Title	The microbiome of professional athletes differs from that of more sedentary subjects in composition and particularly at the functional metabolic level
Authors	Barton, Wiley;Penney, Nicholas C.;Cronin, Owen;Garcia-Perez, Isabel;Molloy, Michael G.;Holmes, Elaine;Shanahan, Fergus;Cotter, Paul D.;O'Sullivan, Orla
Publication date	2018
Original Citation	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
Type of publication	Article (peer-reviewed)
Link to publisher's version	http://gut.bmj.com/content/67/4/625 - 10.1136/gutjnl-2016-313627
Rights	© 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: https://dx.doi.org/10.1136/gutjnl-2016-313627 - http://www.bmj.com/company/products-services/rights-and-licensing/
Download date	2025-08-16 18:35:45
Item downloaded from	https://hdl.handle.net/10468/6514

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

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

[1]Alimentary Pharmabiotic Centre Microbiome Institute, University College Cork, National University of Ireland, Cork, Republic of Ireland

[2]Teagasc Food Research Centre, Moorepark, Fermoy, Cork, Republic of Ireland

[3]Department of Medicine, University College Cork, National University of Ireland, Cork, Republic of Ireland

[4]Section of Biomolecular Medicine, Division of Computational Systems Medicine, Department of Surgery and Cancer, Imperial College London, UK

[5]Division of Surgery, Department of Surgery and Cancer, Imperial College London, UK

Corresponding author:

Prof. Fergus Shanahan
APC Microbiome Institute,
University College Cork,
National University of Ireland, Cork,
Republic of Ireland
T: +353 21 490 1226
E: f.shanahan@ucc.ie

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

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

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

Word count: 3518/4000

ABSTRACT

Objective

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

Design

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

Results

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

Conclusion

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

SUMMARY

What is already known about this subject?

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

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

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

What are the new findings?

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

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

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

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

INTRODUCTION

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

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

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

RESULTS

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

Functional structure of the enteric microbiome correlates with athletic state

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

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

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

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

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

Distinct differences between host and microbial metabolites in athletes and controls

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

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

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

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

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

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

Correlating metabonomic and metagenomic results

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

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

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

DISCUSSION

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

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

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

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

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

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

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

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

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

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

MATERIALS AND METHODS

Study population

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

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

Acquisition of clinical, exercise and dietary data

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

Preparation of Metagenomic libraries

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

Metagenomic statistical and bioinformatic analysis

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

profiling by the most recent iteration of the Human Microbiome Project (HMP) Unified Metabolic Analysis Network (HUMAnN2 v. 0.5.0) pipeline.[44] The functional profiling performed by HUMAnN2 composed tabulated files of microbial metabolic pathway abundance and coverage derived from the Metacyc database.[45] Microbial pathway data was statistically analysed in the R software environment (v. 3.2.2) (for further details see supplementary methods).[46] All presented p values were corrected for multiple comparisons using the Benjamini-Hochberg False Discovery Rate (pFDR) method.[47]

Metabolic profiling

Urine and faecal samples were prepared for metabonomic analysis as previously described.[48, 49] Utilising established methods, urine samples underwent ^{1}H-NMR, reversed-phased (RP) and hydrophilic interaction chromatography (HILIC) profiling experiments. Faecal samples underwent ^{1}H-NMR, hydrophilic interaction chromatography (HILIC) and bile acid UPLC-MS profiling experiments and GC-MS targeted SCFA analysis.[49-51]

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

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

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

Acknowledgements

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

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

Conflicts of interest

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

Author Contributions

W.B. prepared DNA samples for metagenomic sequencing.

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

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

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

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

Results discussed by all authors.

Ethics approval

Cork Clinical Research Ethics Committee.

Provenance and peer review

Reprints and permissions information

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

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

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

Data Access: In conformation of data accessibility protocol, metagenomic raw sequence data from this study are deposited in EMBL Nucleotide Sequence Database (ENA) (http://www.ebi.ac.uk/ena/data/), accession number **PRJEB15388**

REFERENCES

1 Harkin A. Muscling in on depression. *The New England journal of medicine* 2014;371(24):2333-4 doi:10.1056/NEJMcibr1411568

2 Benatti FB, Pedersen BK. Exercise as an anti-inflammatory therapy for rheumatic diseases-myokine regulation. *Nature reviews. Rheumatology* 2015;11(2):86-97 doi:10.1038/nrrheum.2014.193 [2014/11/26].

3 Hawley JA, Krook A. Metabolism: One step forward for exercise. *Nature reviews. Endocrinology* 2016;12(1):7-8 doi:10.1038/nrendo.2015.201

4 Hoffman-Goetz L, Pervaiz N, Packer N, Guan J. Freewheel training decreases pro- and increases anti-inflammatory cytokine expression in mouse intestinal lymphocytes. *Brain, behavior, and immunity* 2010;24(7):1105-15 doi:10.1016/j.bbi.2010.05.001 [2010/06/01].

5 Barton W, Shanahan F, Cotter PD, O'Sullivan O. The metabolic role of the microbiota. *Clinical Liver Disease* 2015;5(4):91-93 doi:10.1002/cld.455

6 Szuhany KL, Bugatti M, Otto MW. A meta-analytic review of the effects of exercise on brain-derived neurotrophic factor. *Journal of psychiatric research* 2015;60:56-64 doi:10.1016/j.jpsychires.2014.10.003 [2014/12/03].

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

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

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

FIGURE LEGENDS

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

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

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

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

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

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

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