lymphocytes. *Brain, behavior, and*  
434 *immunity* 2010;24(7):1105-15 doi:10.1016/j.bbi.2010.05.001 [2010/06/01].
- 435 5 Barton W, Shanahan F, Cotter PD, O'Sullivan O. The metabolic role of the microbiota. *Clinical Liver*  
436 *Disease* 2015;5(4):91-93 doi:10.1002/cld.455
- 437 6 Szuhany KL, Bugatti M, Otto MW. A meta-analytic review of the effects of exercise on brain-  
438 derived neurotrophic factor. *Journal of psychiatric research* 2015;60:56-64  
439 doi:10.1016/j.jpsychires.2014.10.003 [2014/12/03].

- 440 7 Ryan SM, Nolan YM. Neuroinflammation negatively affects adult hippocampal neurogenesis and  
441 cognition: can exercise compensate? *Neuroscience and biobehavioral reviews* 2016;61:121-  
442 31 doi:10.1016/j.neubiorev.2015.12.004 [2015/12/24].
- 443 8 Johannesson E, Simren M, Strid H, Bajor A, Sadik R. Physical activity improves symptoms in irritable  
444 bowel syndrome: a randomized controlled trial. *The American journal of gastroenterology*  
445 2011;106(5):915-22 doi:10.1038/ajg.2010.480 [2011/01/06].
- 446 9 Robsahm TE, Aagnes B, Hjartaker A, Langseth H, Bray FI, Larsen IK. Body mass index, physical  
447 activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis  
448 of cohort studies. *European journal of cancer prevention : the official journal of the European*  
449 *Cancer Prevention Organisation (ECP)* 2013;22(6):492-505  
450 doi:10.1097/CEJ.0b013e328360f434 [2013/04/18].
- 451 10 Schwingshackl L, Missbach B, Dias S, Konig J, Hoffmann G. Impact of different training modalities  
452 on glycaemic control and blood lipids in patients with type 2 diabetes: a systematic review  
453 and network meta-analysis. *Diabetologia* 2014;57(9):1789-97 doi:10.1007/s00125-014-  
454 3303-z [2014/07/06].
- 455 11 Biswas A, Oh PI, Faulkner GE, et al. Sedentary time and its association with risk for disease  
456 incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis.  
457 *Annals of internal medicine* 2015;162(2):123-32 doi:10.7326/M14-1651
- 458 12 Same RV, Feldman DI, Shah N, et al. Relationship Between Sedentary Behavior and Cardiovascular  
459 Risk. *Current cardiology reports* 2016;18(1):6 doi:10.1007/s11886-015-0678-5
- 460 13 Wilmut EG, Edwardson CL, Achana FA, et al. Sedentary time in adults and the association with  
461 diabetes, cardiovascular disease and death: systematic review and meta-analysis.  
462 *Diabetologia* 2012;55(11):2895-905 doi:10.1007/s00125-012-2677-z
- 463 14 Chen YC, Tu YK, Huang KC, Chen PC, Chu DC, Lee YL. Pathway from central obesity to childhood  
464 asthma. Physical fitness and sedentary time are leading factors. *American journal of*  
465 *respiratory and critical care medicine* 2014;189(10):1194-203 doi:10.1164/rccm.201401-  
466 0097OC
- 467 15 Koeth RA, Wang Z, Levison BS, et al. Intestinal microbiota metabolism of L-carnitine, a nutrient in  
468 red meat, promotes atherosclerosis. *Nature medicine* 2013;19(5):576-85  
469 doi:10.1038/nm.3145
- 470 16 Tang WHW, Wang ZE, Levison BS, et al. Intestinal microbial metabolism of phosphatidylcholine  
471 and cardiovascular risk. *New England Journal of Medicine* 2013;368(17):1575-84  
472 doi:10.1056/NEJMoa1109400
- 473 17 Tang WHW, Hazen SL. Microbiome, trimethylamine N-oxide, and cardiometabolic disease.  
474 *Translational research : the journal of laboratory and clinical medicine* 2016  
475 doi:10.1016/j.trsl.2016.07.007
- 476 18 Woting A, Pfeiffer N, Loh G, Klaus S, Blaut M. Clostridium ramosum promotes high-fat diet-  
477 induced obesity in gnotobiotic mouse models. *mBio* 2014;5(5):e01530-14  
478 doi:10.1128/mBio.01530-14
- 479 19 Utzschneider KM, Kratz M, Damman CJ, Hullar M. Mechanisms linking the gut microbiome and  
480 glucose metabolism. *The Journal of clinical endocrinology and metabolism*  
481 2016;101(4):jc20154251 doi:10.1210/jc.2015-4251
- 482 20 Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut  
483 microbiome with increased capacity for energy harvest. *Nature* 2006;444(7122):1027-31  
484 doi:10.1038/nature05414
- 485 21 Williams NC, Johnson MA, Shaw DE, et al. A prebiotic galactooligosaccharide mixture reduces  
486 severity of hyperpnoea-induced bronchoconstriction and markers of airway inflammation.  
487 *The British journal of nutrition* 2016;116(5):798-804 doi:10.1017/S0007114516002762
- 488 22 Wang Z, Klipfell E, Bennett BJ, et al. Gut flora metabolism of phosphatidylcholine promotes  
489 cardiovascular disease. *Nature* 2011;472(7341):57-63 doi:10.1038/nature09922

- 490 23 Zhernakova A, Kurilshikov A, Bonder MJ, et al. Population-based metagenomics analysis reveals  
491 markers for gut microbiome composition and diversity. *Science* 2016;352(6285):565-9  
492 doi:10.1126/science.aad3369
- 493 24 Cronin O, Molloy MG, Shanahan F. Exercise, fitness, and the gut. *Current opinion in*  
494 *gastroenterology* 2016;32(2):67-73 doi:10.1097/MOG.0000000000000240
- 495 25 Estaki M, Pither J, Baumeister P, et al. Cardiorespiratory fitness as a predictor of intestinal  
496 microbial diversity and distinct metagenomic functions. *Microbiome* 2016;4(1):42  
497 doi:10.1186/s40168-016-0189-7
- 498 26 Clarke SF, Murphy EF, O'Sullivan O, et al. Exercise and associated dietary extremes impact on gut  
499 microbial diversity. *Gut* 2014;63(12):1913-20 doi:10.1136/gutjnl-2013-306541
- 500 27 O'Sullivan O, Cronin O, Clarke SF, et al. Exercise and the microbiota. *Gut microbes* 2015;6(2):131-6  
501 doi:10.1080/19490976.2015.1011875
- 502 28 Cronin O, O'Sullivan O, Barton W, Cotter PD, Molloy MG, Shanahan F. Gut microbiota:  
503 implications for sports and exercise medicine. *British journal of sports medicine* 2017  
504 doi:10.1136/bjsports-2016-097225
- 505 29 Rankin A, O'Donovan C, Madigan SM, O'Sullivan O, Cotter PD. 'Microbes in sport' –The potential  
506 role of the gut microbiota in athlete health and performance. *British journal of sports*  
507 *medicine* 2017:bjsports-2016-097227 doi:10.1136/bjsports-2016-097227
- 508 30 Petriz BA, Castro AP, Almeida JA, et al. Exercise induction of gut microbiota modifications in  
509 obese, non-obese and hypertensive rats. *BMC genomics* 2014;15(1):511 doi:10.1186/1471-  
510 2164-15-511
- 511 31 Maffetone PB, Laursen PB. Athletes: Fit but Unhealthy? *Sports Med Open* 2016;2(1):24  
512 doi:10.1186/s40798-016-0048-x
- 513 32 Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates  
514 in the gut. *Gut microbes* 2012;3(4):289-306 doi:10.4161/gmic.19897
- 515 33 Koh A, De Vadder F, Kovatcheva-Datchary P, Backhed F. From Dietary Fiber to Host Physiology:  
516 Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* 2016;165(6):1332-45  
517 doi:10.1016/j.cell.2016.05.041
- 518 34 Ridaura VK, Faith JJ, Rey FE, et al. Gut microbiota from twins discordant for obesity modulate  
519 metabolism in mice. *Science* 2013;341(6150):1241214 doi:10.1126/science.1241214
- 520 35 Hamer HM, Jonkers DM, Bast A, et al. Butyrate modulates oxidative stress in the colonic mucosa  
521 of healthy humans. *Clinical nutrition* 2009;28(1):88-93 doi:10.1016/j.clnu.2008.11.002
- 522 36 den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM. The role of short-  
523 chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism.  
524 *Journal of lipid research* 2013;54(9):2325-40 doi:10.1194/jlr.R036012
- 525 37 Bennett BJ, de Aguiar Vallim TQ, Wang Z, et al. Trimethylamine-N-oxide, a metabolite associated  
526 with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell metabolism*  
527 2013;17(1):49-60 doi:10.1016/j.cmet.2012.12.011
- 528 38 Holmes E, Loo RL, Stalder J, et al. Human metabolic phenotype diversity and its association with  
529 diet and blood pressure. *Nature* 2008;453(7193):396-400 doi:10.1038/nature06882
- 530 39 Brancaccio P, Limongelli FM, Maffulli N. Monitoring of serum enzymes in sport. *British journal of*  
531 *sports medicine* 2006;40(2):96-7 doi:10.1136/bjsm.2005.020719
- 532 40 Stratton SL, Bogusiewicz A, Mock MM, Mock NI, Wells AM, Mock DM. Lymphocyte propionyl-CoA  
533 carboxylase and its activation by biotin are sensitive indicators of marginal biotin deficiency  
534 in humans. *The American journal of clinical nutrition* 2006;84(2):384-8
- 535 41 Wilson GJ, Wilson JM, Manninen AH. Effects of beta-hydroxy-beta-methylbutyrate (HMB) on  
536 exercise performance and body composition across varying levels of age, sex, and training  
537 experience: A review. *Nutrition & metabolism* 2008;5:1 doi:10.1186/1743-7075-5-1
- 538 42 Holmes E, Li JV, Athanasiou T, Ashrafian H, Nicholson JK. Understanding the role of gut  
539 microbiome-host metabolic signal disruption in health and disease. *Trends in microbiology*  
540 2011;19(7):349-59 doi:10.1016/j.tim.2011.05.006

- 541 43 Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE. Validity and repeatability of  
542 the EPIC-Norfolk Physical Activity Questionnaire. *International Journal of Epidemiology*  
543 2002;31(1):168-74 doi:10.1093/ije/31.1.168
- 544 44 Abubucker S, Segata N, Goll J, et al. Metabolic reconstruction for metagenomic data and its  
545 application to the human microbiome. *PLoS computational biology* 2012;8(6):e1002358  
546 doi:10.1371/journal.pcbi.1002358
- 547 45 Caspi R, Altman T, Billington R, et al. The MetaCyc database of metabolic pathways and enzymes  
548 and the BioCyc collection of Pathway/Genome Databases. *Nucleic acids research*  
549 2014;42(Database issue):D459-71 doi:10.1093/nar/gkt1103
- 550 46 R Development Core Team. R: A Language and Environment for Statistical Computing (R  
551 Foundation for Statistical Computing, Vienna, 2012). URL: <http://www.R-project.org> 2015
- 552 47 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach  
553 to Multiple Testing. *J Roy Stat Soc B Met* 1995;57(1):289-300
- 554 48 Dona AC, Jimenez B, Schafer H, et al. Precision high-throughput proton NMR spectroscopy of  
555 human urine, serum, and plasma for large-scale metabolic phenotyping. *Anal Chem*  
556 2014;86(19):9887-94 doi:10.1021/ac5025039
- 557 49 Garcia-Villalba R, Gimenez-Bastida JA, Garcia-Conesa MT, Tomas-Barberan FA, Carlos Espin J,  
558 Larrosa M. Alternative method for gas chromatography-mass spectrometry analysis of short-  
559 chain fatty acids in faecal samples. *J Sep Sci* 2012;35(15):1906-13  
560 doi:10.1002/jssc.201101121
- 561 50 Want EJ, Wilson ID, Gika H, et al. Global metabolic profiling procedures for urine using UPLC-MS.  
562 *Nat Protoc* 2010;5(6):1005-18 doi:10.1038/nprot.2010.50
- 563 51 Sarafian MH, Lewis MR, Pechlivanis A, et al. Bile acid profiling and quantification in biofluids using  
564 ultra-performance liquid chromatography tandem mass spectrometry. *Anal Chem*  
565 2015;87(19):9662-70 doi:10.1021/acs.analchem.5b01556
- 566 52 Veselkov KA, Vingara LK, Masson P, et al. Optimized preprocessing of ultra-performance liquid  
567 chromatography/mass spectrometry urinary metabolic profiles for improved information  
568 recovery. *Anal Chem* 2011;83(15):5864-72 doi:10.1021/ac201065j
- 569 53 Poma JM, Garcia-Perez I, De Iorio M, et al. Subset optimization by reference matching (STORM):  
570 an optimized statistical approach for recovery of metabolic biomarker structural information  
571 from 1H NMR spectra of biofluids. *Analytical chemistry* 2012;84(24):10694-701  
572 doi:10.1021/ac302360v
- 573 54 Cloarec O, Dumas ME, Craig A, et al. Statistical total correlation spectroscopy: an exploratory  
574 approach for latent biomarker identification from metabolic 1H NMR data sets. *Analytical*  
575 *chemistry* 2005;77(5):1282-9 doi:10.1021/ac048630x
- 576 55 Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical  
577 analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).  
578 *Metabolomics* 2007;3(3):211-21 doi:10.1007/s11306-007-0082-2

## 579 **FIGURE LEGENDS**

580 **Figure 1 | Comparison of phylogenetic constructions from metagenomic and 16S rRNA**  
581 **gene sequencing sourced from all participants.** Phylogenetic trees derived from (A)  
582 metagenomic sequencing and (B) 16S rRNA amplicon sequencing. Taxonomic levels are  
583 assigned from centre out with kingdom level assignment in centre and strain level assignment  
584 in outer most ring. Dark blue radial highlights correspond to poorly identified taxonomies  
585 (i.e. ‘unknown’ and ‘unassigned’ database entries). Number of assignments at each level of  
586 phylogeny is displayed below the respective graph. Taxonomic trees derived from the two  
587 sequencing approaches illustrate an advantage of metagenomic sequencing in the number of

588 predictions of lower taxonomic levels and the frequency of full identification of taxa, while  
589 16S rRNA sequencing grants greater insight of high level phylogenies within the population.

590

591 **Figure 2 | Group-wise comparison of microbial metagenomic and metabolomic profiles.**  
592 **(A)** Shannon index of diversity for metabolic pathways from all three groups. Pathway  
593 diversity is increased in the athlete group when compared to low BMI and high BMI controls.  
594 Diversity measures are statistically significant between low BMI and athletes ( $p < 0.049$ ),  
595 with statistical significance between all groups (Kruskal Wallis  $p < 0.05$ ). **(B)** Principle  
596 coordinate analysis (PCoA) of Bray-Curtis compiled distance matrix of all microbial  
597 metabolic pathway relative abundances. Groups show significant variation from one another  
598 (Adonis PERMANOVA  $p < 0.05$ ). **(C & D)** Cross validated orthogonal partial least squares  
599 regression discriminant analysis (OPLS-DA) of full Nuclear Magnetic Resonance ( $^1\text{H-NMR}$ )  
600 spectra from urine ( $R^2Y=0.86$ ,  $Q^2Y=0.60$ )(**C**) and faecal water ( $R^2Y=0.86$ ,  $Q^2Y=0.52$ )(**D**)  
601 samples. OPLS-DA displays robust separation between athletes and controls. Models are  
602 comprised of 1 predictive (tcv[1]) and 1 orthogonal (tocv[1]) principal component.

603

604 **Figure 3 | Group variation of microbial metabolic function and associations between**  
605 **pathways and clinical and dietary variables.** **(A)** Mean relative abundance values of  
606 statistically significant (Kruskal Wallis  $p < 0.05$ ) metabolic pathways binned according to  
607 categories of metabolic function. **(B)** Number of metabolic pathways significantly  
608 (Benjamini-Hochberg corrected  $p < 0.05$ ) correlated with dietary constituents and blood  
609 serum metabolites.

610

611 **Figure 4 | Athletes display a profile of Short Chain Fatty Acids that alters from that of**  
612 **the controls.** **(A)** Heat map of bacterial taxa (family, genus, and species level) that correlate  
613 with faecal short-chain fatty acid levels using Spearman's correlation. Cool colours represent  
614 positive correlations; hot colours represent negative correlations ( $r$ ). All taxa shown had a  
615 correlation  $p$ -value  $< 0.01$ . Those marked \* represent correlations with a pFDR  $< 0.01$  after  
616 Benjamini-Hochberg multiple testing corrections. **(B)** Median concentrations of GC-MS  
617 derived faecal short chain fatty acid. Quantitative analysis of SCFAs in faecal samples shows  
618 significant increase in measured concentrations of acetate, propionate, butyrate, and valerate  
619 in athletes. Error bars represent 95% confidence intervals. \* Data statistically significant ( $p <$   
620  $0.05$  after Benjamini-Hochberg corrections). **(C)** Quantification of statistically relevant  
621 correlations of metabolic pathways to GC-MS derived faecal SCFA concentrations ( $\mu\text{M}$ ).

622

623 **Figure 5 | Distinctive association profiles of metabolic pathways to metabolites in**  
624 **athletes and controls.** **(A)** Significant correlations of faecal water derived metabolites and  
625 metabolic pathways, represented by number of correlations for each metabolite. **(B)** Urine  
626 metabolites significantly correlated to pathways and displayed as number of correlations. **(C)**

- 627 Significant correlations shown in **(A)** displayed as proportions of total associations. **(D)**
- 628 Correlations presented in **(B)** given as proportions of total associations.