Dairy consumption and physical fitness tests associated with fecal microbiome in a Chinese cohort

Jie, Zhuye; Liang, Suisha; Ding, Qiuxia; Li, Fei; Sun, Xiaohuan; Lin, Yuxiang; Chen, Peishan; Cai, Kaiye; Wang, Xiaohan; Zhang, Tao; Zhou, Hongcheng; Lu, Haorong; Xiao, Liang; Yang, Huanming; Wang, Jian; Hou, Yong; Kristiansen, Karsten; Jia, Huijue; Xu, Xun

Published in:
Medicine in Microecology

DOI:
10.1016/j.medmic.2021.100038

Publication date:
2021

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Dairy consumption and physical fitness tests associated with fecal microbiome in a Chinese cohort

Zhuye Jie a,b,c,*, Suisha Lianga,1, Qiuxia Ding a,1, Fei Li a, Xiaohuan Sun a, Yuxiang Lin a, Peishan Chena, Kaiye Caia, Xiaohan Wang a, Tao Zhang b,c, Hongcheng Zhou d, Haoong Lu d,h, Liang Xiao a,c,f, Huannming Yang a,g, Jian Wang a,g, Yong Hou a, Karsten Kristiansena,c,f, Huijue Jia a,b,c,**, Xun Xua,h,***

Abstract

Gut microbiome influenced many aspects of host physiology and psychology. Vice versa, lifestyles factors such as exercise and healthy diet are ways to shape the gut microbiota towards balance. We observed two distinct microbial groups characterized by physical fitness in a multi-omic cohort of 2183 young subjects with metagenomics, national physique comprehensive test, lifestyle and metabolome data. The panel of bacterial taxa including Clostridium bolteae, Escherichia coli, Ruminococcus gnavus, Clostridium clostridioforme, Clostridium innocuum, Bacteroides cellulosilyticus and Oscillospiraceae, were consistently associated with most of the physical fitness. Clostridium species and trace element both increased in the individuals those tend to stay up late. Yogurt consumption was associated with Streptococcus thermophilus and Bifidobacterium animalis subsp. lactis in feces, which differed from potentially endogenous Bifidobacterium species that was associated with milk intake. Our large-scale analyses were poised to advise for a healthy gut microbiome through behavioural changes.

1. Introduction

Gut microbiota is essential to maintain homeostasis in relation to physiology and psychology. Various natural environments has been identified as the external and internal factors impacting the gut microbiota. Accumulating evidence suggests that exercise influenced the gut microbiota and host variables has been found to be associated with distinct microbial compositions. However most of the biomarkers for exercise, sports habits and athletes were controversial due to the smaller sample size with weaker signals. Physical fitness, which represents the health of physical and outcome of exercise, was no report for the link between the gut microbiome and how these associations impact health. Circadian rhythm may also affect the gut microbiota. Break of circadian rhythm, sleep fragmentation and short sleep duration have been reported to promote overgrowth of specific gut bacteria and induce metabolic disturbances. However, it is not known how long-term late sleeping habits affect the gut microbiome in young and healthy individuals.

Here in our 2183 subjects 4D-SZ (trans-omic, with more time points in future studies, from Shenzhen, China) cohort with an average age of 29.6, we presented eight kinds of tests to evaluate the physical fitness conditions of volunteers. A detailed questionnaire was designed to simultaneously collect the lifestyle information on dairy consumption and sleeping habit. At the same time, we performed fecal metagenomic shotgun sequencing and plasma metabolites measurements to decipher the associations between physical fitness, lifestyle and gut microbiome.
Specific microbial taxa in this cohort were observed to be associated with physical fitness, exercise, sleep, and dairy consumption. In addition, our large cohort showed that commercial yogurt strains, especially *Streptococcus thermophilus* and *Bifidobacterium animalis* in feces which differed from potentially endogenous *Bifidobacterium* species, were associated with milk intake, suggesting potentially beneficial effects in cardiometabolic health.

2. Materials and methods

2.1. Study cohort

As a part of 4D-SZ, all the >2000 volunteers in the first cohort were recruited between May 2017 and July 2017 during a physical examination in Shenzhen. The 1400 volunteers in the second cohort were also recruited in 2017 until late September in multiple cities in China, including Wuhan, Qingdao, Tianjin. Detailed sample information regarding the cohorts has been reported [12]. The samples in each omics were shown in Supplementary Table 1c [12]. Baseline characteristics of the cohort were shown in Supplementary Tables 1b and 1d [12].

The study was approved by the Institutional Review Boards (IRB) at BGI-Shenzhen (BGI-IRB19121), and all participants provided written informed consent at enrolment.

2.2. Demographic data collection

The lifestyle questionnaire contained 56 entries involving age, marital status, disease history of the volunteer and his/her family, eating and exercise habits [Supplementary Tables 1b and 1d [12]]. The psychological questionnaire contained 18 entries to evaluate irritability, dizziness, frustration, fear, appetite, self-confidence, resilience [Supplementary Table 1b [12]].

2.3. Samples collection, DNA extraction and metagenomic shotgun sequencing

Fecal samples were self-collected by the volunteers using a kit containing a room temperature stabilizing reagent, which can preserve the fecal metagenome for at least two weeks at room temperature to preserve the metagenome [13]. The samples were frozen at ~80 °C on the same day and only thawed for DNA extraction. The overnight fasting blood samples were collected from a cubital vein of volunteers by medical doctors.

DNA extraction of the stored fecal samples within the next few months after the collection was performed as previously described [14]. This manual extraction protocol has been described in detail [15], and used in many studies [16] [-] [19]. Briefly, a frozen aliquot of fecal sample was suspended in 250 μl of guanidine thiocyanate, 0.1 M Tris (pH 7.5) and 40 μl of 10% N-lauroyl sarcosine. Then 500 μl 5% N-lauroyl sarcosine was added. After 1 h incubation, 500 μl of glass beads (0.1 mm) and 500 μl of TENP were added to the tube for vortexing, followed by centrifugation. The supernatant was transferred to a new tube, and DNA was precipitated by isopropanol [13]. Metagenomic sequencing was performed on the BGISEQ-500 platform (PCR-free without size selection, 100bp of single-end reads for fecal samples), and quality-controlled as previously reported [20].

2.4. Amino acid measurements

40 μl plasma was deproteinized with 20 μl 10% (w/v) sulfosalicylic acid (Sigma) containing internal standards, then 120 μl aqueous solution was added. After centrifugation, the supernatant was used for analysis. The analysis was performed by ultra high pressure liquid chromatography (UHPLC) coupled to an AB Sciex Qtrap 5500 mass spectrometry (AB Sciex, US) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.8 μm, 2.1 × 100 mm) was used for amino compound separation with a flow rate at 0.5 ml/min and column temperature of 55 °C. The mobile phases were (A) water containing 0.05% and 0.1% formic acid (v/v), (B) acetonitrile containing 0.05% and 0.1% formic acid (v/v). The gradient elution was 2% B kept for 0.5 min, then changed linearly to 10% B during 1 min, continued up to 35% B in 2 min, increased to 95% B in 0.1 min and maintained for 1.4 min. Multiple Reaction Monitoring (MRM) was used to monitor all amino compounds. The mass parameters were as follows, Curtain gas flow 35 L/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 60 l/min, Ion Source Gas 2 (GS 2) flow 60 l/min, IonSpray Voltage (IS) 5500 V, temperature 600 °C. All amino compound standards were purchased from sigma and Toronto research chemical (TRC).

2.5. Hormone measurements

250 μl plasma was diluted with a 205 μl aqueous solution. For SPE experiments, HLB (Waters, USA) was activated with 1.0 ml of dichloromethane, acetonitrile, methanol, respectively and equilibrated with 1.0 ml of water. The pretreated plasma sample was loaded onto the cartridge using gravity. Clean up was accomplished by washing the cartridges with 1.0 ml of 25% methanol in water. After drying under vacuum, samples on the cartridges were eluted with 1.0 ml of dichloromethane. The eluted extract was dried under nitrogen and the residual was dissolved with 25% methanol in water and was transferred to an autosampler vial prior to LC-MS/MS analysis. The analysis was performed by UHPLC coupled to an AB Sciex Qtrap 5500 mass spectrometry (AB Sciex, US) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Phenomenex Kinetix C18 column (2.6 μm, 2.1 × 50 mm) was used for steroid hormone separation with a flow rate at 0.8 ml/min and column temperature of 55 °C. The mobile phases were (A) water containing 1 mM Ammonium acetate, (B) Methanol containing 1 mM Ammonium acetate. The gradient elution was 25% B kept for 0.9 min, then changed linearly to 40% B during 0.9 min, continued up to 70% B in 2 min, increased to 95% B in 0.1 min and maintained for 1.6 min. Multiple Reaction Monitoring (MRM) was used to monitor all steroid hormone compounds. The mass parameters were as follows, Curtain gas flow 35 L/min, Collision Gas (CAD) was medium, Ion Source Gas 1 (GS 1) flow 60 l/min, Ion Source Gas 2 (GS 2) flow 60 l/min, Nebulizer Current (NC) 5, temperature 500 °C. All steroid hormone profiling compound standards were purchased from sigma, Toronto research chemical (TRC), Cerilliant and DR. Ehrenstorfer.

2.6. Trace element measurements

200 μl of whole blood were transferred into a 15 ml polyethylene tube and diluted 1:25 with a diluent solution consisting of 0.1% (v/v) Triton X-100, 0.1% (v/v) HNO3,2 mg/L AU, and internal standards (20 μg/L). The mixture was sonicated for 10 min before ICP-MS analysis. Multi-element determination was performed on an Agilent 7700x ICP-MS (Agilent Technologies, Tokyo, Japan) equipped with an octupole reaction system (ORS) collision/reaction cell technology to minimize spectral interferences. The continuous sample introduction system consisted of an autosampler, a quartz torch with a 2.5-mm-diameter injector with a Shield Torch system, a Scott double-pass spray chamber and nickel cones (Agilent Technologies, Tokyo, Japan). A glass concentric MicroMist nebulizer (Agilent Technologies, Tokyo, Japan) was used for the analysis of diluted samples.

2.7. Water-soluble vitamins measurements

200 μl plasma were deproteinized with 600 μl methanol (Merck), water, acetic acid (9:1:0.1) containing internal standards, thiamine-(4-methyl-13C-hizazol-5-yl-13C3) hydrochloride (Sigma-Aldrich), levomefolic acid-13C, d3, riboflavin-13C, 15N2, 4-pyridoxic acid-d3 and pantothenic acid-13C3,15 N heme calcium salt (Toronto Research...
Chemicals), 500 μl supernatant was dried by nitrogen flow. 60 μl water were added to the residues, vortexed, the mixture was centrifuged and the supernatant was used for analysis. The analysis was performed by UPLC coupled to a Waters Xevo TQ-S Triple Quad mass spectrometry (Waters, USA) with the electrospray ionization (ESI) source in positive ion mode. A Waters ACQUITY UPLC HSS T3 column (1.7 μm, 2.1 × 50 mm) was used for water-soluble vitamins separation with a flow rate at 0.45 ml/min and column temperature of 45 °C. The mobile phases were (A) 0.1% formic acid in the water, (B) 0.1% formic acid in methanol. The following elution gradient was used: 0–1 min, 99.0%–99.0% A; 1–1.5 min, 99.0%–A–97.0% A; 1.5–2 min, 97.0%–A–70.0% A; 2–3.5 min, 70.0%–A–30.0% A; 3.5–4.0 min, 30%–A–10.0% A; 4.0–4.8 min, 10%–A–10.0% A; 4.9–6.0 min, 99.0%–A–99.0% A. Multiple Reaction Monitoring (MRM) was used to monitor all water-soluble vitamins. The mass parameters were as follows, the capillary voltages of 3000 V and source temperature of 150 °C were adopted. The desolvation temperature was 500 °C. The collision gas flow was set at 0.10 ml/min. The cone gas and desolvation gas flow were 150 l/h and 1000 l/h, respectively. All water-soluble vitamins standards were purchased from Sigma-Aldrich (USA).

2.8. Fat-soluble vitamins measurements

250 μl plasma were deproteinized with 1000 μl methanol and acetonitrile (v/v, 1:1) (Fisher Chemical) containing internal standards, all-trans-Retinol-d5, 25-HydroxyVitamin-D2-d6, 25-HydroxyVitamin-D3-d6, vitamin K1-d7, a-Tocopherol-d6 (Toronto Research Chemicals). 900 μl supernatant was dried by nitrogen flow. 80 μl 80% acetonitrile was added to the residues, vortexed, the mixture was centrifuged, and the supernatant was used for analysis. The analysis was performed by UPLC coupled to an AB Sciex Qtrap 4500 mass spectrometry (AB Sciex, USA) with the atmospheric pressure chemical ionization (APCI) source in positive ion mode. A Waters ACQUITY UPLC BEH C18 column (1.7 μm, 2.1 × 50 mm) was used for fat-soluble vitamins separation with a flow rate at 0.50 ml/min and column temperature of 45 °C. The mobile phases were (A) 0.1% formic acid in water, (B) 0.1% formic acid in acetonitrile. The following elution gradient was used: 0–0.5 min,60.0%-60.0% A; 0.5–1.5 min, 60.0%–A–20.0% A; 1.5–2.5 min, 20.0%–A–0% A, 2.5–4.5 min, 0%–A–0% A; 4.5–4.6 min, 0%–A–60.0% A; 4.6–5.0 min, 60.0%–A–60.0% A. Multiple Reaction Monitoring (MRM) was used to monitor all fat-soluble vitamins. The mass parameters were as follows, Curtain gas flow 30 l/min, Curtain Gas (C) was medium, Ion Source Gas 1 (GS 1) flow 40 l/min, Ion Source Gas 2 (GS 2) flow 50 l/min, Nebulizer Current (NC) 5, temperature 400 °C. All fat-soluble vitamins standards were purchased from Sigma-Aldrich (USA), Toronto research chemical (TRC).

2.9. Medical parameters

All the volunteers were recruited during the physical examination. The medical test included blood tests, urinalysis, routine examination of cervical secretion. All the medical parameters were measured by the physical examination center and shown in Supplementary Table 1b [12], 1d [12]. InBody (InBody Co. Ltd. U.S.A.) was used to estimate body composition.

2.10. Physical fitness test

Eight kinds of tests were performed to evaluate volunteers’ physical fitness condition (Supplementary Table 1b [12]). Vital capacity was measured by HK6800-PH (Hengkangjiaye, China). Eye-closed and single-legged standing was measured by HK6800-ZL. Choice reaction time was measured by HK6800-FY. Grip strength was measured by HK6800-WL. Sit and reach was measured by HK6800-TQ. Sit-ups was measured by HK6800-YW. Step index was measured by HK6800-TJ. Vertical jump was measured by HK6800-ZT. We got a measured value from each test. Then each measure value score was assigned 1 through 5 based on its corresponding age-matched national standards (Supplementary Table 5 [12]). Both the direct measurements and the scores were used for analyses (Supplementary Table 2 [12], Supplementary Table 3 [12]).

2.11. Quality control, taxonomic annotation and abundance calculation

The sequencing reads were quality-controlled as described previously [20] and implementation of the pipeline is available at https://github.com/jiezhuye/cOMG. Briefly, adapter trimming/filtering is automatically processed by the BGISEQ-500 sequencing platform. The raw sequences with low quality were filtered and trimmed by overall accuracy (OA) control strategy [20] using OAs1 (-Qsys = 33, -minLen = 30, -Scut = 0.9, -Qcut = 0.8). Then the high-quality reads were aligned to hg19 by SOAP2.22 (identity ≥ 0.9) to remove human-source reads by removeHost (-D 4 -s 30 -r 1 -v 7 -i 0.9). The retained high-quality clean reads were aligned to the integrated gene catalog (IGC) with 9,879,896 genes by SOAP2.22 (identity ≥ 0.95) to get the assignment [15]. Taxonomy of the fecal MGs/MLGs was then determined from their constituent genes, as previously described [14,21,22]. The profile of this MGs/MLG was the trimmed mean (remove the highest and lowest 5% abundance genes) of genes abundance within this MGs/MLG. We also compute the Metaphlan2 [23] profile (v2.6.0) with the default parameter.

2.12. The factors in each type of omics predicted by other types of omics

Every factor in each omics was regressed on the relative abundances of MGs profile (found in at least 10% of the samples) in the fecal samples using default parameters in the RFCV function from randomForest package in R(cv.fold = 5, ntrees = 500). Dichotomous variables (such as gender), and unordered categorical variable (such as region) were re-coding into dummy variables. Frequency items such as yogurt eating habits were assigned to integers. RFCV R was defined as the Spearman correlation between the measured value and 5-fold cross-validation predicted value. Then the top 5 important predict factors in each omics type was record (Supplementary Fig 1). The same prediction process was done between any two types of omics. Then ggplot2 package in R was used to boxplot predict the power of target omics factors by all kinds of other predictor omics. R pheatmap and barplot were used to make the heatmap plot for representative factor. Mantel test was done with mantel.rtest function(overall composition was measured with bray distance). 4999 times permutation was used to compute the p-value, and 4999 times bootstrap was used to compute the 95% confidence interval.

2.13. Statistical adjustment for factors that may influence the gut microbiome

Associations between gut microbiome MGs, functional modules, Shannon diversity, and other types of omics data were calculated by linear model or partial Spearman correlation, adjusted for potential confounders including gender, age, BMI, health products (amino acid, vitamin, calcium), antivirus, antibiotics, drugs (currently using antihypertensive drugs, hypoglycemic drugs, hypolipidemic drugs), days since last menstrual bleeding, pregnant, lactation, bowel problem (defecation), as reported in recent studies [24]. Besides the above basic set of confounders, we also show the results adjusted for more potential confounders including dietary (dietary taste spicy, sweet, salty, oil, or light, high sugar and high-fat diet habit, fruit and vegetable intake, favors fat meal), exercise (exercise frequency, exercise intensity, average time per exercise), drinking, smoking and Bristol’s stool score.

2.14. Mathematical transformation for compositional data

We use mmvec [25,26] to re-rank and highlight associations picked up by average ranks methods (see methods 2.16), which account for the
composition effect. Gut microbe relative abundance profile is a compositional data (sum to 1). “mmvec” (microbe–metabolite vectors) was a compositional technique that can handle multi-omics datasets. It reports co-occurrence probabilities between microbes and metabolites by a two-layer neural network. Softmax transform enforces scale invariance to remove the composition effect. It has several users defined parameters. In our study, we set “-latent-dim 3 -min-feature-count 10 -learning-rate 1e-5 -beta1 0.9 -beta2 0.99”. Both the logloss and cv_rmse curves decaying, and plateau is close to zero, which suggested a good model. Other parameter set can achieve the same good fitting too, such as “-latent-dim 3 -min-feature-count 10 -learning-rate 1e-5 -beta1 0.9 -beta2 0.95” and “-latent-dim 5 -min-feature-count 10 -learning-rate 1e-5 –beta1 0.85 –beta2 0.95”. But the ranks will vary little across different parameter sets. The ranks vary by the log(x+1) or tmm transform of metabolite profile in R package edgeR [27]. The best rank for a paired metabolite and microbe among different parameter set and transform was reported in Supplementary Table 4b [12].

2.15. Benjamini-Hochberg multiple hypothesis testing correction

The multiple hypothesis testing Benjamini-Hochberg corrections were done for one source target omics pair each time. In Supplementary Table 2a [12], we showed two versions of Benjamini-Hochberg adjustment results for Shannon and other types of omics data. Q-value was the BH adjustment within one omics each time. Q-value(all) was done overall on all omics.

2.16. Robust association network construction between any two omics data type including fecal microbial MGSs

An rank average method [27] was used to combine the results of multiple inference methods to make a robust omics association network. We combined two non-linear models, one-to-many random forest and one-to-one partial Spearman’s correlation, to test the association between factor from any two types omics.

Step 1. Data preprocessing.

Dichotomous variables (such as gender) and unordered categorical variables (such as region) were re-coding into dummy variables. Frequency items such as yogurt eating habits were assigned to integers. We kept bacterial taxa with larger than 10% in all the samples for further analysis. For other type omics data, we removed variables following these rules: (i) Near zero variance. (ii) With more than 70% missing value. After filtering, we got 98 metabolites, 69 items of lifestyle questionnaire, 73 medical test parameter, 18 items of psychological questionnaire, 24 physical fitness test scores and 2352 ML/G/MGS. Missing values were filled with the median. Outliers were defined as outside of the 95% quartiles and outliers samples are removed.

Step 2. Computation of associations using multiple inference methods.

For each factor in one omics, we performed regression using RFCV function with default parameter based on all factors in one other omics and calculated RFCV R [28]. 5-fold average variable importance was extracted for step3. That is average of five RF importance score for each fold. Partial Spearman’s correlation (ppcor R package) between factors from any two types of omics was also extracted. Potential confounders were considered as described above. We also show generalized linear model results from MaAsLin R package [29]) with default parameters after adjusting for the above confounders.

Step 3. Robust networks construction.

To get the robust and strongest association between factors from any two types of omics, in other words, to filter predictor factors and target factors, we did it in two steps. First, the top 20 target factors with the highest RFCV R were selected as target factors. Then, choose predictor factors for every selected target factor: we kept predictor factors with top 30 average ranks and retained edges with partial Spearman’s correlation BH-adjusted pvalue <0.05. The average rank was computed as the sum of the ranks across the RFCV importance and absolute partial Spearman rho. For example, metabolites as target and gut microbe as the source. We regressed gut microbes against the metabolites and compute the 5-fold cross-validation predict power (RFCV R) and partial Spearman correlation for each metabolite. 20 metabolites with the highest RFCV were kept. For each of the 20 select metabolites such as VA, average ranks across RFCV and partial Spearman were done. Gut microbe biomarker for VA was found with average rank top 30th and passed the partial Spearman BH-adjusted pvalue <0.05.


For each target factor, top 5-10 average ranks source factor in each source omics type were selected as representative factors to make bar-plots using ggplot2 package (Fig. 3). The ComplexHeatmap package in R was used to plot omics triadic relation (Figs. 2–3).

3. Results

3.1. Alpha diversity and composition of the gut microbiome associated with physical fitness, exercise and body composition

The Factors related to sports were first investigated. Three exercises factors, exercise intensity level, sports habits, average time per exercise, were significantly associated (q < 0.05) with gut microbiome diversity (Shannon index) (Fig. 1 and Supplementary Table 2a). Moreover, gut microbiome diversity was impressively associated with favourable scores in most fitness tests (q < 0.05, Fig. 1a and Supplementary Table 2a). The connection between microbiota composition and host fitness can be identified by means of the PerMANOVA analysis (adonis). Most fitness tests were significantly associated with overall microbiota composition variations (called microbe covariate, q < 0.05, Fig. 1c and Supplementary Table 2b). The Inbody score (see Materials and methods 2.9), an overall body composition evaluation, was also a significant microbe covariate(q = 4.66E-03). We then used a 5-fold cross-validation random Forest to investigate the prediction power for fitness based on bacterial taxa. Interestingly, handgrip strength, vital capacity and national physique comprehensive score could be predicted by the gut microbiome (R = 0.052, 0.119 and 0.144 respectively), which was comparable with the predicting result using lifestyle factors including sports habit and exercise intensive (R = 0.255) (Supplementary Fig. 1, Supplementary Table 3d). However, exercise associated factors such as sports habits (especially resistance training) were weekly predicted (R < 0.1).

3.2. Two distinct bacterial taxa groups characterized the level of physical fitness

Based on the correlation pattern, bacterial taxa were clustered into two distinct groups: higher score in physical fitness test and more exercise, or lower score in physical fitness test and less exercise, which also indicated that gut biomarkers were consistently covariates in most physical fitness tests. Novel taxa those were associated with higher scores of vital capacity (a commonly used index to assess lung function) included Alastipes shahii (q = 0.001), Faecalibacterium prausnitzii (q = 0.006) and Bifidobacterium adolescentis (q = 0.006), whereas the other three taxa those were negatively associated with vital capacity were disease-related bacteria, including Clostridium clostridioforme (q = 0.005), Ruminococcus gravis (q = 0.038) and Escherichia coli (q = 0.045), regardless of statistical adjustments (Fig. 2, Supplementary Tables 3d and 4c). Handgrip strength (a protective factor for cardiovascular casualty [30]) was negatively associated with E. coli (q = 0.013) (Fig. 2). Age and sex-stratified vertical jump score (Supplementary Table 5) were negatively associated with E. coli (q = 0.004), while positively associated with Bacteroides cellulosolvensis (q = 0.027), Bacteroides intestinalis (q = 0.007), etc. B. cellulosolvensis and B. stercorisoris (Fig. 2). National physique comprehensive score (a summary of all the
The association between gut microbiome alpha diversity and composition with physical fitness and lifestyle. a-b, Associations between alpha diversity (Shannon) of gut microbiome and host physical fitness and lifestyle. Fig. 1. The length of the bars represents the pcor coefficient, and the marks on the bar represent the BH adjusted pcor p-value. (Shannon) of gut microbiome and host physical test including vital capacity, handgrip strength, jump score, etc.) were correspondingly associated (q < 0.1) with the above reported bacterial including *B. cellulosilyticus* and *E. coli* (Fig. 2, Supplementary Table 3d).

The higher-score-fitness-associated taxa were significantly associated with more exercise (Fig. 2). The highest positive correlation were *B. cellulosilyticus*, *B. stercorirosoris* with exercise intensity (q = 0.011, 0.059, respect), follow by faster reaction time (q = 0.032, 0.003). In addition, a trans-omics analysis between the plasma metabolites and the gut microbiota revealed that these two fitness-associated taxa were correlated with aldosterone (q = 0.002 for *B. cellulosilyticus* with aldosterone) and folic acid (q = 0.197 for *B. stercorirosoris* with folic acid), which were healthy components in both the initial and the validation cohorts (Supplementary Table 3a).

To understand the function of fitness-associated taxa, we calculated the Kyoto Encyclopedia of Genes and Genomes (KEGG) ontology (KO) abundance in the sto metabolomics data. The abundance of K04334, a curli subunit which was produced by *E. coli* and other enteric bacteria [31–33] with the potential to influence cerebral amyloid aggregation and neuroinflammation, were depleted in the individuals with higher physical fitness score(national physique comprehensive score, q = 0.023), vital capacity(q = 0.102) and jump score(q = 0.055). The abundance of K07811(trimethylamine-N-oxide reductase, an enzyme that has been linked to heart and kidney disease [34]), were also depleted in higher fitness score individuals(q < 0.1). In contrast, the abundance of K00016(lactate dehydrogenase) was positively correlated with sports habits(q = 0.082). *B. cellulosilyticus* mainly converts cellulose and sugars to acetate, propionate and succinate [35]. Propionate was sufficient to enhance sports performance in a human study [36], suggesting the link between *B. cellulosilyticus* and physical fitness. These results invite more intervention studies on physical exercise and the gut microbiome in the future.

3.3. Basic daily behavior could be predicted by gut microbes

We next evaluated another two pillars of the daily life of human, dairy consumption and sleep. In dairy consumption factors, dietary taste(salty and light), high-sugar and high-fat dietary habit and milk consumption...
Fig. 2. The association between gut microbiome with physical fitness and exercise. The colour of the heat map indicated the partial Spearman’s correlation adjusted for factors that may influence the gut microbiome, as shown in Materials and methods 2.13. BH adjusted p-value is denoted: +, q-value<0.1; *, q-value<0.05; **, q-value<0.01.

habits significantly associated with microbiota compositional variation (q < 0.1, Fig. 1d, Supplementary Table 2b). The dietary structure was the only dairy consumption factors significantly associated with alpha diversity (q < 0.1, Fig. 1b, Supplementary Table 2a). While in the prediction model, yogurt consumption was the top predictable factors (R = 0.249), following by dietary structure, high-sugar and high-fat dietary preference, smoking, milk consumption, drinking (especially low concentration alcohol, <15% alcohol concentration). Yogurt consumption habit could also be predicted by metabolite data (R = 0.223) and medical test data (R = 0.178).

In sleep factors, average daily sleep time was a significant microbe covariate (q < 0.1, Fig. 1d, Supplementary Table 2b). Staying up until midnight was the top predictable sleep factors (R = 0.135, Supplementary Fig. 1, Supplementary Table 3b). However, nap habits and sleep quality in the last month were not significantly associated with gut microbiome. Staying up until midnight could still be predicted by metabolite (R = 0.254) and medical test data (R = 0.119).

Dietary taste salty were validated to be significantly associated with microbiota compositional variation (q < 0.1, Supplementary Table 2b) in the validation cohort. Yogurt consumption also can be validated as the top predictable dairy factors (R = 0.229) in the prediction model, similarly with the observation in the discovery cohort.

3.4. Dairy consumption associated with specific microbial taxa in the healthy gut microbiome

The correlation between yogurt consumption with metabolite (R = 0.223) and medical test data (R = 0.178) motivated a more detailed investigation on the gut microbiome. Besides defecation frequency and sex, yogurt consumption could also be partially predicted by gut microbiome composition (Fig. 3, Supplementary Table 2b). A recent study cast doubts on the health benefits of probiotics, concluding that colonization of the bacteria was highly variable between individuals [37]. In both our large cohorts, Streptococcus thermophilus, a species included in commercial yogurt mainly for its thermal stability and metabolic support for other strains, was consistently detected in yogurt eaters and scaled with self-reported frequency of yogurt consumption (q = 1.13E-23, Fig. 3). The fecal bacterium mostly came from the strain S. thermophilus MN-ZLW-002, sharing 97.6% coverage and 99% identity and B. animalis subsp. lactis BB-12, sharing 99.9% coverage and 100% identity with the commonly commercial used strains from local yogurt companies. Fecal relative abundance of B. animalis associated with less stress, less bilirubin, and lower diastolic blood pressure (Fig. 3d). In the validation cohort, the association with yogurt consumption was stronger for Veillonella than B. animalis (Supplementary Table 3b), while the other parameters were unfortunately unavailable.

In contrast to S. thermophilus, B. animalis, and Veillonella, no significant increase in any Lactobacillus strains (Fig. 3). Those who used to take yogurt also showed less Clostridium botulinum(q = 0.012), a bacterium known to be elevated in a number of cardiometabolic diseases [16,21]. Intriguingly, fecal C. botulinum associated with plasma triglyceride(q = 1.719E-06), uric acid(q = 0.004), phosphoserine(q = 0.001), vitamin A(q = 0.005), and mercury(q = 0.011) (Fig. 3e), offering an explanation for the epidemiological evidence of yogurt consumption and reduced risk of gout [38].

In the validation cohort, C. botulinum also associated with mercury(q = 0.080) and to a lesser extent with vitamin A (the vitamin A association was sensitive to covariates, Supplementary Table 3a). Besides, yogurt consumption was associated with a number of favourable measurements such as higher HDL (high-density lipoprotein) cholesterol, lower uric acid and triglycerides, less cysteine, mercury and hydroxyproline (Fig. 3a).

Regarding Bifidobacterium in the gut microbiome, B. longum(q = 0.009), B. catenulatum(q = 3.775E-05) and
B. pseudocatenulatum (q = 0.006) were enriched in individuals who consumed milk (Fig. 3a and b), implying that some of the yogurt-associated differences came from its exogenous strains such as S. thermophilus and B. animalis, as well as less C. bolteae. The higher abundance of Bifidobacterium spp. and lower abundance of Ruminococcus sp. 5_1_39BFAA associated with milk intake were validated in the additional 1404 individuals (q < 0.1, Supplementary Table 3). Milk drinking was also associated with vitamin B2, B5, B6, HDL, lymphocyte counts in the blood, vital capacity, and psychological scores (q < 0.1, Fig. 3a and b, Fig. S3).

Of note, sleep late novel showed negative correlations with Veillonella atypical (q = 0.006) and 25-hydroxy vitamin D3/D (q = 4.17E-08/7.23E-09), and showed positively correlated with Clostridium hatheway (q = 0.035), Clostridium phoceensis (q = 0.011), Holdemania filiformis (q = 0.047), mercury (q = 6.12E-08), selenium (q = 3.77E-08), arsenic (q = 0.052), vitamin A (q = 8.18E-07), hydroxyproline (q = 8.00E-09) and phosphoserine (q = 0.016) (Supplementary Fig. 2a, Supplementary Fig. 3, Supplementary Table 3b). Sleep late-associated taxa were further found to positive correlate with sleep late-associated metabolites, for example enrichment of Clostridium hatheway with mercury (q = 0.056) and arsenic (q = 0.0560) (Supplementary Table 3a). We confirmed the present of tyrosinase (TYR; 1.14.18.1) for dopamine in Clostridium hatheway and Clostridium phoceensis genomes by eggnog mapper. Dopamine could induce neurotoxicity [39]. Veillonella atypical, a performance-enhancing microbe [36], was the best predictor for staying up until after midnight (Supplementary Table 3b). Thus, including sleeping as a perceivable point is essential to get a complete understanding of the gut microbiome.

4. Discussion

Although effects of sleep fragmentation on hemopoiesis have been seen despite antibiotic treatment [40], our results nonetheless suggested the additional role of gut microbiome in late sleep individuals, together with trace elements, vitamins, and host genetics [41]. Less hypocretin in mice subjected to sleep fragmentation promoted atherosclerosis [40]. Increased adiposity and decreased lean mass with sleep loss also involved toll-like receptors (TLRs) [42,43]. Here in our study, we identified novel cardiometabolic disease-associated bacterial taxa, including C. hatheway was associated with staying up late. Another Clostridium species, C. saccharolyticum have been reported enriched in the sleep-deprived mice [44]. However, the C. oroticum, in the same study, have been reported more abundant in the control group compared with sleep-deprived group.

Moreover, trace elements, including mercury, selenium, arsenic were positively associated with staying up late. Higher Hg exposure in mid-childhood has been found to be related to later sleep timing in adolescence [45]. Trimethylarsine oxide also has been linked to sleep disorders, defined as wake-up at night and leg jerks [46]. Urinary arsenic acid but not other types of urinary speciated arsenic (arsenic acid, arsenobetaine, arsphenolcholine, dimethylarsinic acid, monomethylasracnic acid) was associated with an increased trouble sleeping [47]. Even though the supplementary of selenium have been reported associated with a 20% lower risk of having difficulty falling asleep [48], in animal models,
selenium is significantly increasing wakefulness through inhibition of the enzyme prostaglandin D synthase [49]. These contradictory results indicated there might be more than one pathway between trace elements supplements and sleep habits. Gut microbiota may play a role in the pathway. More interventional and mechanistic studies are needed to figure out the conclusion.

The potential influence of physical activity on the gut microbiota has been analyzed in small cohorts of rugby athletes [3] and colorectal cancer patients [19]. Although more detailed information for physical activity is preferable, recordings compliance used mobile products, such as Fitbit, is notoriously bad in healthy individuals [50]. This large cohort at least suggested that exercising might help to improve cardio-pulmonary function (grip strength, vital capacity) and decrease the incidence of cardiometabolic diseases. Intense exercise explored for application to individuals with diseases such as prediabetes and Alzheimer’s [51,52], maybe no less important than endurance or resistance training. Our results suggested that different types of exercise could have differential impacts on the gut microbiome, and the microbiome changes could monitor for the training effect. Endurance training actually lowers testosterone [53] and could lead to hyperuricemia, especially if combined with high-fructose food and drinks and lack of dairy consumption [38].

Our large-scale analyses provided substantial support for the health benefits of yogurt consumption. The universally present bacterial taxa were Streptococcus thermophilus and Bifidobacterium animalis instead of commonly tested probiotics from Lactobacillus. An orally administered strain of B. longum has been shown to persist in 30% of individuals for at least six months [54], while we failed to detect in feces an L. casei strain gavaged to rats [55,56], suggesting general colonization differences between Bifidobacterium and Lactobacillus. The strains used by Zmora et al., including a number of Lactobacillus, Bifidobacterium, as well as Streptococcus and Lactococcus, were all detectable in various gastrointestinal sites despite laxative use and colonoscopy [37]. One hypothesis for the association with desirable cardiometabolic and psychological scores observed in our study for yogurt or milk is the production of metabolites such as folate and GABA by S. thermophilus, Bifidobacterium and Lactobacillus [57,58]. The positive association of endogenous Bifidobacterium species with milk intake is likely due to the live bacteria which help metabolize the lactose in this largely lactose-intolerant population.

Moreover, Lactobacilli have been reported to sequester heavy metals, including lead and cadmium [59]. These live or dead probiotics could potentially exert functions on the immune system or even the brain. The Veillonellaceae family which could degrade lactic acid also associated with yogurt intake and was higher in our cohort than in Hadza hunter-gathers (3.406 ± 2.373% versus 0.0002 ± 0.003%) [18,36,60]. It remains unclear whether and how diary consumption could affect the gut microbiome in other cohorts with regional differences in China.

This study provides a reference for the young gut microbiome with the physical fitness test and questionnaire data and reveals interrelationship with other omics such as trace elements and hormones that have not been included in other study designs (Supplementary Fig. 4). Interventional and mechanistic studies will be needed to see how physical activity, well-timed sleeping and dietary interventions such as yogurt, milk and vegetables might improve the gut microbiome, hormone levels, cardiometabolic and mental health.

Data and materials availability

Metagenomic sequencing data for all samples have been deposited to the CNSA (https://db.cnsg.org/cnsa/) of (CNGB) database under the accession code CNP0000426 [61,62], CNP0000289 [61,62]. Please refer to all the Supplementary Tables in our first paper of the same cohort [12].

Declaration of competing interest

The authors declare no competing financial interest.

Acknowledgments

This study was funded by the Guangdong key areas R&D projects, China (No. 2019B020229003), and supported by Guangdong Provincial Key Laboratory of Genome Read and Write(No. 2017B030301011). The authors are very grateful to colleagues at BGI-Shenzhen and China National GeneBank (CNGB) in Shenzhen for sample collection, DNA extraction, library construction, sequencing, and discussions. We thank Dr. Qiang Sun (University of Toronto), our colleagues Chen Chen and Yanmei Ju, Jinghua Wu for helpful comments.

Credit author statement


Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.medmic.2021.100038.

References
