Understanding the holobiont: crosstalk between gut microbiota and mitochondria during endurance

Microbiota-mitochondria crosstalk

Núria Mach[1*], Marco Moroldo[1], Andrea Rau[1,2], Jérôme Lecardonnel[1], Laurence Le Moyec[3,4], Céline Robert[1,5], and Eric Barrey[1]

[1] GABI UMR 1313, INRAE, Université Paris-Saclay, AgroParisTech, Jouy-en-Josas, France

[2] BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne, Estrées-Mons, France

[3] Université d'Évry Val d'Essonne, Université Paris-Saclay, Évry, France

[4] MCAM UMR7245, CNRS, Muséum National d'Histoire Naturelle, Paris, France

[5] École Nationale Vétérinaire d'Alfort, Maisons-Alfort, France

Correspondence to Núria Mach

Email address: nuria.mach@inrae.fr

ABSTRACT

Endurance exercise has a dramatic impact on the functionality of mitochondria and on the composition of the intestinal microbiome, but the mechanisms regulating the crosstalk between these two components are still largely unknown. Here, we sampled 20 elite horses before and after an endurance race and used blood transcriptome, blood metabolome and fecal microbiome to describe the gut-mitochondria crosstalk. A subset of mitochondria-related differentially expressed genes involved in pathways such as energy metabolism, oxidative stress and inflammation was discovered and then shown to be associated with

butyrate-producing bacteria of the *Lachnospiraceae* family, especially *Eubacterium*. The mechanisms involved were not fully understood, but through the action of their metabolites likely acted on *PPARγ*, the *FRX-CREB* axis and their downstream targets to delay the onset of hypoglycemia, inflammation and extend running time. Our results also suggested that circulating free fatty acids may act not merely as fuel but drive mitochondrial inflammatory responses triggered by the translocation of gut bacterial polysaccharides following endurance. Targeting the gut-mitochondria axis therefore appears to be a potential strategy to enhance athletic performance.

Keywords: endurance/ holobiont/ horse/ microbiota/ mitochondria

INTRODUCTION

To keep up with energy demand and to maintain homeostasis, endurance exercise modifies multiple systems, ranging from the whole-body level to the molecular level (Mach & Fuster-Botella, 2017; Clark & Mach, 2017). In recent years, our understanding of the role played by mitochondria during this kind of challenge has expanded far beyond its bioenergetic capacity, which is represented by well characterized pathways such as oxidative phosphorylation (OXPHOS), fatty acid β-oxidation (FAO) and the tricarboxylic acid (TCA) cycle (Pfanner *et al*, 2019). Indeed, it is now widely accepted that mitochondria regulate cytosolic calcium homeostasis and cellular redox status, that they generate much of the cell reactive oxygen species (ROS), and that they are involved in steroid and heme biosynthesis, urea degradation, apoptosis and initiation of inflammation through inflammasomes (Vezza *et al*, 2020; Jackson & Theiss, 2020; Wong *et al*, 2016; Chinnery & Hudson, 2013).

It has also been established that the mitochondrial genome (mtDNA) and the nuclear genome are constantly communicating with each other to regulate the aforementioned pathways. For example, most of the proteins involved in OXPHOS and mitochondrial functions like

mtDNA replication and expression, mtDNA repair, redox and energy regulation are encoded by the nuclear genome and require specific targeting signals to be directed from the cytosol to mitochondrial surface receptors and then to the proper mitochondrial sub-compartments (Pfanner *et al*, 2019). The transcriptional programs of mitochondria comprise over 1,600 nuclear-encoded mitochondrial proteins (Pagliarini *et al*, 2008; Richter-Dennerlein *et al*, 2016). Any alteration in the OXPHOS and FAO processes, in mitochondrial membrane potential ($\Delta\Psi m$), mitochondrial biogenesis, ROS production and inflammation can have a deep impact on the response to endurance exercise. For instance, increased mitochondrial biogenesis improves muscle endurance performance due to higher rates of OXPHOS and FAO (Hood *et al*, 2011). On the contrary, lower metabolic rates, increased ROS production and acidosis during prolonged exercise are associated with fatigue and inability to maintain speed (Rapoport, 2010). Endurance or long exercise and mitochondrial functions are strictly intertwined and influence each other.

The gut microbiota is considered a central organ because of its direct and indirect roles in host physiology, including improved metabolic health and athletic performance (Keohane *et al*, 2019; Barton *et al*, 2017; Scheiman *et al*, 2019). A healthy ecosystem in horse includes numerous, highly dominant taxa along with a multitude of minor players with lower representation, but important metabolic activity (Mach *et al*, 2020; Costa & Weese, 2018; Kauter *et al*, 2019).

The interdependence of gut microbiota and mitochondria is being increasingly recognized, with many diseases originating from mitochondrial dysfunctions linked to well-described changes in gut microbiota (Cardoso & Empadinhas, 2018; Bajpai *et al*, 2018; Karlsson *et al*, 2013; Yardeni *et al*, 2019; Franco-Obregón & Gilbert, 2017; Saint-Georges-Chaumet *et al*,

2015; Mottawea *et al*, 2016; Saint-Georges-Chaumet & Edeas, 2018). This complex interplay occurs principally through endocrine, immune and humoral signalling (Mottawea *et al*, 2016). Enteric short-chain fatty acids (SCFAs), the major class of metabolites produced from bacterial fermentation of non-digestible carbohydrates, are widely thought to mediate the relationship between the gut microbiota and the mitochondria in different tissues. Branched-chain amino acids (BCAAs), secondary bile acids, ROS, nitric oxide (NO), and hydrogen sulfide (H_2S) are also thought to play at least a partial role in this molecular interchange (Clark & Mach, 2017).

While formal proof is missing in horse athletes, a prevailing hypothesis is that the gut microbiota and its metabolites regulate crucial transcription factors and coactivators involved in mitochondrial functions that underpin endurance performance (Hawley *et al*, 2018). In mice models, gut microbiota depletion via broad-spectrum antibiotics showed reduced production of SCFAs, lower bioavailability of serum glucose, decreased endurance capacity and impairment of the *ex vivo* skeletal muscle contractile function (Nay *et al*, 2019). In close agreement, gut microbiota depletion also triggered a reduction of both faecal SCFA content and circulating concentration of SCFAs coupled to a drop in running capacity in mice (Okamoto *et al*, 2019). In contrast, mice with *Veillonella* in their intestinal ecosystem showed significantly increased submaximal treadmill run time to exhaustion (Scheiman *et al*, 2019), prompting the authors to speculate that the lactate generated during sustained bouts of exercise could be accessible to the microbiota and converted into SCFAs that ultimately enhanced energetic resilience and stamina. Alternatively, there may be other mechanisms through which gut microbiota and its metabolites relate to mitochondria, including but not limited to the regulation of mitochondrial oxidative stress (Jones & Neish, 2014; Franco-Obregón & Gilbert, 2017), as well as the activation of the inflammasome and the production

of inflammatory cytokines, all of which are key players in the adaptation to endurance exercise (Clark & Mach, 2017; Mach & Fuster-Botella, 2017).

We recently presented a three-pronged association study, connecting horse gut microbiota with untargeted serum metabolome data and measures of host physiology and performance in the context of endurance (Plancade *et al*, 2019). We found no significant associations between the gut ecosystem and serum metabolites, especially those relying heavily on mitochondrial OXPHOS, FAO, TCA and gluconeogenesis. The number of annotated metabolites in the study was likely insufficient to reliably encompass the given mitochondrial functions. To further advance our knowledge of the molecular basis for the gut-mitochondria crosstalk that support the adaptation to long exercise, in this work we tethered whole blood transcriptome profiling to our previous metabolome and metagenome data. In doing so, we sought to identify the ways in which mitochondrial and nuclear transcriptomes coordinate with each other, and how gut microbiota and circulating metabolites can dynamically modulate this process. By jointly characterizing the whole blood transcriptome, metabolome, fecal microbiota and SCFAs of 20 elite horses competing in an endurance race, we aim to provide a functional readout of microbial activity and improve our understanding of the gut microbiota-mitochondria axis during long exercise.

RESULTS

To elucidate how mitochondria and gut microbiota are linked during endurance exercise, we studied 20 healthy endurance horses selected from the cohort already described by Plancade *et al*. (2019) and Le Moyec *et al*. (2019) (Fig 1). All of the animals were of similar age and performance level (Table EV1). They performed a long exercise during about 8 hours at an average speed of 17.1 ± 1.67 km/h with some rest periods every 30-40 km.

Whole transcriptome profiles, proton nuclear magnetic resonance (^{1}H NMR) metabolome profiles and biochemical assay data were obtained from blood samples collected at both T0 (pre-ride) and T1 (post-ride), while SCFAs measurements, 16S rRNA data and the concentration of bacteria, anaerobic fungi and ciliate protozoa were generated from fecal samples at T0 alone.

Blood transcriptome profiles and mitochondrial related genes

To gather information about the global structure of the blood transcriptome, a scaled principal component analysis (PCA) was carried out on the expressed genes (n =11,232). The first component accounted for 43% of the total variability and revealed a marked separation of the two time points (Fig 2A). We then carried out a standard differential analysis between the two time points. After Bonferroni correction of the raw p-values, a total of 6,021 differentially expressed genes (DEGs) was obtained at an adjusted $p < 0.05$, of which 2,658 were upregulated and 3,363 downregulated at T1 respect to T0 (Table EV2; Fig 2B).

These results were then complemented using a weighted gene co-expression network analysis (WGCNA) (Langfelder & Horvath, 2008) on the expressed genes. WGCNA identified three gene modules, corresponding to 7,914 genes, that were correlated to the ^{1}H NMR and biochemical assay metabolites (Table EV3). These genes strongly overlapped with the set of DEGs, 91.1% of which (i.e., 5,486 out of 6,021 genes) were included among them. The metabolites which showed the highest levels of correlation with the gene modules were bilirubin, non-esterified fatty acids (NEFAs), tyrosine, lactate and, to a lesser extent, β-hydroxybutyrate (BHB) (see next paragraph, Fig EV1).

Because we were especially interested in understanding the role played by mitochondria in our biological system, we then decided to study in more detail the features related to these

organelles. To this end, we used a literature-based meta-analytic approach to build a non-redundant consensus list of 2,082 genes related to mitochondria, based on the information available in the Integrated Mitochondrial Protein Index (IMPI) (Smith & Robinson, 2019), the Mitocarta Inventory (Calvo *et al*, 2016) and the literature (Expanded View Information and Table EV4). This consensus list was also descriptively annotated to gain global insight into the main biological functions represented within it. A total of 80 third-level KEGG hierarchies were identified, with a strong representation of pathways related to carbohydrate and lipid metabolism (8 and 9 ontology terms, respectively). Eight pathways were associated with amino acid metabolism, while five were linked to the apoptosis process (Fig EV2). This subset was crossed with each of the DEG and WGCNA module gene lists. In the case of the DEG list, the intersection with this subset yielded a total of 801 genes (Table EV5 and Fig 2C). Both the Fisher's exact test and the hypergeometric test showed strong levels of over-representation, with *p*-values of 1.0×10^{-5} and 9.02×10^{-7}, respectively. In the case of the WGCNA gene modules, the intersection included 1,011 genes (Fig 2C). Again, both statistical tests indicated strong enrichment, with *p*-values of 1.0×10^{-5} and 1.07×10^{-5}, respectively.

The set of 801 genes in the intersection of the mitochondrial consensus and DEG list, which will be referred to hereafter simply as "mt-related genes", was selected for the downstream steps of analysis. All of the mt-related genes were encoded by the nuclear genome except for *MT-ND6* (MT-NADH dehydrogenase, subunit 6), which was encoded by the antisense strand of the mt-DNA (Fig 2D). These mt-related genes were further characterized to gather information about their molecular function. The functional analysis showed that roughly 75 % of these genes were directly involved in energy metabolism (i.e., pathways such as OXPHOS and FAO) and metabolite synthesis and degradation (Fig EV3). For instance, we observed an

enrichment of genes related to nutrient transport across the mitochondrial inner membrane (*TOM/TIM* units, *VDAC*, *MPC1*, *ACAA1*, members of the mitochondrial carrier family *SLC25* and of the pyruvate dehydrogenase kinase isozyme), fatty acid metabolism (*ELOVL7*, *SIRT5*, and *ACAD* members), lipogenesis *(FASN* and *PPARγ)*, and fatty acid channelling into oxidation (*CPT1B*, *ACADVL*, *ACOT9*, *ACOX1*, *ACSL1*, *ACSL4*, *ACSS3*). Additionally, key genes involved in the mitochondrial biogenesis (*POLG* and *POLG2*), mitochondrial fission (*PINK1*), mitochondrial fusion (*OPA1*), mitophagy (*BNIP3* and *PINK1)*, oxidative stress (*SOD1*, *SOD2*, glutathione S-transferases and glutathione peroxidases families), and the resolution of lipopolysaccharide induced pro-inflammatory pathway (*C1QBP*) were also found in this list. *CREB*, the most potent activator of *PGC-1α* (Wu *et al*, 2006) was also differentially expressed upon endurance exercise.

Our data further indicated that among the mt-related genes, at least 21 genes encoding rps and rpl proteins of the small and large subunits of ribosomes (*rpl3*, *rpl4*, *rpl5*, *rpl6*, *rpl18*, *rpl23*, *rpl27*, *rpl36*, *rps3*, *rps8*, *rps9*, *rps10*, *rps11*, *rps12*, *rps13*, *rps14*, *rps15*, *rps16*, *rps17*, *rps18*, *rps19*) (Janouškovec *et al*, 2017; Esser *et al*, 2004; Maier *et al*, 2013) were common with the α-Proteobacteria. This was also the case for the methionine sulfoxide reductase A (*MSRA*) and NAD(P)H dehydrogenase quinone 1 (*NQO1*) (Crisp *et al*, 2015), consistent with a vertical origin in the mitochondrial endosymbiont.

^{1}H NMR metabolome, biochemical assay and acetylcarnitine profiles

The ^{1}H NMR metabolome and biochemical assay profiles used in this paper were gleaned from our previous works (Plancade *et al*, 2019; Le Moyec *et al*, 2019). Briefly, a total of 50 ^{1}H NMR known metabolites was detected in the plasma, including several amino acids, energy metabolism-related metabolites and organic osmolytes (Table EV6). Three well-

known microbial derived metabolites were ascertained, including formate, dimethyl sulfone and trimethyl amine oxide (TMAO).

The relative abundance values of these circulating metabolites fell within the normal reference range for healthy horses. However, the concentration of lactate (a proxy for glycolytic stress and disturbances in cellular homeostasis (Hawley *et al*, 2018)) was significantly increased after the race, as well as the levels of fatty acids from lipoproteins and of certain amino acids, namely alanine, branched amino acids such as leucine, valine and *iso*-valerate, glutamate, glutamine and aromatic amino acids such as tyrosine and phenylalanine. Ketone bodies were slightly increased after the race (i.e., acetoacetate and acetate; Table EV6). In the case of biochemical profiles, all of the horses showed above-average concentrations for total bilirubin**,** creatine kinase, aspartate transaminase, and serum amyloid A after the race. The NEFA and BHB concentrations also showed similar patterns (Table EV7).

Fecal short chain fatty acids measurements, 16S rRNA data and microorganism concentrations

The microbiota composition and derived-metabolites were obtained from Plancade *et al*. (2019), but it is important to note that the 16S rRNA raw sequences were re-analysed using the QIIME 2 plugin, which quantitatively improved results over QIIME 1 by enhancing the pre-processing of sequenced reads, the taxonomy assignment, the phylogenetic insertion and the generation of amplicon sequence variants (ASVs) (Callahan *et al*, 2017). A total of 519,866 high-quality sequence reads were obtained (mean per subject: 21,131 $\pm$ 15,625, range: 6,036 – 57,389). Reads were clustered into 3,384 chimera- and singleton-filtered ASVs at 99% sequence similarity (Table EV8). The intestinal microbial community found in

the total set of 20 individuals as a whole was made up of a core of 23 genera, the core being defined as the genera shared by 99% of all sampling events with a minimum 0.1% mean relative abundance. Overall, 61% of the core genera belonged to the Firmicutes phylum, mainly to the *Lachnospiraceae* and *Ruminococcaceae* families (Fig EV4A). A total of 100 unique genera was identified in the microbiota (Table EV9). The majority of these genera (80%) fell among the 20 most abundant (Fig EV4B) and accounted for more than 75% of the sequences in the data (Table EV9). To deeper understand how the gut microbiota functioned, SCFAs, pH measurements and the loads of anaerobic fungi, protozoa and total bacteria in feces were also investigated. The main products of microbial fermentation were acetate, propionate and butyrate, with ratios ranging from 60:32:8 to 76:19:5. Small amounts of branched chain fatty acids (*iso*-butyrate, valerate and *iso*-valerate) were also detected (Table EV10). Although bacteria represented the major portion of the fecal microbiota in our horses, the relative concentrations of anaerobic fungi and ciliate protozoa were 0.82 and 0.76, respectively (Table EV11).

Integration of transcriptome, ^{1}H NMR metabolites, biochemical parameters, fecal microbiota, SCFA and microorganism concentrations

After the identification of the 801 mt-related genes that were regulated by endurance exercise, it remained to be determined which genes were interconnected to the gut microbiota and responded to specific circulating molecules. To this aim, we applied four independent statistical methods using the mt-related genes as the response variable and the other data sets, namely ^{1}H NMR metabolome, biochemical assay profiles, fecal microbiota, fecal SCFAs and the concentrations of bacteria, anaerobic fungi and protozoa as exploratory variables.

We first used global non-metric multidimensional scaling (NMDS) ordinations to visualize the structure of mt-related gene expression (ordination stress = 6%, k = 2, non-metric fit r^2 =

0. 0.996, linear fit r^2 = 0.988) and we then fitted all sets of explanatory variables to the ordination to find the most influential variables (Fig 3A). Bacteria such as *Oribacterium*, *Rikenellaceae* RC9, *Ruminococcaceae* NK4A214, unclassified rumen bacterium and *Clostridium sensu stricto* showed the strongest correlation to all ordinations, together with some microbial-derived metabolites (i.e., dimethyl sulfone, formate, valerate and *iso*-valerate) and the plasmatic NEFA (adjusted $p < 0.05$; Fig 3A-B).

To control for spatial variance and to identify the minimal combination of non-redundant covariates that would best fit with the mt-related gene profiles, a more rigorous multivariate distance-based redundancy analysis (db-RDA) was used on constrained NMDS ordinations. In agreement with the aforementioned results, the expression of mt-related genes responded most strongly to bacteria such as *Treponema* (r^2_{adj} = 0.48, p = 0.007), followed by *Butyrivibrio* (r^2_{adj} = 0.45, p = 0.002), plasmatic NEFA (r^2_{adj} = 0.38, p = 0.003), *Fibrobacter* (r^2_{adj} = 0.31, p = 0.0008) and *Oribacterium* (r^2_{adj} = 0.24, p = 0.003; Fig 3C). The ^{1}H NMR metabolites and fecal SCFAs and the concentration of fecal microorganism did not directly contribute to the variation of mt-related genes. Therefore, they were not selected by the dbRDA model.

These findings were further confirmed by an RDA forward-selection model based on the Akaike information criterion. Specifically, *Oribacterium* (F = 7.26, $p < 0.005$), *Fibrobacter* (F = 2.96, $p < 0.005$), *Butyrivibrio* (F = 2.91, $p < 0.005$), *Agathobacter* (F = 2.12, $p < 0.005$), *Treponema* (F = 7.15, $p < 0.01$), unclassified rumen bacterium (F = 1.96, $p < 0.01$), and the concentration of plasmatic NEFA (F = 2.91, $p < 0.005$) explained most of the variance observed in mt-related genes (Fig 3D). However, the ^{1}H NMR metabolites, the fecal SCFAs and concentration of microorganisms were not found to contribute significantly to the variability in mt-related gene expression. The first constrained axis (RDA1) explained 44% of the variance in mt-related gene expression, and the second (RDA2) explained 9.4% of the

276 variance; on the other hand, the two first unconstrained axes (PC1 and PC2) represented less
277 than 8% of the total variance, i.e., much less than that explained by the explanatory variables
278 together.
279 To uncover other potential underlying mechanisms of mitochondrial regulation, we then
280 sought to examine the relationships existing among all aforementioned data sets by adding a
281 further categorical variable, namely the racing performance of horses. To perform this task,
282 we used the DIABLO framework from mixOmics (Singh *et al*, 2019). While the mt-related
283 genes showed high levels of covariation with the fecal microbiota ($r^2 > 0.91$, Fig 4A), it was
284 not possible to identify a tight relationship with the other data sets. A more fine-grained view
285 of this biological system was then obtained by focusing on pairwise correlations between
286 variables. The first component of the DIABLO analysis highlighted a significant link
287 between a subset of 45 mt-related genes (all encoded by the nuclear genome) and four gut
288 taxa (i.e., the genus *Mogibacterium*, the species *Eubacterium coprostanoligenes* and the
289 groups *Rikenellaceae* RC9 and *Ruminococcaceae* NK4A214). A link to blood metabolites
290 related to energy supply (i.e., methyl groups of FAs and choline-containing compounds) and
291 metabolites related to the TCA cycle such as glutamine, glutamate, and α-glucose was also
292 unveiled (Fig 4B-C). Moreover, mt-related genes co-occurred with pronounced variations of
293 microbiota derived metabolites, including fecal acetate and valerate, and plasmatic
294 concentrations of TMAO, dimethyl sulfone, and formate. The subset of 45 mt-related genes
295 was functionally enriched in pathways related to fatty acid β-oxidation, mitochondrial
296 apoptosis and biogenesis, respiratory electron transport and signalling and innate immune
297 system response (Table EV12).
298 Finally, an rCCA analysis was also carried out to study in a more targeted way the
299 relationships between mt-related genes and gut microbiota. In this case, the most relevant
300 associations were represented by 90 genes and 9 bacterial genera (Table EV13). Overall, this

301 method largely validated the associations already detected with the other aforementioned
302 approaches. First, all of the four genera highlighted by DIABLO were confirmed, as well as
303 the genus *Fibrobacter*, which had already been detected using NMDS and the db-RDA
304 method. Second, three more taxa found with rCCA appeared to be functionally related to
305 other previously identified microorganisms, thus providing indirect support to those findings.
306 This was the case for *Ruminococcaceae* UCG-002, *Pseudobutyrivibrio* and the *Eubacterium*
307 *hallii* group.
308
309 **DISCUSSION**
310 In this work, we present an integrative study that combines the whole blood transcriptomic
311 with untargeted serum metabolome data, blood biochemical assay profiles and gut
312 metagenome in 20 equine-athletes. We assumed the adaptive response to extreme endurance
313 exercise was essentially explained through the gut-mitochondria axis. Indeed, the different
314 and complementary statistical approaches that we used confirmed this hypothesis,
315 highlighting that the two main omic layers at play were the mt-related genes and the gut
316 microbiota composition ($r^2 > 0.91$).
317
318 Whole blood transcriptome underlines in a clear manner the global response to exercise in
319 equines (Mach *et al*, 2017b, 2016; Capomaccio *et al*, 2013; Barrey *et al*, 2006; Ropka-Molik
320 *et al*, 2017), including the inflammatory response of the muscle associated with sarcolemma
321 permeability and rhabdomyolysis (Barrey *et al*, 2006) (displayed in our study by the high
322 levels of plasma creatine kinase and aspartate aminotransferase after the race). Yet, it remains
323 to be explored whether the whole blood transcriptome reflects the physiological events
324 occurring at the mitochondrial level, notably in the tissues that are highly solicited under
325 endurance, like for instance, the skeletal muscles, the heart and the liver (Gunn, 1987). In

addition, whether the transcription of nuclearly-encoded and mitochondrially-encoded genes are regulated in a coordinated way is much less well understood. We therefore compared the transcriptome profile in equine-athletes before and after exercise and we used a meta-analytical approach that allowed us to identify 801 differentially expressed genes that were putatively linked to mitochondria. These genes fit neatly into the well-characterized context of adaptive mitochondrial regulation to endurance and mostly belonged to molecular pathways such as mitochondrial biogenesis, energy metabolism through OXPHOS and FAO, resistance to oxidative stress, mitophagy and inflammation regulation.

We then specifically focused on the mechanisms underlying the biological links between the aforementioned mt-related genes and the gut microbiota to untangle the gut-mitochondria crosstalk. The interdependence of mitochondria and gut microbiota is underscored by several lines of evidences (Yardeni *et al*, 2019; Gruber & Kennedy, 2017; Mottawea *et al*, 2016; Zhang *et al*, 2020; Han *et al*, 2017; Qi & Han, 2018; Ruiz *et al*, 2020; Saint-Georges-Chaumet & Edeas, 2018), although the range and extent of this interplay are largely unknown. For example, (Mottawea *et al*, 2016) showed that butyrate-producing bacteria and mitochondrial proteins were positively correlated, suggesting a signalling role for butyrate in mitochondrial gene expression. In support of this observation, our results revealed that several functionally redundant butyrate-producing bacterial families were associated with the mt-related genes, namely *Lachnospiraceae* (*Oribacterium, Butyrivibrio, Agathobacter* and *Eubacterium* spp.*), Ruminococcaceae, Spirochaetaceae* (*Treponema* spp.) (Vital *et al*, 2015; Vacca *et al*, 2020; Gharechahi *et al*, 2020) and *Rikenellaceae* (Vital *et al*, 2015). The bioavailability of butyrate is obviously related to endurance performance because of the role played by this molecule in energy metabolism (Mollica *et al*, 2017). Beyond the scope of its energy producing capacity, butyrate is also known to induce the expression of *PPARγ* gene

(Gao *et al*, 2009) and downstream targets in different cells. Our blood transcriptomic analysis indicated an upregulation of *PPARγ* following exercise, raising the possibility that the enrichment in butyrate-producing bacteria increased the expression of this transcription factor and downstream signaling, leading to fatty acid shuttling into and oxidation by the mitochondria. Notably, the redox imbalance during strenuous exercise might be also attenuated by butyrate (Mottawea *et al*, 2016; Dobashi *et al*, 2011). The potential of butyrate to improve exercise capacity has been further posited by Gao *et al*. (2009) and Henagan *et al*. (2015), who observed that the supplementation of this molecule improved the oxidative skeletal muscle phenotype, its mitochondrial content and its proportion of type I fibers. Concomitantly, it is possible that serum lactate, which appeared to be significantly increased in our horses upon prolonged exercise, entered the gut lumen, where it was subsequently transformed into butyrate by *Eubacterium hallii* (Duncan *et al*, 2004; Scheiman *et al*, 2019). Indeed, *Eubacterium hallii* was associated to a subset of mt-related genes according to our rCCA analysis. *Eubacterium*-derived butyrate could then be absorbed into the portal vein and serve as an energy source to the different organs. Another plausible mechanisms employed by microbiota to communicate with mitochondria involved valerate, a branched SCFA formed from protein and amino acid degradation (Fernandes *et al*, 2014). It may be debated whether the urea produced by the host during endurance could be hydrolyzed by commensal gut microbiota resulting in valerate.

Beyond SCFAs, the secondary bile acids could also play an important role in gut-microbiota crosstalk. The genera *Eubacterium* and *Clostridium,* which contributed significantly to our biological system, have the capacity to degrade 5–10% of the primary bile acids forming secondary bile acids (Gérard, 2013). Secondary bile acids might interact with the mitochondria via the activation *FXR-CREB* axis. *CREB,* which was significantly increased in

376 our horses after the endurance race, is a sensor of energy charge and other stress signals, and
377 is a regulator of metabolism that activates autophagy and lipid catabolic functions (Seok *et al*,
378 2014). Therefore, a picture emerges that, under conditions that foster an increased
379 colonization by these microorganisms, the production of butyrate, valerate and secondary bile
380 acids in the intestine is likely increased, with potential effects on the mitochondria
381 functionality and endurance performance (Fig 5). Yet, evidence of causality of microbiome-
382 derived metabolites on the gut-mitochondria crosstalk remains elusive.
383
384 The coordination between mitochondria and gut microbiota was presumably regulated by the
385 circulating free fatty acids, the so-called NEFAs. It is becoming clearer that NEFAs not only
386 serves as energy source in the working muscles but act as extracellular signaling molecules
387 that modulate the production of chemokines and cytokines, and the synthesis of pro-
388 inflammatory lipid-derived species (Rodríguez-Carrio *et al*, 2017). Thus, a provocative
389 extension of our work suggests that increased release of NEFAs participated in the
390 mitochondrial regulation of the inflammatory processes elicited by oxidative stress, microbial
391 dissemination and microbial lipopolysaccharides translocation outside of the gastrointestinal
392 tract, commonly observed in endurance athletes (Fielding & Dechant, 2012). Increased
393 release of free fatty acids may dampen the inflammatory response and prevent or mitigate the
394 negative effects of redox imbalance. Supporting this notion, the mitochondrial sirtuin (*SIRT5*)
395 and *SIRT1*, which have an anti-inflammation function (Wang *et al*, 2017), were found to be
396 increased during endurance.
397
398 The observations presented herein indicate that the horse could be considered as an
399 interesting *in vivo* model for research in the field of human exercise given its large body size,
400 the aptitude for endurance exercise (Votion *et al*, 2012; van der Kolk *et al*, 2020), that is,

high baseline maximal oxygen uptake ($VO_{2max;}$ ~120 $mL \cdot min^{-1} \cdot kg^{-1}$) and the ability to sustain work at a high percentage of VO_{2max} without either the accumulation of exponential levels of blood lactate or skeletal muscle fatigue and the exercise economy (van der Kolk *et al*, 2020; Cottin *et al*, 2010; Goachet & Julliand, 2015). The level of exercise performed by a horse during an endurance competition is similar to that of a human marathon runner (Mach *et al*, 2016; Capomaccio *et al*, 2013) or ultramarathon runner (Scott *et al*, 2009). Nevertheless, despite the usefulness of this model, the differences between the microbiota of horses and humans are relevant. In contrast to what happens in humans, in horses the cecum is large relative to the total gastrointestinal tract and it is an important site for the fermentation of plant materials. Our study presents other limitations. Although the omic approaches used here are considered robust and generate high quality data, they still present several limitations. For instance, 16S rRNA sequencing measures the relative abundance of bacterial genera contained in it, but it does not give any information about its actual functionality, which should be therefore evaluated using other methods, such as for instance metatranscriptomics. Moreover, in our case, ^{1}H NMR has been able to detect only metabolites at high concentrations, like in the cases of amino acids, lipids, choline and N-acetylglucosamine. In this regard, the combination of ^{1}H NMR and mass spectrometry should result in better coverage of metabolites derived from bacteria, metabolites that are produced by the host and then modified by bacteria and metabolites that are *de novo* synthesized by bacteria. Lastly, it still remains to be determined how the individual components of blood, including plasma, platelets, erythrocytes, nucleated blood cells and exosomes reflect the transcriptomic profiles in horses. Upon endurance, contracting muscles release proteins and metabolites that have endocrine-like properties (Hawley, 2020), but they might also release long non-coding RNAs, myo-miRs and circulating cell-free respiratory competent mitochondria (Al Amir Dache *et al*,

2020; Song *et al*, 2020) that might participate in the aforementioned crosstalk by using the microbiota-derived metabolites.

Taken together, the present study offers extensive novel insight into the mitochondria-gut microbiota axis and opens the way for mechanistic studies that will lead to a better understanding of the orchestrated molecular pathways that underpin endurance adaptations and contribute to the holobiont biology. This is the first description of how metabolites derived from commensal gut microbiota (SCFAs and secondary bile acids), or produced by the host and biochemically modified by gut bacteria (lactate and urea), might influence the genes related to the mitochondria and involved in energy production, redox balance and inflammatory cascades, making them a potential therapeutic target for the endurance. The activation of *PPARγ* and the *FRX-CREB* axis are likely key mechanisms through which SCFAs and bile acids coordinately engage multiple converging pathways to regulate mitochondrial functions, including fatty acid uptake and oxidation to forestall hypoglycemia and ensure longer running time.

Our results also suggest that free fatty acids may not only serve as an important fuel for skeletal muscle during endurance, but may also regulate mitochondrial inflammatory responses through a plethora of mechanisms, the principal one likely being the modulation of the intestinal barrier-ROS production and lipopolysaccharide translocation. Further research focusing on the role that gut microbiota plays on the mitochondrial function across a wide range of tissues and cell types may be highly informative to improve the athlete's energy metabolism, redox status and inflammatory response.

MATERIALS AND METHODS

Ethics approval

The study protocol was reviewed and approved by the local animal care and use committee (ComEth EnvA-Upec-ANSES, reference: 11-0041, dated July 12th 2011) for horse study. All the protocols were conducted in accordance with EEC regulation (n^{o} 2010/63/UE) governing the care and use of laboratory animals, which has been effective in France since the 1st of January 2013. In all cases, the owners and riders provided their informed consent prior to the start of study procedures with the animals.

Animals

Twenty pure-breed or half-breed Arabian horses (7 females, 3 male, and 10 geldings; mean ± SD age: 10 ± 1.69) were selected from the cohort used by Plancade *et al*. (2019) (Table EV1). The 20 horses were selected following these criteria: (1) enrolment in the 160 km or 120 km category; (2) blood sample collection before and after the race; (3) feces collection before the race; (4) absence of gastrointestinal disorders during the four months prior to enrolment; (5) absence of antibiotic treatment during the four months prior to enrolment and absence of anthelmintic medication within 60 days before the race; (6) a complete questionnaire about diet composition and intake.

Among the 20 horses selected for this study, 16 horses were enrolled for the 160 km category and four for the 120 km category. In the 160 km category, two animals were eliminated due to tiredness after 94 km and 117 km, respectively, and five horses failed a veterinary gate check due to lameness after 94 km (n = 1), and after 117 km (n = 4). In the 120 km category, one horse was eliminated due to metabolic troubles after 90 km (Table EV1).

The weather conditions, terrain difficulty and altitude were the same for all the participants enrolled in the study as all races (120 and 160 km) took place during October 2015 in

475 Fontainebleau (France). The average air temperature was 15 ºC, with a maximum of 20 ºC
476 and a minimum of 11 ºC, the average air humidity was 88%, and no rain was recorded.
477
478 As detailed by Plancade *et al*. (2019), to ensure sample homogeneity, the participating horses
479 were subject to the same management practices throughout the endurance ride and passed the
480 International Equestrian Federation (FEI) compulsory examination before the start. Animals
481 were fed 2–3 hours before the start of the endurance competition with *ad libitum* hay and 1
482 kg of concentrate pellets. During the endurance competition, all the animals underwent
483 veterinary checks every 20 to 40 km, followed by recovery periods of 40 to 50 minutes (in
484 accordance with the FEI endurance rules). After each veterinary gate check, the animals were
485 provided with *ad libitum* water and hay and a small amount of concentrate pellets.
486
487 **Transcriptomic microarray data production and pre-processing**
488 Blood samples for RNA extraction were collected from each animal at T0 and T1 using
489 Tempus Blood RNA tubes (Thermo Fisher). Because blood interacts with every organ and
490 tissue in the body and has crucial roles in immune response, inflammation and physiological
491 homeostasis (Mohr & Liew, 2007), blood-based transcriptome was carried-out as a means for
492 exploring the response to endurance.
493 Total RNAs were then isolated using the Preserved Blood RNA Purification Kit I (Norgen
494 Biotek Corp., Ontario, Canada), according to the manufacturer's instructions. RNA purity
495 and concentration were determined using a NanoDrop ND-1000 spectrophotometer (Thermo
496 Fisher) and RNA integrity was assessed using a Bioanalyzer 2100 (Agilent Technologies,
497 Santa Clara, CA, USA). All the 40 RNA samples were processed.
498 Transcriptome profiling was performed using an Agilent 4X44K horse custom microarray
499 (Agilent Technologies, AMADID 044466). All of the steps were performed by the

@BRIDGe facility (INRAE Jouy-en-Josas, France, http://abridge.inra.fr/), as described previously (Mach *et al*, 2017b, 2016).

The horse array was annotated as described by Mach *et al*. (2017b, 2016). In a limited number of cases, a manual annotation step was also included. Probe intensities were background-corrected using the "normexp" method, $\log_2$ scaled and quantile normalized using the limma package (version 3.1.42.2) (Smyth, 2004) in the R environment (version 3.6.1). Only the probes which presented, on at least two arrays, intensity values at least 10% higher than the 95% percentile of all the negative control probes were kept. Subsequently, controls were discarded and the probes corresponding to genes were summarized. The obtained expression matrix "E1" was processed with the arrayQualityMetrics R package (version 3.42.0) (Kauffmann *et al*, 2009) for quality assessment. No outliers were detected.

Transcriptome statistical analysis

The differential analysis was performed using the limma R package. A linear model was fitted for each gene, setting the time, the sex, the distance and whether the animal was eliminated from the race as fixed effects, and comparing T1 to T0. The individual was included as a random effect using the "duplicateCorrelation" function (Smyth *et al*, 2005). The *p*-values were Bonferroni corrected setting a threshold of 0.05. The expression matrix "E1" was then used to perform a scaled PCA analysis with FactoMineR R package (version 2.4) (Lê *et al*, 2008).

To confirm the results of the DE analysis, the WGCNA method was also run on the "E1" matrix using the WGCNA R package (version 1.69) (Langfelder & Horvath, 2008). The parameters for the analysis were set as follows: "corFnc" = bicor, "type" = signed hybrid, "beta" = 10, "deepSplit" = 4, "minClusterSize" = 30, and "cutHeight" = 0.1. The eigengenes corresponding to each identified module were correlated individually to all the ^{1}H NMR and

525 biochemical assay metabolites, i.e., a set of 56 different molecules (see next paragraphs). A
526 module was then considered positively or negatively associated to this set of molecules if the
527 Pearson *r* correlation values were ≥ |0.65| for at least 5 molecules and if all the corresponding
528 *p*-values were ≤ 1e-05. The positively and the negatively correlated modules defined in this
529 way were merged to obtain a single gene list, which was subsequently compared to the DEG
530 list using a Venn diagram.
531 Afterwards, a literature-based meta-analytical enrichment test was carried out to assess
532 whether the DEG list and the gene list obtained using WGCNA were enriched in genes
533 related to mitochondria. To this aim, first a consensus list of genes related to these organelles
534 was created (Expanded View Information and Table EV4). This gene list was then annotated
535 in a descriptive way. First, gene symbols were converted into the corresponding KEGG
536 Orthology (KO) codes using the "db2db" tool (https://biodbnet-
537 abcc.ncifcrf.gov/db/db2db.php) from the bioDBnet suite and using the then up-to-date
538 underlying databases. Then, the retrieved KO codes were processed with the "Reconstruct
539 Pathway" tool (https://www.genome.jp/kegg/tool/map_pathway.html) from the KEGG
540 Mapper suite. Eventually, the obtained KEGG pathways underwent some manual editing step
541 to make them easier to interpret, namely (1) only first- and third-level hierarchies including at
542 least 15 genes were kept for visualization; (2) the "Human diseases" first-level hierarchy,
543 with all its child taxonomic terms, was removed; (3) the "Metabolic pathways" third-level
544 hierarchy was discarded as it was redundant with respect to the other third-level hierarchies
545 of the "Metabolism" term (Fig EV4).
546 Then, the same gene list was intersected with the genes found expressed on the microarray
547 (i.e., the "E1" matrix). The subset thus obtained was separately intersected with the DEG list
548 and the WGCNA gene modules. A Fisher's exact test and hypergeometric test were then used
549 to evaluate the overrepresentation of genes related to mitochondrial functions in each list.

The intersection between the genes related to mitochondria found on the microarray and the DEGs included 801 genes and was referred to as "mt-related genes". It was functionally annotated using ClueGO (version 2.5.7) (Bindea *et al*, 2009) by carrying out a right-sided test. Significance was set at a Benjamini-Hochberg adjusted *p*-value of 0.05 and the k-score was fixed at 0.4. Only the "KEGG 30.01.2019" ontology was selected.

Proton magnetic resonance (^{1}H NMR) metabolite analysis in plasma

As described by Plancade *et al*. (2019) and Le Moyec *et al.* (2019) to characterize the metabolic phenotype of endurance horses in detail, we measured ^{1}H NMR spectra at 600 MHz for plasma samples. Blood was collected from each horse the day before the event (T0) and within 30 minutes from the end of the endurance race (T1) using sodium fluoride and oxalate tubes in order to inhibit further glycolysis that may increase lactate levels after sampling. All the samples were immediately refrigerated at 4 ºC to minimize the metabolic activity of cells and enzymes and to keep metabolite composition as stable as possible, and clotting time was strictly controlled to avoid cell lysis. After clotting, plasma was separated from blood cells and subsequently transported to the laboratory at 4 ºC and then frozen at -80 °C (no more than 5 h later in all cases). Plasma samples were subsequently thawed at room temperature. Using 5 mm NMR tubes, 600 µL of plasma were added to 100 µL deuterium oxide for field locking. The ^{1}H NMR spectra were acquired at 500 MHz with an AVANCE III (Bruker, Billerica, MA, USA) equipped with a 5 mm reversed QXI Z-gradient high-resolution probe. Water signal was suppressed with a pre-saturation pulse (3.42×10^{-5} W) during a 3s-relaxation delay at the water resonance frequency. The spectrum was divided into 0.001 ppm regions (bins) over which the signals were integrated to obtain intensities. The high- and low-field ends were removed, leaving only the data between 9.5 to 0.0 ppm. The region between 4.5 and 5.0 ppm, which corresponded to the signal of residual water, was

also removed. The data were normalized according to the spectra using the probabilistic quotient method (Dieterle *et al*, 2006) and the bins, corresponding to the variables for the statistical analysis, were scaled to unit variance. Further details on sample preparation, data acquisition, data quality control, spectroscopic data pre-processing, and data pre-processing including bin alignment, scaling and normalization are broadly discussed elsewhere (Le Moyec *et al*, 2014).

As specified in Plancade *et al.* (2019), metabolite identification was then performed by using information acquired from other available biochemical databases, namely HMD (http://www.hmdb.ca/), KEGG (https://www.genome.jp/kegg/), METLIN (http://metlin.scripps.edu/), ChEBI (http://www.ebi.ac.uk/Databases/), and LIPID MAPS (http://www.lipidmaps.org/) and the literature (Le Moyec *et al*, 2014; Mach *et al*, 2017b; Le Moyec *et al*, 2019; Jang *et al*, 2017). Each peak was assigned to a metabolite when chemical shifts of peaks in the samples were the same as in the publicly available reference databases or literature (with a shift tolerance level of ± 0.005 ppm), in order to counteract the effects of measurements and pre-processing variability introduced by factors such as pH and solvents. A manual curation for identified compounds was carried out by an expert in horse metabolomics (Le Moyec *et al*, 2014). Eventually, the relative abundance of each metabolite was calculated as the area under the peak (Zheng *et al*, 2011). A total of 50 metabolites was identified, which belonged to the following broad categories: amino acids, including aromatic and branched-chain amino acids, energy metabolism-related metabolites, saccharides, and organic osmolytes (Table EV6). We refer to our previous work (Plancade *et al*, 2019) for more detailed descriptions of the pre-processing and main results of the plasma metabolome data that were used to generate the input files provided with this study.

Biochemical assay data production

Blood samples for biochemical assays were collected at T0 and T1 using 10 mL BD Vacutainer EDTA tubes (Becton Dickinson, Franklin Lakes, NJ, USA). As detailed in Plancade *et al*. (2019), after clotting the tubes were centrifuged and the harvested serum was stored at 4 °C until analysis (no more than 48 later, in all cases). Sera were assayed for total bilirubin, conjugated bilirubin, total protein, creatinine, CK, β-hydroxybutyrate, ASAT, γ-glutamyltransferase and serum amyloid A levels on a RX Imola analyzer (Randox, Crumlin, UK). The biochemical values obtained are reported in the Table EV7.

Fecal measurements, 16S data production and analysis

As described by Plancade *et al*. (2019), fresh fecal samples were obtained while monitoring the horses before the race (no more than 24 h before starting the race, in all cases). One fecal sample from each animal was collected off the ground immediately after defecation as described by Mach *et al*. (2017a) and Plancade *et al*. (2019), and three aliquots (200 mg) were prepared. Aliquots for pH determination were kept at room temperature, while aliquots for SCFA analysis and DNA extraction were snap-frozen. Since most of the horses experienced dehydration after the race, the gastrointestinal emptying was significantly delayed and consequently we were not able to recover the feces after the race.

Fecal pH was immediately determined after 10% fecal suspension (wt/vol) in saline solution (0.15 M NaCl solution). SCFAs concentrations were measured as previously described in Mach *et al*. (2017a) . The values obtained are described in the Table EV10.

Total DNA was extracted using the EZNA Stool DNA Kit (Omega Bio-Tek, Norcross, Georgia, USA), and following the manufacturer's instructions. DNA was then quantified using a Qubit and a dsDNA HS assay kit (Thermo Fisher).

The V3-V4 hyper-variable region of the 16S rRNA gene was amplified as previously reported by our team (Mach *et al*, 2017a; Plancade *et al*, 2019; Clark *et al*, 2018; Mach *et al*,

2020; Massacci *et al*, 2019). The concentration of the purified amplicons was measured using a Nanodrop 8000 spectrophotometer (Thermo Fisher) and their quality was checked using DNA 7500 chips onto a Bioanalyzer 2100 (Agilent Technologies). All libraries were pooled at equimolar concentration, and the final pool had a diluted concentration of 5 nM and was used for sequencing. The pooled libraries were mixed with 15% PhiX control according to the protocol provided by Illumina (Illumina, San Diego, CA, USA) and sequenced on a single MiSeq (Illumina, USA) run using a MiSeq Reagent Kit v2 (500 cycles).

The Divisive Amplicon Denoising Algorithm (DADA) was implemented using the DADA2 plug-in for QIIME 2 (version 2019.10) to perform quality filtering and chimera removal and to construct a feature table consisting of read abundance per amplicon sequence variant (ASV) by sample (Callahan *et al*, 2016). DADA2 models the amplicon sequencing error in order to identify unique ASV and infers sample composition more accurately than traditional Operational Taxonomic Unit (OTU) picking methods that identify representative sequences from clusters of sequences based on a % similarity cut-off (Callahan *et al*, 2016). The output of DADA2 was an abundance table, in which each unique sequence was characterized by its abundance in each sample. Taxonomic assignments were given to ASVs by importing SILVA 16S representative sequences and consensus taxonomy (release 132, 99% of identity) to QIIME 2 and classifying representative ASVs using the naive Bayes classifier plug-in (Bokulich *et al*, 2018). The feature table, taxonomy, and phylogenetic tree were then exported from QIIME 2 to the R statistical environment and combined into a phyloseq object (McMurdie & Holmes, 2013). Prevalence filtering was applied to remove ASVs with less than 1% prevalence and in fewer than three individuals, decreasing the possibility of data artifacts affecting the analysis (Callahan *et al*, 2016). To reduce the effects of uncertainty in ASV taxonomic classification, we conducted most of our analysis at the microbial genus level.

650 The phyloseq (version 1.32.0) (Mcmurdie & Holmes, 2012), vegan (version 2.5.6) (Dixon,
651 2003) and microbiome packages (version 1.10.0) were used in R (version 4.0.2) for the
652 downstream steps of analysis. The minimum sampling depth in our data set was 10,423 reads
653 per sample. Reads were clustered into 3,385 chimera- and singleton-filtered Amplicon
654 Sequence variants (ASVs) at 99% sequence similarity. ASV counts per sample and ASV
655 taxonomical assignments are available in Table EV8. Data were aggregated at genus, family,
656 order, class and phyla levels throughout the taxonomic-agglomeration method in the phyloseq
657 R package, which merges taxa of the same taxonomic category for a user-specific taxonomic
658 level. The final table obtained after these steps was called "G1" and included 100 genera
659 (Table EV9).
660
661 **qPCR quantification of bacterial, fungal and protozoan concentration**
662 As detailed by Plancade *et al*. (2019), concentrations of bacteria, anaerobic fungi and
663 protozoa in fecal samples were quantified by qPCR using a QuantStudio 12K Flex platform
664 (Thermo Fisher Scientific, Waltham, USA). Primers for real-time amplification of bacteria
665 (FOR: 5'-CAGCMGCCGCGGTAANWC-3'; REV: 5'-CCGTCAATTCMTTTRAGTTT-3'),
666 anaerobic fungi (FOR: 5'-TCCTACCCTTTGTGAATTTG-3'; REV: 5'-
667 CTGCGTTCTTCATCGTTGCG-3') and protozoa (FOR: 5'-
668 GCTTTCGWTGGTAGTGTATT-3'; REV: 5'-CTTGCCCTCYAATCGTWCT-3'), are
669 described in Mach *et al*. (2015) and Clark *et al*. (2018) and were purchased from Eurofins
670 Genomics (Ebersberg, Germany).
671 Amplified fragments of the target amplicons were used to create a seven-point 10-fold
672 standard dilution series. The dilution points ranged from 2.25×10^7 to 2.25×10^{13} copies per
673 µg of DNA for bacteria and protozoa and from 3.70×10^6 to 3.70×10^{12} copies per µg of
674 DNA for anaerobic fungi. qPCR reactions were performed in a final volume of 20 µL,

containing 10 µL of Power SYBR Green PCR Master Mix (Thermo Fisher), 2 µL of standard or DNA template at 0.5 ng/µL and 0.6 µM of each primer to a final concentration of 200 mM for bacteria and anaerobic fungi and 150 mM for protozoa.

In all the cases, the thermal cycling conditions were as follows: initial denaturation at 95 °C for 10 min; 40 cycles of denaturation at 95 °C for 15 sec, annealing and extension at 60 °C for 60 sec. To check for the absence of nonspecific signals, a dissociation step was added after each amplification. It was carried out by ramping the temperature from 60 ºC to 95 °C. All qPCR runs were performed in triplicate, and the standard curve obtained using the target amplicons was used to calculate the number of copies of microorganisms in feces.

Taking into account the molecular mass of nucleotides and the amplicon length, the number of copies was obtained using the following equation: copies per nanogram = $(NL \times A \times 10^{-9})/(n \times mw)$, where "NL" is the Avogadro constant (6.02×10^{23} molecules per mole), "A" is the molecular weight of DNA molecules (ng), "n" is the length of the amplicon in base pairs, and "mw" is the molecular weight per base pair. The final values obtained are described in the Table EV11.

Integrative statistical analyses

Data integration was carried out using several approaches and different combinations of data sets. Prior to integration, each data set underwent a specific set of pre-processing steps.

In the case of the ^{1}H NMR and biochemical assays, data were processed by subtracting the T0 values from the T1 values. The two data sets included 50, and 9 molecules, respectively.

In the case of the blood transcriptome, a new expression matrix ("E2") was created including only the differentially expressed mt-related genes and by subtracting the T0 from the T1 expression matrix values, i.e., by calculating the ratio between T1 and T0 log scaled expression values from the two matrices. A total of 801 genes was retained (Table EV5).

The fecal pH values and the proportions of the 6 SCFAs in feces did not undergo any specific pre-processing. In the case of fecal microbiota, the genera table "G1" was modified using the mixMC framework of the mixOmics R package (version 6.10.9) (Rohart *et al*, 2017). First, raw data were pre-filtered by removing the genera for which the percentage of the sum of counts was lower than 1% compared to the total sum of all counts. Then, the pre-filtered data were transformed using the Centered Log Ratio transformation (CLR) and applying an offset of 1. The filtered genera matrix "G2" obtained in this way included 85 genera (Table EV14). Finally, the concentration of microorganisms in feces did not undergo any specific editing.

After the pre-processing steps, a first round of integration was performed using three different methods and working with all six data sets available, namely: (1) mt-related genes (i.e., the "E2" matrix); (2) ^{1}H NMR metabolites; (3) biochemical assay metabolites; (4) the concentrations of fecal SCFAs; (5) fecal 16S rRNA gene sequencing data (i.e., the "G2" matrix); and (6) the concentration of fecal microorganisms.

As a first integration approach, a global NMDS ordination was used to extract and summarize the variation in mt-related genes (the "response variable") using the "metaMDS" function in vegan R package. To determine the number of dimensions for each NMDS, stress values were calculated. Stress values are a measure of how much the distances in the reduced ordination space depart from the distances in the original p-dimensional space. High stress values indicate a greater possibility that the structuring of observations in the ordination space is entirely unrelated to that of the original full-dimensional space.

The other five data sets (the "explanatory variables") were then fitted to the ordination plots using the "envfit" function in the vegan R package (Clarke & Ainsworth, 1993) with 10,000 permutations. The "envfit" function performs a multivariate analysis of variance (MANOVA) and linear correlations for categorical and continuous variables. The effect size and significance of each covariate were determined comparing the difference in the centroids of

each group relative to the total variation, and all of the *p*-values derived from the "envfit" function were Benjamini-Hochberg adjusted. The obtained r^2 gives the proportion of variability (that is, the main dimensions of the ordination) that can be attributed to the explanatory variables.

As a second integration approach, a forward-selection model-building method for redundancy analysis (RDA) (Blanchet *et al*, 2008) was used to extract and summarize the variation in mt-related genes (the "response variable") that could be explained by the other five data sets (the "explanatory variables"). To determine which set of covariates provided the most parsimonious model, automatic stepwise model selection for constrained ordination methods was used as implemented by the "ordistep" function of the vegan R package. To test for robustness, a forward automatic model selection on a distance based RDA was then performed using the "ordiR2step" function of the vegan package in R (Oksanen *et al*, 2013). This provided an estimation of the linear cumulative effect size of all the identified non-redundant covariates and of their independent fraction in the best model. In the case of this latter function, the "E2" matrix was modified using the Hellinger transformation prior to the analysis. These two RDA functions use different criteria for variable selection. The "ordistep" funciton uses the Akaike's information criterion (AIC) and p-value < 0.05 obtained from Monte Carlo permutation tests, while "ordiR2step" uses the adjusted coefficient of determination (r^2_{adj}). In both cases, the procedure begins by comparing a null model containing no variables and a test model containing one variable, where every possible covariate is considered.

As a third integrative approach, the N-integration algorithm DIABLO of the mixOmics R package was used. In this case, the relationships existing among all six data sets were studied by adding a further categorical variable, i.e., the performance of horses. Horses that had a poor performance or that had been eliminated (n = 8) were compared to horses that had

completed the race (n = 12, Table EV1). DIABLO seeks to estimate latent components by modelling and maximizing the correlation between pairs of pre-specified datasets to unravel similar functional relationships between them (Singh *et al*, 2019). A full weighted design was considered and, to predict the number of latent components and the number of discriminants, the "block.splsda" function was used. In both cases, the model was first fine-tuned using the leave-one-out cross-validation by splitting the data into training and testing. Then, classification error rates were calculated using balanced error rates (BERs) between the predicted latent variables with the centroid of the class labels (i.e., eliminated vs non eliminated horses) using the "max.dist" function. BERs account for differences in the number of samples between different categories. Only interactions with $r \geq |0.70|$ were visualized using CIRCOS. To visualize the high-confidence molecule co-associations determined by CIRCOS, only those with $r \geq |0.70|$ and more than 15 connections were automatically visualized using the organic layout algorithm in Cytoscape (version 3.8.1) (http://cytoscape.org).

Finally, we performed a pairwise integration, focusing only on mt-related genes and microbiota data, and using the rCCA method as implemented by the mixOmics R package (https://cran.r-project.org/web/packages/mixOmics/). The penalization parameters were estimated using the "shrinkage" method and setting the "ncomp" parameter to two. The correlation matrix thus obtained was filtered by retaining only the genes and the genera for which at least one association value data point presented $r \geq |0.55|$.

ACKNOWLEDGMENTS

We are grateful to Marine Beinat, Julie Rivière, Emmanuelle Rebours, Jordi Estellé, Caroline Morgenthaler and Arni Janssen for participating in the sample collection and organization during the project. We particularly thank Valentine Ballan and Fanny Blanc who helped to

isolate the blood RNA. We also thank Catherine Philippe for the SCFA analysis in feces, Diane Esquerré who prepared the libraries and performed the MiSeq sequencing at the GeT-PlaGe genomics core facility (INRAE, Toulouse, https://get.genotoul.fr/) and Estel Blasi, who created the picture of horses. Lastly, we are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020, Migale bioinformatics facility, doi: 10.15454/1.5572390655343293E12) for providing computing and storage resources and to all the horse owners, riders and endurance race organizers who participated in the study.

This work was funded by grants from the *Fonds Eperon*, the *Institut Français du Cheval et de l'Equitation* (IFCE), the *Association du Cheval Arabe* (ACA) and the *Institute National de la Recherche pour l'Agriculture, l'alimentation et l'environnement* (INRAE).

AUTHOR CONTRIBUTIONS

NM and EB conceived of the presented idea. NM designed the experiment, performed the PCR and the RT-qPCR experiments, carried out most of the bioinformatics analysis and the integrative biology approaches. NM wrote the main manuscript text and prepared the figures with support from MM and EB. MM carried out the transcriptome analyses, the rCCA integrative analysis and prepared most of the tables. MM and EB created the consensus list of genes related to mitochondria. AR verified the statistical methods and CR verified the blood biochemical and the performance data. JL performed all the laboratory steps related to microarray analyses. LLM performed the metabolomic experiment and analyzed the metabolite peaks. CR and EB were in charge of the organization and sampling management during the race. All authors reviewed the manuscript and approved the final version.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY SECTION

The datasets produced in this study are available in the following databases:

- Microarray expression data: Gene Expression Omnibus (GEO) repository under the accession number GSE163767; (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163767)
- Metabolomic data: NIH Common Fund's Data Repository and Coordinating Center UrqK1489; (http://dev.metabolomicsworkbench.org:22222/data/DRCCMetadata.php?Mode=Study&StudyID=ST000945)
- Gut metagenome 16S rRNA targeted locus data: DDBJ/EMBL/GenBank under the accession KBTQ00000000, version KBTQ00000000.1; (locus KBTQ01000000). The corresponding BioProject is PRJNA438436, and the accession numbers of the BioSamples included in it were SAMN08715709 to SAMN08715760.

REFERENCES

Al Amir Dache Z, Otandault A, Tanos R, Pastor B, Meddeb R, Sanchez C, Arena G, Lasorsa L, Bennett A, Grange T, *et al* (2020) Blood contains circulating cell-free respiratory competent mitochondria. *FASEB J* 34: 3616–3630

Bajpai P, Darra A & Agrawal A (2018) Microbe-mitochondrion crosstalk and health: An emerging paradigm. *Mitochondrion* 39: 20–25

Barrey E, Mucher E, Robert C, Amiot F & Gidrol X (2006) Gene expression profiling in blood cells of endurance horses completing competition or disqualified due to metabolic disorder. *Equine Vet J* 38: 43–49

Barton W, Penney NC, Cronin O, Garcia-Perez I, Molloy MG, Holmes E, Shanahan F, Cotter

825 PD & O'Sullivan O (2017) The microbiome of professional athletes differs from that of
826 more sedentary subjects in composition and particularly at the functional metabolic
827 level. *Gut*: gutjnl-2016-313627

828 Bianchessi V, Vinci MC, Nigro P, Rizzi V, Farina F, Capogrossi MC, Pompilio G, Gualdi V
829 & Lauri A (2016) Methylation profiling by bisulfite sequencing analysis of the mtDNA
830 Non-Coding Region in replicative and senescent Endothelial Cells. *Mitochondrion* 27:
831 40–47

832 Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH,
833 Pagès F, Trajanoski Z & Galon J (2009) ClueGO: A Cytoscape plug-in to decipher
834 functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*
835 25: 1091–1093

836 Blanchet FG, Legendre P & Borcard D (2008) Forward selection of explanatory variables.
837 *Ecology* 89: 2623–2632

838 Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA &
839 Gregory Caporaso J (2018) Optimizing taxonomic classification of marker-gene
840 amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6: 90

841 Callahan BJ, McMurdie PJ & Holmes SP (2017) Exact sequence variants should replace
842 operational taxonomic units in marker-gene data analysis. *ISME J* 11: 2639–2643

843 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA & Holmes SP (2016)
844 DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*
845 13: 581

846 Calvo SE, Clauser KR & Mootha VK (2016) MitoCarta2.0: An updated inventory of
847 mammalian mitochondrial proteins. *Nucleic Acids Res* 44: D1251–D1257

848 Capomaccio S, Vitulo N, Verini-Supplizi A, Barcaccia G, Albiero A, D'Angelo M,
849 Campagna D, Valle G, Felicetti M, Silvestrelli M, *et al* (2013) RNA sequencing of the

850 exercise transcriptome in equine athletes. *PLoS One* 8: e83504
851 Cardoso SM & Empadinhas N (2018) The microbiome-mitochondria dance in prodromal
852 Parkinson's disease. *Front Physiol* 9: 471
853 Chinnery PF & Hudson G (2013) Mitochondrial genetics. *Br Med Bull* 106: 135–139
854 Clark A & Mach N (2017) The Crosstalk between the Gut Microbiota and Mitochondria
855 during Exercise. *Front Physiol* 8: 319
856 Clark A, Sallé G, Ballan V, Reigner F, Meynadier A, Cortet J, Koch C, Riou M, Blanchard A
857 & Mach N (2018) Strongyle Infection and Gut Microbiota: Profiling of Resistant and
858 Susceptible Horses Over a Grazing Season. *Front Physiol* 9: 1–18
859 Clarke KR & Ainsworth M (1993) A method of linking multivariate community structure to
860 environmental variables. *Mar Ecol Prog Ser* 92: 205–219
861 Cosson P, Marchetti A, Ravazzola M & Orci L (2012) Mitofusin-2 Independent Juxtaposition
862 of Endoplasmic Reticulum and Mitochondria: An Ultrastructural Study. *PLoS One* 7:
863 e46293
864 Costa MC & Weese JS (2018) Understanding the Intestinal Microbiome in Health and
865 Disease. *Vet Clin North Am - Equine Pract* 34: 1–12
866 Cottin F, Metayer N, Goachet AG, Julliand V, Slawinski J, Billat V & Barrey E (2010)
867 Oxygen consumption and gait variables of Arabian endurance horses measured during a
868 field exercise test. *Equine Vet J* 42: 1–5
869 Crisp A, Boschetti C, Perry M, Tunnacliffe A & Micklem G (2015) Expression of multiple
870 horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes.
871 *Genome Biol* 16: 50
872 Dieterle F, Ross A, Schlotterbeck G & Senn H (2006) Probabilistic Quotient Normalization
873 as Robust Method to Account for Dilution of Complex Biological Mixtures. Application
874 in 1H NMR Metabonomics. *Anal Chem* 78: 4281–4290

875 Dixon P (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* 14: 927

876 Dobashi Y, Miyakawa Y, Yamamoto I & Amao H (2011) Effects of intestinal microflora on
877 superoxide dismutase activity in the mouse cecum. *Exp Anim* 60: 133–139

878 Duncan SH, Louis P & Flint HJ (2004) Lactate-utilizing bacteria, isolated from human feces,
879 that produce butyrate as a major fermentation product. *Appl Environ Microbiol* 70:
880 5810–5817

881 Esser C, Ahmadinejad N, Wiegand C, Rotte C, Sebastiani F, Gelius-Dietrich G, Henze K,
882 Kretschmann E, Richly E, Leister D, *et al* (2004) A genome phylogeny for mitochondria
883 among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes.
884 *Mol Biol Evol* 21: 1643–1660

885 Fernandes J, Su W, Rahat-Rozenbloom S, Wolever TMS & Comelli EM (2014) Adiposity,
886 gut microbiota and faecal short chain fatty acids are linked in adult humans. *Nutr*
887 *Diabetes* 4: e121

888 Fielding CL & Dechant JE (2012) Colic in competing endurance horses presenting to referral
889 centres: 36 cases. *Equine Vet J* 44: 472–475

890 Franco-Obregón A & Gilbert JA (2017) The Microbiome-Mitochondrion Connection:
891 Common Ancestries, Common Mechanisms, Common Goals. *mSystems* 2: e00018-17

892 Gao Z, Yin J, Zhang J, Ward RE, Martin RJ, Lefevre M, Cefalu WT & Ye J (2009) Butyrate
893 improves insulin sensitivity and increases energy expenditure in mice. *Diabetes* 58:
894 1509–1517

895 Gérard P (2013) Metabolism of cholesterol and bile acids by the gut microbiota. *Pathogens* 3:
896 14–24

897 Gharechahi J, Farhad M, Mohammad V & Han J (2020) Metagenomic analysis reveals a
898 dynamic microbiome with diversified adaptive functions to utilize high lignocellulosic
899 forages in the cattle rumen. *ISME J (accepted)*

900 Goachet AG & Julliand V (2015) Implementation of field cardio-respiratory measurements to
901 assess energy expenditure in Arabian endurance horses. *Animal* 9: 787–792
902 Gruber J & Kennedy BK (2017) Microbiome and Longevity: Gut Microbes Send Signals to
903 Host Mitochondria. *Cell* 169: 1168–1169
904 Gunn HM (1987) Muscle, bone and fat proportions and muscle distribution of thoroughbreds
905 and other horses. In *International Conference on Equine Exercise Physiology* pp 253–
906 264. J.R. Gillespie and N.E. Robinson Eds., ICEEP Publications, Davis, California
907 Gustafsson CM, Falkenberg M & Larsson NG (2016) Maintenance and Expression of
908 Mammalian Mitochondrial DNA. *Annu Rev Biochem* 85: 133–160
909 Gustafsson CM & Samuelsson T (2001) Mediator - A universal complex in transcriptional
910 regulation. *Mol Microbiol* 41: 1–8
911 Han B, Sivaramakrishnan P, Lin CCJ, Neve IAA, He J, Tay LWR, Sowa JN, Sizovs A, Du G,
912 Wang J, *et al* (2017) Microbial Genetic Composition Tunes Host Longevity. *Cell* 169:
913 1249-1262.e13
914 Hawley JA (2020) Microbiota and muscle highway — two way traffic. *Nat Rev Endocrinol*
915 16: 71–72
916 Hawley JA, Lundby C, Cotter JD & Burke LM (2018) Maximizing Cellular Adaptation to
917 Endurance Exercise in Skeletal Muscle. *Cell Metab* 27: 962–976
918 Henagan TM, Stefanska B, Fang Z, Navard AM, Ye J, Lenard NR & Devarshi PP (2015)
919 Sodium butyrate epigenetically modulates high-fat diet-induced skeletal muscle
920 mitochondrial adaptation, obesity and insulin resistance through nucleosome
921 positioning. *Br J Pharmacol* 172: 2782–2798 (2015).
922 Hood DA, Uguccioni G, Vainshtein A & D'souza D (2011) Mechanisms of exercise-induced
923 mitochondrial biogenesis in skeletal muscle: Implications for health and disease. *Compr*
924 *Physiol* 1: 1119–1134

925 Jackson DN & Theiss AL (2020) Gut bacteria signaling to mitochondria in intestinal
926 inflammation and cancer. *Gut Microbes* 11: 285–304
927 Jang HJ, Kim DM, Kim KB, Park JW, Choi JY, Hyeog Oh J, Song KD, Kim S & Cho BW
928 (2017) Analysis of metabolomic patterns in thoroughbreds before and after exercise.
929 *Asian-Australasian J Anim Sci* 30: 1633–1642
930 Janouškovec J, Tikhonenkov D V., Burki F, Howe AT, Rohwer FL, Mylnikov AP & Keeling
931 PJ (2017) A New Lineage of Eukaryotes Illuminates Early Mitochondrial Genome
932 Reduction. *Curr Biol* 27: 3717–3724
933 Jones RM & Neish AS (2014) Redox signaling mediates symbiosis between the gut
934 microbiota and the intestine. *Gut Microbes* 5: 250–253
935 Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, Nielsen J &
936 Bäckhed F (2013) Gut metagenome in European women with normal, impaired and
937 diabetic glucose control. *Nature* 498: 99–103
938 Kauffmann A, Gentleman R & Huber W (2009) arrayQualityMetrics - A bioconductor
939 package for quality assessment of microarray data. *Bioinformatics* 25: 415–416
940 Kauter A, Epping L, Semmler T, Antao E-M, Kannapin D, Stoeckle SD, Gehlen H, Lübke-
941 Becker A, Günther S, Wieler LH, *et al* (2019) The gut microbiome of horses: current
942 research on equine enteral microbiota and future perspectives. *Anim Microbiome* 1: 1–15
943 Keohane DM, Woods T, O'Connor P, Underwood S, Cronin O, Whiston R, O'Sullivan O,
944 Cotter P, Shanahan F & Molloy MGM (2019) Four men in a boat: Ultra-endurance
945 exercise alters the gut microbiome. *J Sci Med Sport* 22: 1059–1064
946 van der Kolk JH, Thomas S, Mach N, Ramseyer A, Burger D, Gerber V & Nuoffer JM
947 (2020) Serum acylcarnitine profile in endurance horses with and without metabolic
948 dysfunction. *Vet J* 255: 105419
949 Langfelder P & Horvath S (2008) WGCNA: An R package for weighted correlation network

950 analysis. *BMC Bioinformatics* 9: 559

951 Lê S, Josse J & Husson F (2008) FactoMineR: An R Package for Multivariate Analysis. *J*
952 *Stat Softw* 25: 1–18

953 Lee C, Zeng J, Drew BG, Sallam T, Martin-Montalvo A, Wan J, Kim SJ, Mehta H, Hevener
954 AL, De Cabo R, *et al* (2015) The mitochondrial-derived peptide MOTS-c promotes
955 metabolic homeostasis and reduces obesity and insulin resistance. *Cell Metab* 21: 443–
956 454

957 Mach N, Berri M, Estellé J, Levenez F, Lemonnier G, Denis C, Leplat JJ, Chevaleyre C,
958 Billon Y, Dore J, *et al* (2015) Early-life establishment of the swine gut microbiome and
959 impact on host phenotypes. *Environ Microbiol Rep* 7: 554–569

960 Mach N, Foury A, Kittelmann S, Reigner F, Moroldo M, Ballester M, Esquerré D, Rivière J,
961 Sallé G, Gérard P, *et al* (2017a) The effects of weaning methods on gut microbiota
962 composition and horse physiology. *Front Physiol* 8: 535

963 Mach N & Fuster-Botella D (2017) Endurance exercise and gut microbiota: A review. *J Sport*
964 *Heal Sci* 6: 179–197

965 Mach N, Plancade S, Pacholewska A, Lecardonnel J, Rivière J, Moroldo M, Vaiman A,
966 Morgenthaler C, Beinat M, Nevot A, *et al* (2016) Integrated mRNA and miRNA
967 expression profiling in blood reveals candidate biomarkers associated with endurance
968 exercise in the horse. *Sci Rep* 6: 22932

969 Mach N, Ramayo-Caldas Y, Clark A, Moroldo M, Robert C, Barrey E, López JM & Le
970 Moyec L (2017b) Understanding the response to endurance exercise using a systems
971 biology approach: combining blood metabolomics, transcriptomics and miRNomics in
972 horses. *BMC Genomics* 18: 187

973 Mach N, Ruet A, Clark A, Bars-Cortina D, Ramayo-Caldas Y, Crisci E, Pennarun S, Dhorne-
974 Pollet S, Foury A, Moisan M-P, *et al* (2020) Priming for welfare: gut microbiota is

975 associated with equitation conditions and behavior in horse athletes. *Sci Rep* 10: 8311
976 Maier UG, Zauner S, Woehle C, Bolte K, Hempel F, Allen JF & Martin WF (2013)
977 Massively convergent evolution for ribosomal protein gene content in plastid and
978 mitochondrial genomes. *Genome Biol Evol* 12: 2318–2329
979 Massacci FR, Clark A, Ruet A, Lansade L, Costa M & Mach N (2019) Inter-breed diversity
980 and temporal dynamics of the faecal microbiota in healthy horses. *J Anim Breed Genet*
981 00: 1–18
982 Mcmurdie PJ & Holmes S (2012) Phyolseq: A Bioconductor package for handling and
983 Analysis of High-Throughput Phylogenetic Sequence Data. *Pac Symp Biocomput* 0:
984 235–246
985 McMurdie PJ & Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive
986 Analysis and Graphics of Microbiome Census Data. *PLoS One* 8: e61217
987 Mohr S & Liew CC (2007) The peripheral-blood transcriptome: new insights into disease and
988 risk assessment. *Trends Mol Med* 13: 422–432
989 Mollica MP, Raso GM, Cavaliere G, Trinchese G, De Filippo C, Aceto S, Prisco M, Pirozzi
990 C, Di Guida F, Lama A, *et al* (2017) Butyrate regulates liver mitochondrial function,
991 efficiency, and dynamics in insulin-resistant obese mice. *Diabetes* 66: 1405–1418
992 Mottawea W, Chiang CK, Mühlbauer M, Starr AE, Butcher J, Abujamel T, Deeke SA,
993 Brandel A, Zhou H, Shokralla S, *et al* (2016) Altered intestinal microbiota-host
994 mitochondria crosstalk in new onset Crohn's disease. *Nat Commun* 7: 13419
995 Le Moyec L, Robert C, Triba MN, Billat VL, Mata X, Schibler L & Barrey E (2014) Protein
996 catabolism and high lipid metabolism associated with long-distance exercise are
997 revealed by plasma NMR metabolomics in endurance horses. *PLoS One* 9: 1–10
998 Le Moyec L, Robert C, Triba MN, Bouchemal N, Mach N, Rivière J, Zalachas-Rebours E &
999 Barrey E (2019) A first step toward unraveling the energy metabolism in endurance

1000 horses: Comparison of plasma nuclear magnetic resonance metabolomic profiles before
1001 and after different endurance race distances. *Front Mol Biosci* 6: 45
1002 Nay K, Jollet M, Goustard B, Baati N, Vernus B, Pontones M, Lefeuvre-Orfila L, Bendavid
1003 C, Rué O, Mariadassou M, *et al* (2019) Gut bacteria are critical for optimal muscle
1004 function: A potential link with glucose homeostasis. *Am J Physiol - Endocrinol Metab*
1005 317: E158-171
1006 Nicholls TJ & Gustafsson CM (2018) Separating and Segregating the Human Mitochondrial
1007 Genome. *Trends Biochem Sci* 43: 869–881.
1008 Okamoto T, Morino K, Ugi S, Nakagawa F, Lemecha M, Ida S, Ohashi N, Sato D, Fujita Y
1009 & Maegawa H (2019) Microbiome potentiates endurance exercise through intestinal
1010 acetate production. *Am J Physiol - Endocrinol Metab* 316: E956–E966
1011 Oksanen J, Blanchet FG, Kindt R, Minchin PR, O'Hara RB, Simpson GL, Solymos P,
1012 Stevens MHH & Wagner HH (2013) Package 'vegan': a community ecology package.
1013 260
1014 Pagliarini DJ, Calvo SE, Chang B, Sheth SA, Vafai SB, Ong SE, Walford GA, Sugiana C,
1015 Boneh A, Chen WK, *et al* (2008) A Mitochondrial Protein Compendium Elucidates
1016 Complex I Disease Biology. *Cell* 134: 112–123
1017 Pearce SF, Rebelo-Guiomar P, D'Souza AR, Powell CA, Van Haute L & Minczuk M (2017)
1018 Regulation of Mammalian Mitochondrial Gene Expression: Recent Advances. *Trends*
1019 *Biochem Sci* 42: 625–639
1020 Pfanner N, Warscheid B & Wiedemann N (2019) Mitochondrial proteins: from biogenesis to
1021 functional networks. *Nat Rev Mol Cell Biol* 20: 267–284
1022 Plancade S, Clark A, Philippe C, Helbling J-C, Moisan M-P, Esquerré D, Le Moyec L,
1023 Robert C, Barrey E & Mach N (2019) Unraveling the effects of the gut microbiota
1024 composition and function on horse endurance physiology. *Sci Rep* 9: 9620

1025 Qi B & Han M (2018) Microbial Siderophore Enterobactin Promotes Mitochondrial Iron
1026 Uptake and Development of the Host via Interaction with ATP Synthase. *Cell* 175: 571-
1027 582.e11

1028 Rapoport BI (2010) Metabolic factors limiting performance in marathon runners. *PLoS*
1029 *Comput Biol* 6: e1000960

1030 Richter-Dennerlein R, Oeljeklaus S, Lorenzi I, Ronsör C, Bareth B, Schendzielorz AB, Wang
1031 C, Warscheid B, Rehling P & Dennerlein S (2016) Mitochondrial Protein Synthesis
1032 Adapts to Influx of Nuclear-Encoded Protein. *Cell* 671: 471–483

1033 Rizzuto R, De Stefani D, Raffaello A & Mammucari C (2012) Mitochondria as sensors and
1034 regulators of calcium signalling. *Nat Rev Mol Cell Biol* 13: 566–578

1035 Rodríguez-Carrio J, Salazar N, Margolles A, González S, Gueimonde M, de los Reyes-
1036 Gavilán CG & Suárez A (2017) Free fatty acids profiles are related to gut microbiota
1037 signatures and short-chain fatty acids. *Front Immunol* 8: 823

1038 Rohart F, Gautier B, Singh A & Lê Cao KA (2017) mixOmics: An R package for 'omics
1039 feature selection and multiple data integration. *PLoS Comput Biol* 13: e1005752

1040 Ropka-Molik K, Stefaniuk-Szmukier M, Zukowski K, Piórkowska K, Gurgul A & Bugno-
1041 Poniewierska M (2017) Transcriptome profiling of Arabian horse blood during training
1042 regimens. *BMC Genet* 18: 31

1043 Ruiz E, Penrose HM, Heller S, Nakhoul H, Baddoo M, Flemington EF, Kandil E & Savkovic
1044 SD (2020) Bacterial TLR4 and NOD2 signaling linked to reduced mitochondrial energy
1045 function in active inflammatory bowel disease. *Gut Microbes* 11: 350–363

1046 Saint-Georges-Chaumet Y, Attaf D, Pelletier E & Edeas M (2015) Targeting microbiota-
1047 mitochondria inter-talk: Microbiota control mitochondria metabolism. *Cell Mol Biol* 61:
1048 121–124

1049 Saint-Georges-Chaumet Y & Edeas M (2018) Microbiota–mitochondria inter-talk:

1050 Consequence for microbiota–host interaction. *Pathog Dis* 74: 1–5

1051 Scheiman J, Luber JM, Chavkin TA, MacDonald T, Tung A, Pham LD, Wibowo MC, Wurth
1052 RC, Punthambaker S, Tierney BT, *et al* (2019) Meta-omics analysis of elite athletes
1053 identifies a performance-enhancing microbe that functions via lactate metabolism. *Nat*
1054 *Med* 25: 1104–1109

1055 Scott JM, Esch BTA, Shave R, Warburton DER, Gaze D & George K (2009) Cardiovascular
1056 consequences of completing a 160-km ultramarathon. *Med Sci Sports Exerc* 41: 25–33

1057 Seok S, Fu T, Choi SE, Li Y, Zhu R, Kumar S, Sun X, Yoon G, Kang Y, Zhong W, *et al*
1058 (2014) Transcriptional regulation of autophagy by an FXR-CREB axis. *Nature* 516:
1059 108–111

1060 Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ & Cao KAL (2019)
1061 DIABLO: An integrative approach for identifying key molecular drivers from multi-
1062 omics assays. *Bioinformatics* 35: 3055–3062

1063 Smith AC & Robinson AJ (2019) Mitominer v4.0: An updated database of mitochondrial
1064 localization evidence, phenotypes and diseases. *Nucleic Acids Res* 47: D1225–D1228

1065 Smyth GK (2004) Linear models and empirical bayes methods for assessing differential
1066 expression in microarray experiments. *Stat Appl Genet Mol Biol* 3: Article3

1067 Smyth GK (2005) limma: Linear Models for Microarray Data. In *Bioinformatics and*
1068 *Computational Biology Solutions Using R and Bioconductor* pp 397–420.

1069 Smyth GK, Michaud J & Scott HS (2005) Use of within-array replicate spots for assessing
1070 differential expression in microarray experiments. *Bioinformatics* 21: 2067–2075

1071 Song X, Hu W, Yu H, Wang H, Zhao Y, Korngold R & Zhao Y (2020) Existence of
1072 circulating mitochondria in human and animal peripheral blood. *Int J Mol Sci* 21: 2122

1073 Vacca M, Celano G, Calabrese FM, Portincasa P, Gobbetti M & De Angelis M (2020) The
1074 controversial role of human gut lachnospiraceae. *Microorganisms* 8: 1–25

1075 Vezza T, Abad-Jiménez Z, Marti-Cabrera M, Rocha M & Víctor VM (2020) Microbiota-
1076 mitochondria inter-talk: A potential therapeutic strategy in obesity and type 2 diabetes.
1077 *Antioxidants* 9: 1–21
1078 Vital M, Gao J, Rizzo M, Harrison T & Tiedje JM (2015) Diet is a major factor governing the
1079 fecal butyrate-producing community structure across Mammalia, Aves and Reptilia.
1080 *ISME J* 9: 832–843
1081 Votion DM, Gnaiger E, Lemieux H, Mouithys-Mickalad A & Serteyn D (2012) Physical
1082 fitness and mitochondrial respiratory capacity in horse skeletal muscle. *PLoS One* 7:
1083 e34890
1084 Wang F, Wang K, Xu W, Zhao S, Ye D, Wang Y, Xu Y, Zhou L, Chu Y, Zhang C, *et al*
1085 (2017) SIRT5 Desuccinylates and Activates Pyruvate Kinase M2 to Block Macrophage
1086 IL-1β Production and to Prevent DSS-Induced Colitis in Mice. *Cell Rep* 19: 2331–2344
1087 Wang S, Jacquemyn J, Murru S, Martinelli P, Barth E, Langer T, Niessen CM & Rugarli EI
1088 (2016) The Mitochondrial m-AAA Protease Prevents Demyelination and Hair Greying.
1089 *PLoS Genet* 12: e1006463
1090 Wong ML, Inserra A, Lewis MD, Mastronardi CA, Leong L, Choo J, Kentish S, Xie P,
1091 Morrison M, Wesselingh SL, *et al* (2016) Inflammasome signaling affects anxiety- and
1092 depressive-like behavior and gut microbiome composition. *Mol Psychiatry* 21: 797–805
1093 Wu Z, Huang X, Feng Y, Handschin C, Feng Y, Gullicksen PS, Bare O, Labow M,
1094 Spiegelman B & Stevenson SC (2006) Transducer of regulated CREB-binding proteins
1095 (TORCs) induce PGC-1α transcription and mitochondrial biogenesis in muscle cells.
1096 *Proc Natl Acad Sci U S A* 39: 14379–14384
1097 Yardeni T, Tanes CE, Bittinger K, Mattei LM, Schaefer PM, Singh LN, Wu GD, Murdock
1098 DG & Wallace DC (2019) Host mitochondria influence gut microbiome diversity: A
1099 role for ROS. *Sci Signal* 12: eaaw3159

Zhang Y, Kumarasamy S, Mell B, Cheng X, Morgan EE, Britton SL, Vijay-Kumar M, Koch LG & Joe B (2020) Vertical selection for nuclear and mitochondrial genomes shapes gut microbiota and modifies risks for complex diseases. *Physiol Genomics* 52: 1–14

Zheng C, Zhang S, Ragg S, Raftery D & Vitek O (2011) Identification and quantification of metabolites in 1H NMR spectra by Bayesian model selection. *Bioinformatics* 27: 1637–1644

FIGURES

FIGURE LEGENDS

Figure 1 - Description of the cohort and of the data analysis workflow.

A Key features of the experimental design.

B Overview of the data analysis workflow. On the left, the six datasets used in the study are depicted, indicating whether they were obtained at T0 and T1 or at T0 only.

Data information: written permission for publication of the drawings corresponding to endurance event in the panel A was obtained. In the A panel, the pictures of the blood and plasma tubes were download from https://smart.servier.com. In all cases, no changes were made. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.

Figure 2 - Overview of the mt-related genes.

A Plot of the first two components of the PCA obtained using all of the expressed genes.

B Sankey diagram of the differentially expressed genes, showing the numbers of up-regulated, down-regulated and mt-related genes. Only a relatively small fraction of the mt-related genes (249 out of 801) was up-regulated, whereas most (552 out of 801) were down-regulated.

C Venn diagrams illustrating the overlaps among expressed genes (blue), differentially expressed genes (purple), genes included in WGCNA modules (yellow) and mt-related genes (green).

D Illustration of the proportion of mt-related genes encoded by the mitochondrial and by the nuclear genomes. Except for *MT-ND6*, all of the other genes are encoded by the nuclear genome.

Data information: in panel D, the picture of the mitochondria was download from https://smart.servier.com. In all cases, no changes were made. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.

Figure 3 - Associations between mt-related genes, microbiota and circulating metabolites.

A Dissimilarities in mt-related gene expression represented by the non-metric multidimensional scaling (NMDS) ordination plot. The Bray–Curtis dissimilarity index was calculated on normalised data, the samples were coloured according to the total length of the race and the two different shapes of the dots indicate if the horses finished the race or if they were eliminated.

B Effect sizes of gut microbiota, fecal SCFAs, ^{1}H NMR and biochemical assay metabolites over NMDS ordination. Covariates are coloured according to the type of dataset: ^{1}H NMR metabolites are in orange, biochemical assay metabolites in red, fecal SCFAs in violet and bacteria in dark blue. Horizontal bars show the amount of variance (r^2) explained by each covariate in the model as determined by 'envfit' function.

C Grouped bar chart showing the cumulative effect sizes of covariates on mt-related gene expression (coloured bars) compared to individual effect sizes assuming covariate independence (grey bars) using a stepwise model selection using distance-based redundancy

analysis (dbRDA). Covariates are coloured according to the type of dataset: red for biochemical assay metabolites, and blue for bacteria.

D Plot showing the covariates that contribute significantly to the variation of mt-related genes determined by stepwise model selection using redundancy analysis (RDA). The arrows for each variable show the direction of the effect and are scaled by the unconditioned r^2 value. Covariates are coloured according to the type of dataset: red for biochemical assay metabolites and dark blue for bacteria.

Figure 4 - Data integration using mt-related genes, ^{1}H NMR metabolites, biochemical assay metabolites, fecal SCFAs and gut bacteria.

A Matrix scatterplot showing the correlation between the first components related to each dataset in DIABLO according to the input design.

B CIRCOS plot of the final multi-omics final signature. Each dataset is given a different colour: mt-related genes are in green, ^{1}H NMR metabolites in orange, biochemical assay metabolites in red, fecal SCFAs in violet and gut bacteria in dark blue. Red and blue lines indicate positive and negative correlations between two variables, respectively ($r \geq |0.70|$).

C Visualization of the network obtained with Cytoscape using the final DIABLO multi-omics signature as an input. Only features with more than 15 connections are shown. The size of the nodes indicates the number of interacting partners within the network.

Figure 5 - The bidirectional crosstalk between the gut microbiota and mitochondria in endurance horses.

The intertwined communication between mitochondria and gut microbiota was likely mediated by microbiota derived byproducts (SCFA and secondary bile acids), which regulate mitochondrial redox balance, inflammation and energy production during intense exercise.

Among the SCFA, butyrate appeared as a key regulator of mitochondrial energy production and oxidative stress.

Increased lactate and urea concentrations upon prolonged exercise likely entered the gut lumen and were subsequently transformed into SCFA. It is also suggested that circulating free fatty acids participated in the mitochondrial regulation of the inflammatory processes elicited by oxidative stress, microbial dissemination and microbial lipopolysaccharides translocation outside of the gastrointestinal tract, as often occurs in endurance athletes. Whether these mechanisms confer an advantage for endurance performance remains still speculative, but results raise the possibility that gut-microbiota crosstalk is pivotal for greater energy availability, aerobic metabolism, glycogen preservation, resistance to fatigue and to maintain speed during the race.

Written permission for publication of the horse drawings was obtained. The pictures of the mitochondria and gut were downloaded from https://smart.servier.com. In all cases, no changes were made. Servier Medical Art by Servier is licensed under a Creative Commons Attribution 3.0 Unported License.

EXPANDED VIEW INFORMATION

Creation of a consensus list for literature-based meta-analytical analysis of mitochondrial-related genes.

We retrieved four gene lists: (1) the Integrated Mitochondrial Protein Index (IMPI) gene list (Smith & Robinson, 2019). Only the human gene list ('Human IMPI genes', 'MitoMiner version Q2 2018', downloaded from http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/impi.do) was retained (1,626 genes) to facilitate finding horse/human orthologs; (2) the Mitocarta Inventory (Calvo *et al*, 2016) gene list. As before, only the human gene list ('Human MitoCarta 2.0 genes', 'MitoCarta 2.0 genes in MitoMiner' downloaded from

http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/mitocarta.do) was retained (1,158 genes) to facilitate the identification of horse/human orthologs; (3) all of the 618 genes included in the following KEGG pathways: hsa00020, hsa00061, hsa00062, hsa00071, hsa00072, hsa00100, hsa00130, hsa00140, hsa00190, hsa00240, hsa00280, hsa00290, hsa00471, hsa00480, hsa00628, hsa00760, hsa00860, hsa01212, hsa04020, hsa04024, hsa04072, hsa04115, hsa04136, hsa04137, hsa04140, hsa04210, hsa04215, hsa04350, hsa04370, hsa04668 and hsa04979. The genes in each KEGG pathway were retrieved from the KEGG database (release 96.0) using the R package KEGGREST; (4) a custom list including 103 genes found in the literature (Pearce *et al*, 2017; Bianchessi *et al*, 2016; Nicholls & Gustafsson, 2018; Wang *et al*, 2016; Cosson *et al*, 2012; Rizzuto *et al*, 2012; Gustafsson & Samuelsson, 2001; Gustafsson *et al*, 2016; Lee *et al*, 2015). These four gene sets were merged to create a consensus list, which included 2,082 unique genes (Table EV4).

EXPANDED VIEW FIGURES

Figure EV1 - Correlations between eigengene modules and metabolites from WGCNA method.

Each module is labelled with a unique colour as an identifier. A module is considered positively (highlighted in red) or negatively (highlighted in blue) associated to metabolites if the Pearson *r* correlation values were $\geq |0.65|$ for at least 5 molecules (highlighted again in red or blue) and if the corresponding *p*-values are $\leq$ 1e-05. Within each cell, upper values indicate the correlation values between modules and metabolites, while lower values are the corresponding *p*-values.

Figure EV2 - Descriptive KEGG pathway classification bar plot obtained using the consensus list of genes related to mitochondria.

The horizontal bars represent the absolute number of genes found in third-level KEGG pathways, grouped in first-level KEGG pathways using a colour code. The vertical bars on the right indicate the names of first-level pathways.

Figure EV3 - Functional classification of the mt-related genes.

Functional classification of mt-related genes obtained with ClueGO. The chart shows the functional groups found as enriched, and the name of the group is represented by the group leading term. The significant enriched functional groups are marked with a red *.

Figure EV4 - Composition of the core microbiota.

A heatmap showing the core microbiota and its prevalence at different detection thresholds. Only the genera shared by 99% of individuals in the cohort and with a minimum detection threshold of 0.1% are shown.

B Circular stacked barplot of the main genera included in the core genome. Genera are coloured according to the phylum they belong to.

EXPANDED VIEW LARGE TABLES

Table EV1 - Metadata of the horses recruited in the experiment.

The letters between brackets in the “Ranking” column indicate the causes of the elimination of the horse: “L” corresponds to “lameness”, R” to “retired” and “M” to “metabolic issues”.

Table EV2 - List of genes found differentially expressed in whole blood in T1 with respect to T0 horses.

Columns two and three show the mean $\log_2$ normalized expression values in the two time points, while columns four and five show the $\log_2$ FC values and the Bonferroni adjusted *p*-

values. Columns six to nine indicate the genes that are also found in the mt-related consensus list. The gene lists used to create the final consensus are indicated separately.

Table EV3 - List of genes included in each of the three eigengene modules found as correlated to metabolites using WGCNA.

The modules are indicated using the default WGCNA naming convention, and highlighted in blue to indicate negative correlation and red to indicate positive correlation.

Table EV4 - Consensus list of genes related to mitochondria obtained as described in the Expanded View Information.

Gene description, gene localizations and gene types were determined according to IPA (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis).

Table EV5 - List of mt-related genes, defined as the intersection between the list included in the Table EV4 and the set of expressed genes.

Gene description, gene localizations and gene types were determined according to IPA (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Columns from N from AG show the expression values found in the "E2" matrix, which were obtained by subtracting the T0 from the T1 expression matrix values, i.e., by calculating the ratio between T1 and T0 log scaled expression values from the two matrices.

Table EV6 - Relative abundance of metabolites obtained from blood of the 20 horses under study collected before and after the endurance ride.

Table EV7 - Biochemical parameters obtained from the blood of the 20 horses under

study collected before and after the endurance race.

Table EV8 - ASV taxonomical assignments and ASV counts for the 20 horses under study.

Table EV9 - Non-normalized annotated abundance genera table observed in the fecal samples of the 20 horses under study.

Table EV10 - Fecal pH and fecal short chain fatty acids measurements in the 20 horses under study before the endurance race.

Table EV11 - Concentrations of bacteria, ciliate protozoa and anaerobic fungi in the feces of the 20 horses under study.

Table EV12 - Correlated variables obtained using DIABLO on all of the available data sets.

Gene descriptions and localizations were determined using the IPA database (https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis). Molecular pathways were determined using ClueGO 2.5.7 in the case of genes.

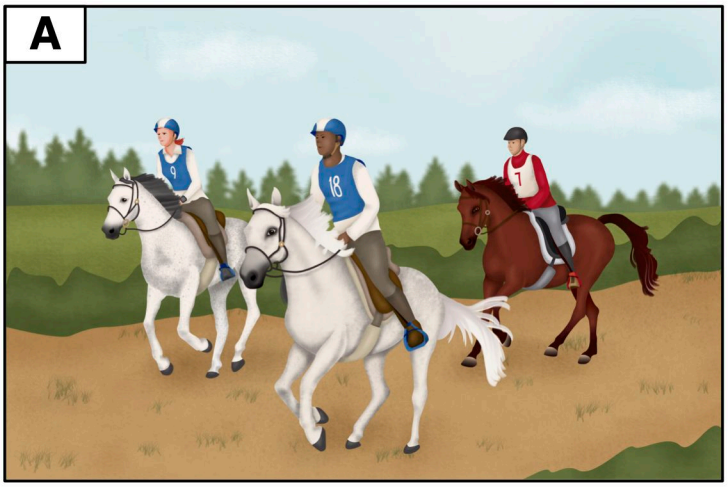
Table EV13 - Correlation matrix of the associations between mt-related genes and bacterial genera obtained using the rCCA method.

Only the genes and the genera for which at least one association value data point presented $r \geq |0.55|$ are shown.

Table EV14 - Genera table obtained using the mixMC framework.

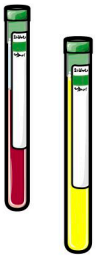
Genera with less than 1% counts with respect to the total number were removed, and subsequently a centered log ratio transformation was applied.

A



Key features of the experimental design

- 20 endurance horses
- T0 (pre-ride) and T1 (post-ride) measurements
- Similar age and performance
- Matched individual data acquisition:
  - Blood transcriptome
  - Blood ^{1}H NMR metabolites
  - Blood biochemical assays
  - Fecal microbiota
  - Fecal SCFAs
  - Fecal microorganisms loads



B

Blood profiling: datasets

Gene expression T0 ↔ Gene expression T1
n = 20 *n* = 20
DE genes
mt-related DE genes

^{1}H NMR metabolites T0 ↔ ^{1}H NMR metabolites T1
n = 20 *n* = 20

Biochemical assays T0 ↔ Biochemical assays T1
n = 20 *n* = 20

Fecal profiling: datasets

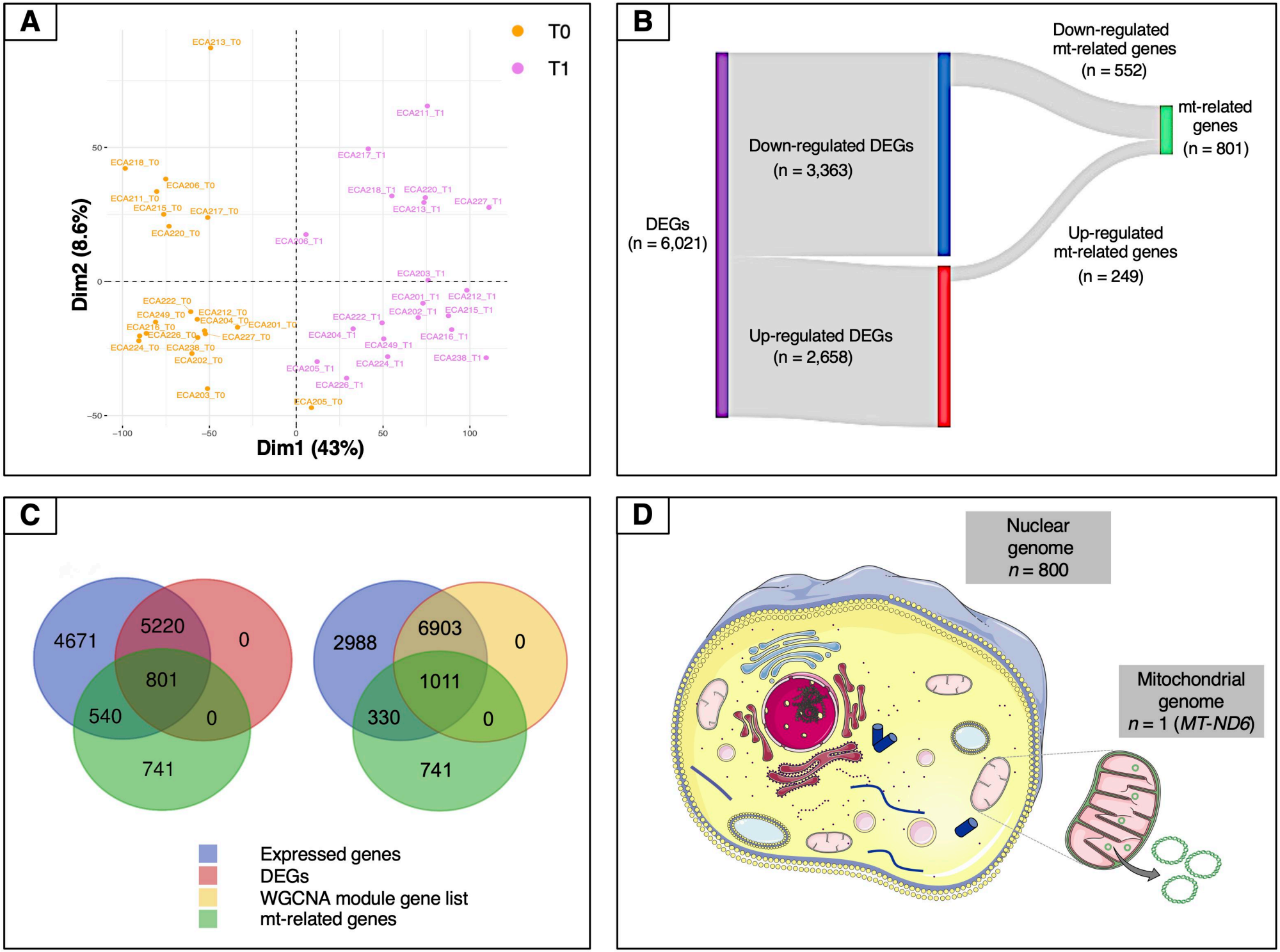
Microbiome T0
n = 20

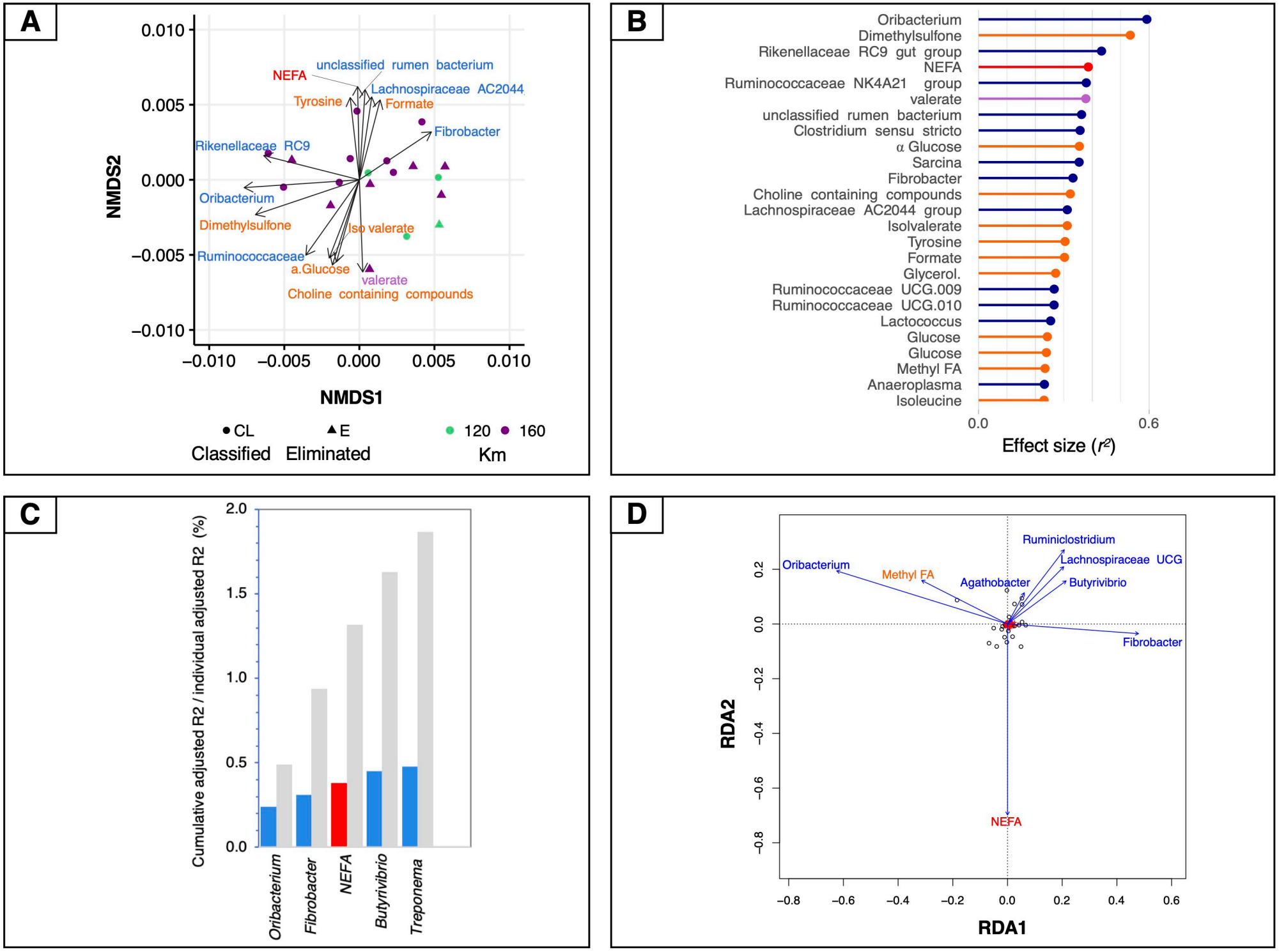
SCFAs T0
n = 20

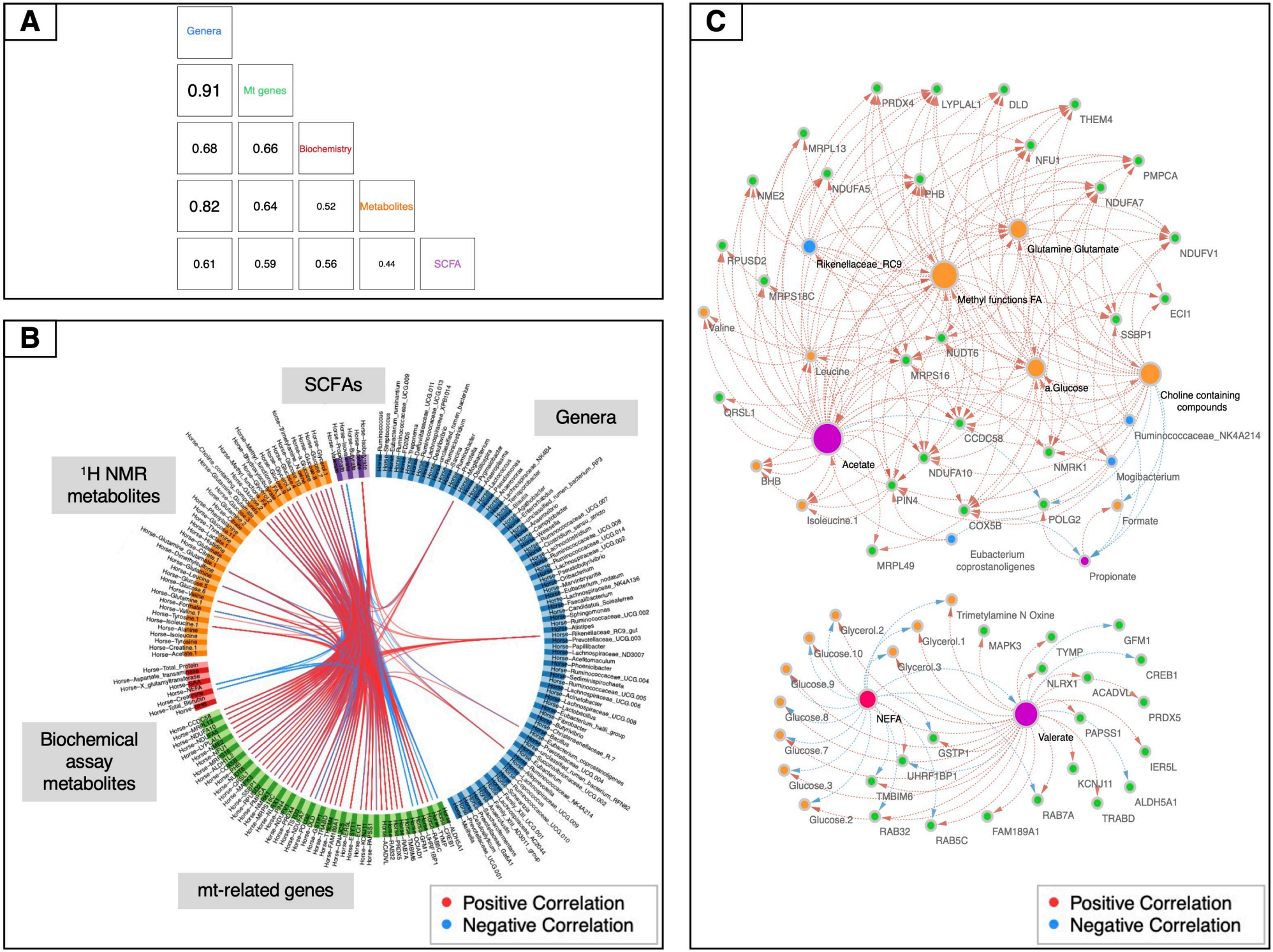
Microorganisms loads T0
n = 20

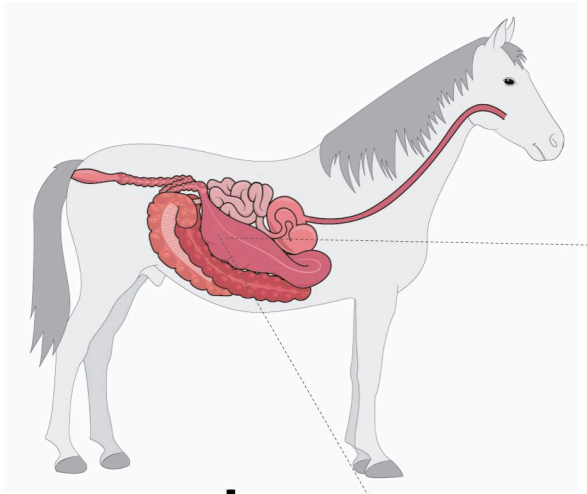
Data integration

- Molecular signatures
- Novel mechanisms: microbiota / mitochondria crosstalk
- Strategies to improve microbiota / mitochondria crosstalk


A
T0
T1
Dim2 (8.6%)
Dim1 (43%)
B
DEGs
(n = 6,021)
Down-regulated DEGs
(n = 3,363)
Up-regulated DEGs
(n = 2,658)
Down-regulated
mt-related genes
(n = 552)
Up-regulated
mt-related genes
(n = 249)
mt-related
genes
(n = 801)
C
4671
5220
0
801
540
0
741
2988
6903
0
1011
330
0
741
Expressed genes
DEGs
WGCNA module gene list
mt-related genes
D
Nuclear
genome
n = 800
Mitochondrial
genome
n = 1 (MT-ND6)

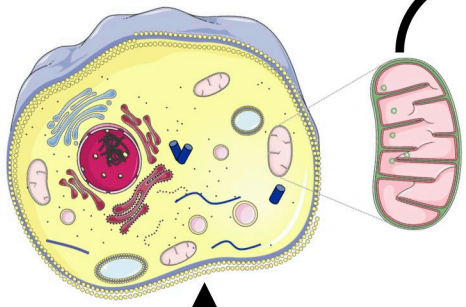

A
0.010
0.005
0.000
−0.005
−0.010
NMDS2
NEFA
unclassified rumen bacterium
Lachnospiraceae AC2044
Tyrosine
Formate
Fibrobacter
Rikenellaceae RC9
Oribacterium
Dimethylsulfone
Iso valerate
Ruminococcaceae
a.Glucose
valerate
Choline containing compounds
−0.010
−0.005
0.000
0.005
0.010
NMDS1
CL
E
120
160
Classified
Eliminated
Km
B
Oribacterium
Dimethylsulfone
Rikenellaceae RC9 gut group
NEFA
Ruminococcaceae NK4A21 group
valerate
unclassified rumen bacterium
Clostridium sensu stricto
α Glucose
Sarcina
Fibrobacter
Choline containing compounds
Lachnospiraceae AC2044 group
Isolvalerate
Tyrosine
Formate
Glycerol.
Ruminococcaceae UCG.009
Ruminococcaceae UCG.010
Lactococcus
Glucose
Glucose
Methyl FA
Anaeroplasma
Isoleucine
0.0
0.6
Effect size (r^2)
C
Cumulative adjusted R2 / individual adjusted R2 (%)
2.0
1.5
1.0
0.5
0.0
Oribacterium
Fibrobacter
NEFA
Butyrivibrio
Treponema
D
Ruminiclostridium
Lachnospiraceae UCG
Oribacterium
Methyl FA
Agathobacter
Butyrivibrio
Fibrobacter
NEFA
RDA2
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−0.8
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6
RDA1


A
Genera
0.91
Mt genes
0.68
0.66
Biochemistry
0.82
0.64
0.52
Metabolites
0.61
0.59
0.56
0.44
SCFA
B
SCFAs
Genera
¹H NMR metabolites
Biochemical assay metabolites
mt-related genes
Positive Correlation
Negative Correlation
C
PRDX4
LYPLAL1
DLD
THEM4
MRPL13
NFU1
PMPCA
NME2
NDUFA5
PHB
NDUFA7
Glutamine Glutamate
NDUFV1
RPUSD2
Rikenellaceae_RC9
Methyl functions FA
MRPS18C
ECI1
Valine
SSBP1
NUDT6
Leucine
MRPS16
a.Glucose
Choline containing compounds
QRSL1
CCDC58
Ruminococcaceae_NK4A214
Acetate
NMRK1
BHB
NDUFA10
Mogibacterium
PIN4
Isoleucine.1
COX5B
POLG2
Formate
MRPL49
Eubacterium coprostanoligenes
Propionate
Trimetylamine N Oxine
Glycerol.2
Glycerol.1
MAPK3
GFM1
Glucose.10
TYMP
Glycerol.3
CREB1
Glucose.9
NLRX1
ACADVL
Glucose.8
NEFA
PRDX5
Valerate
PAPSS1
Glucose.7
GSTP1
IER5L
UHRF1BP1
Glucose.3
KCNJ11
TMBIM6
ALDH5A1
Glucose.2
RAB7A
TRABD
RAB32
FAM189A1
RAB5C



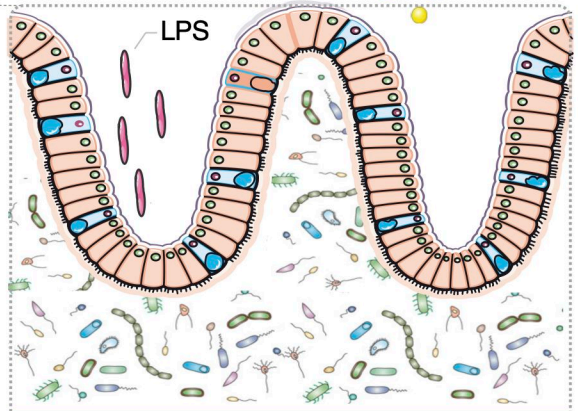
Endurance response

- Lactate and urea production
- NEFA release
- Systemic inflammation
- Increased endotoxin flux
- Leaky intestinal barrier



Secondary bile acids

SCFAs

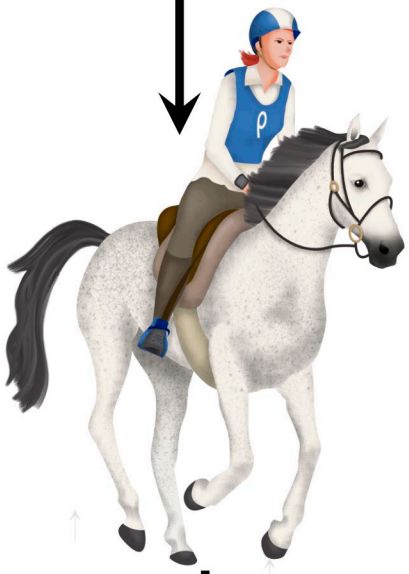


Gut microbiota

- Fermentation to SCFA
- Bile acid transformation
- Conversion of lactate and urea
- LPS translocation

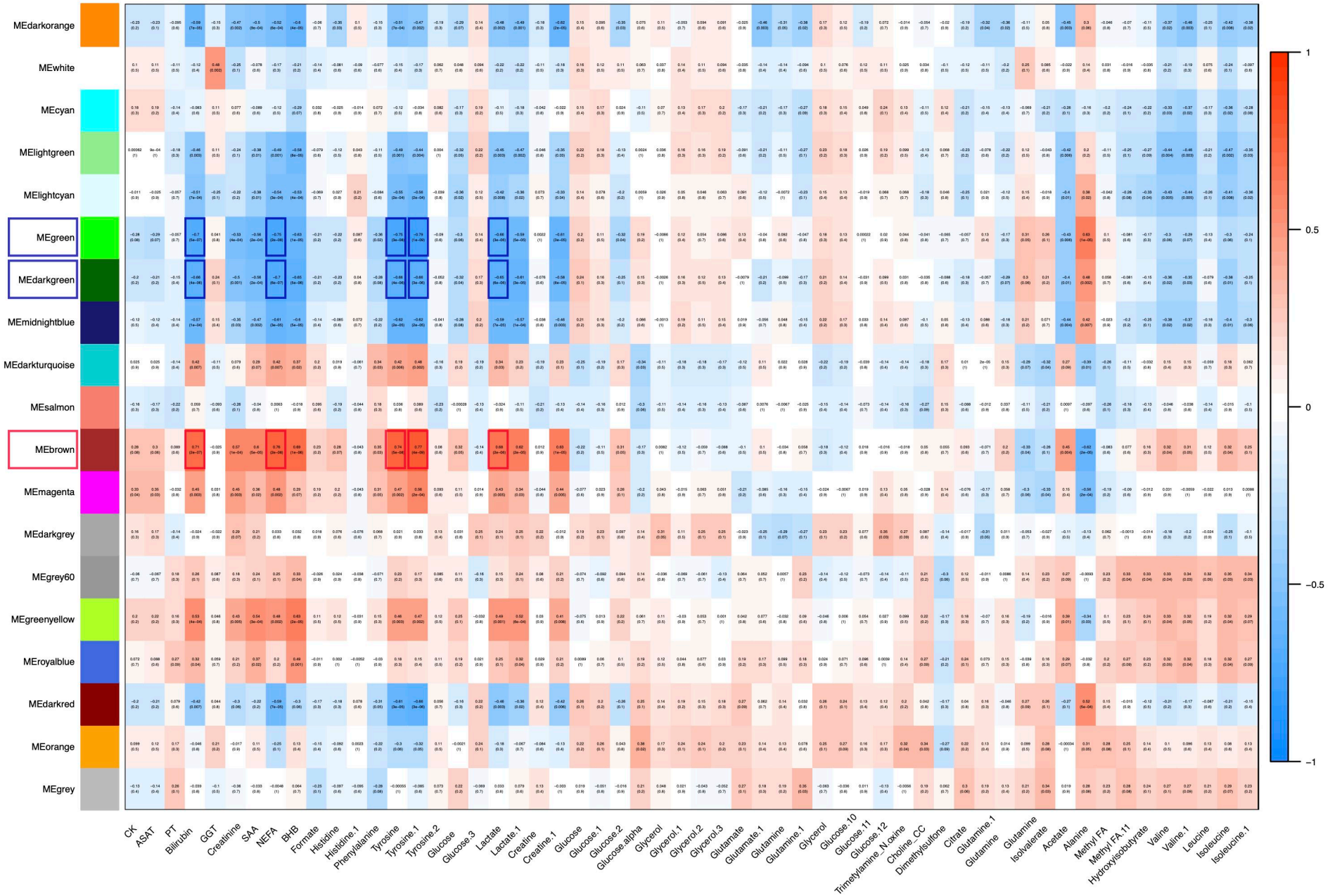
Mitochondrial functions

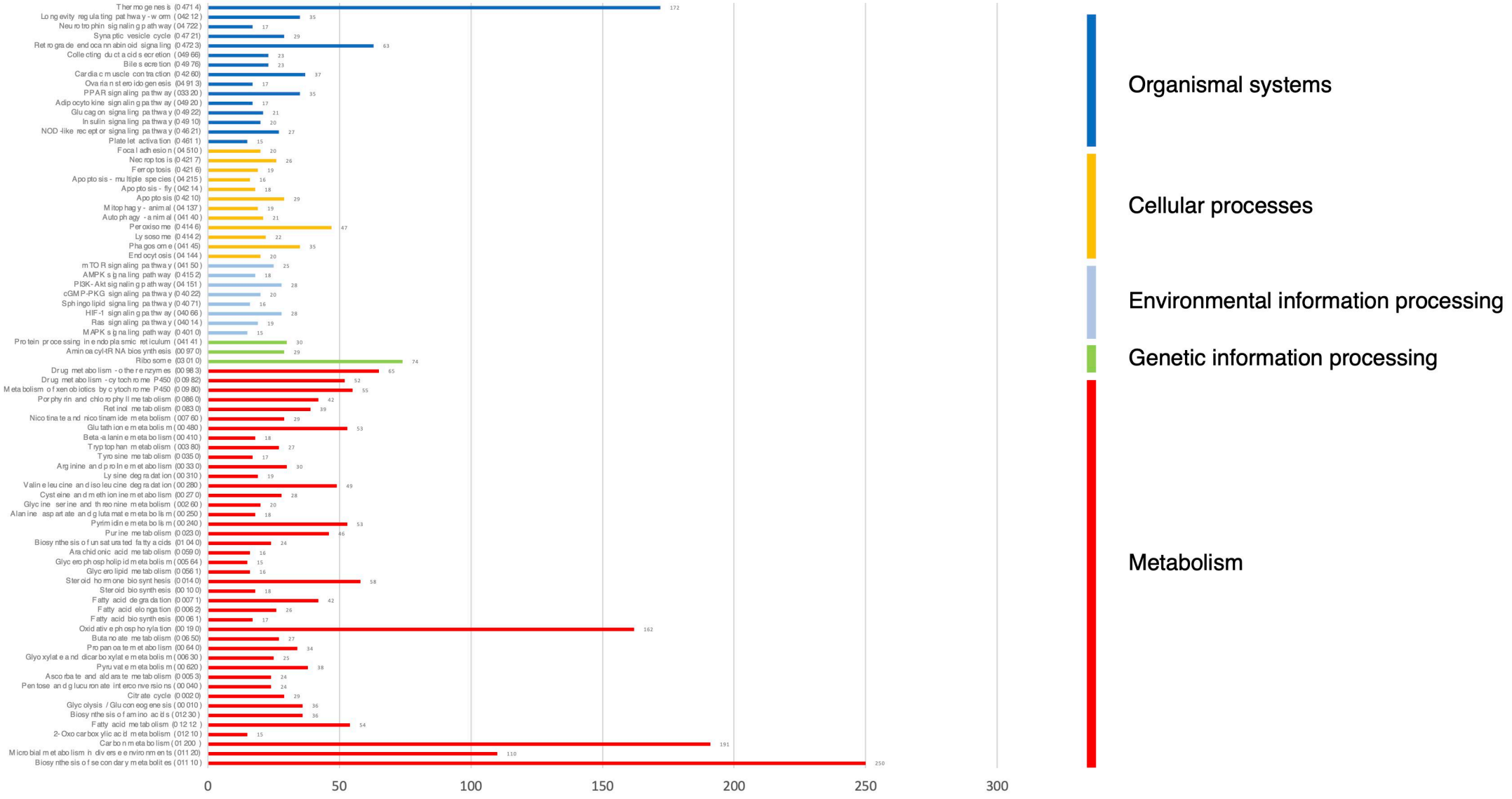
- Increased FAO
- Increased OXPHOS
- Mitochondrial biogenesis
- Apoptosis
- Inflammasome

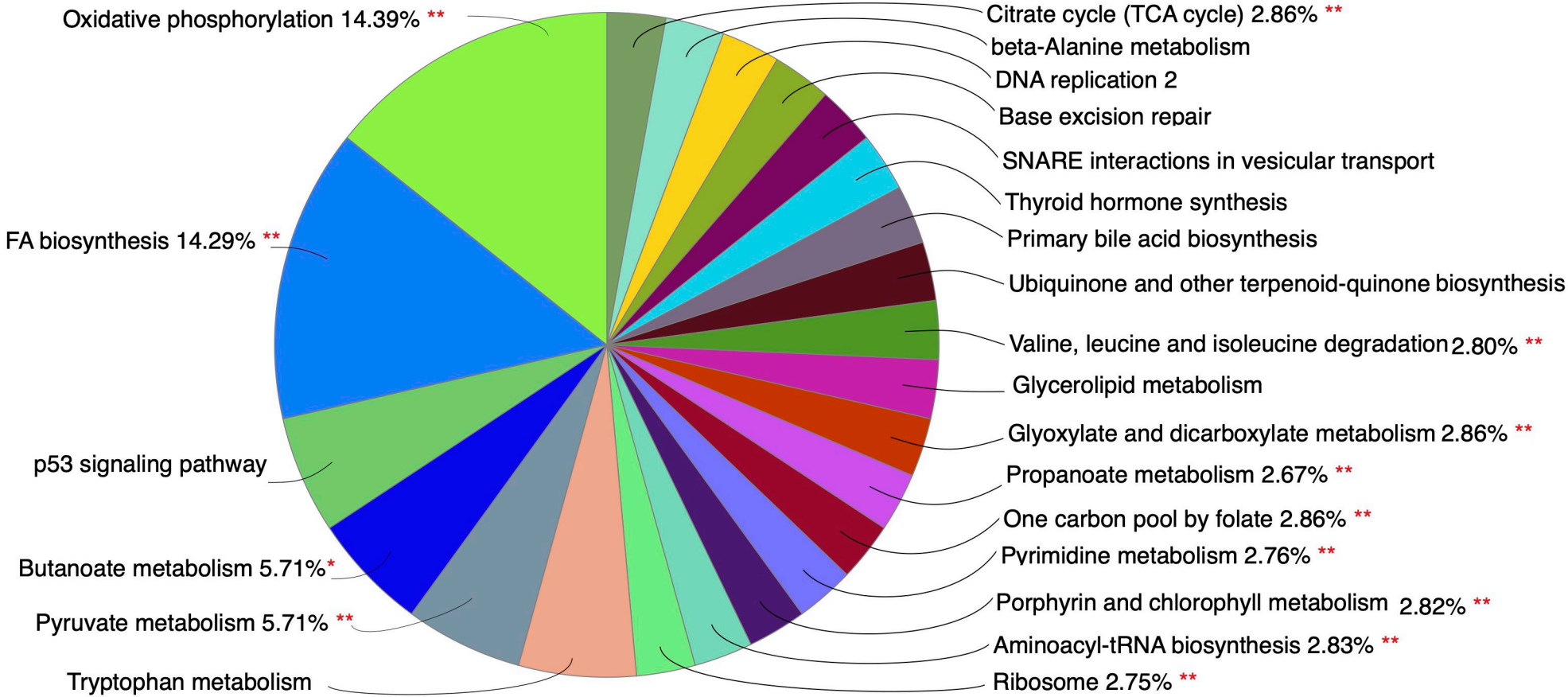


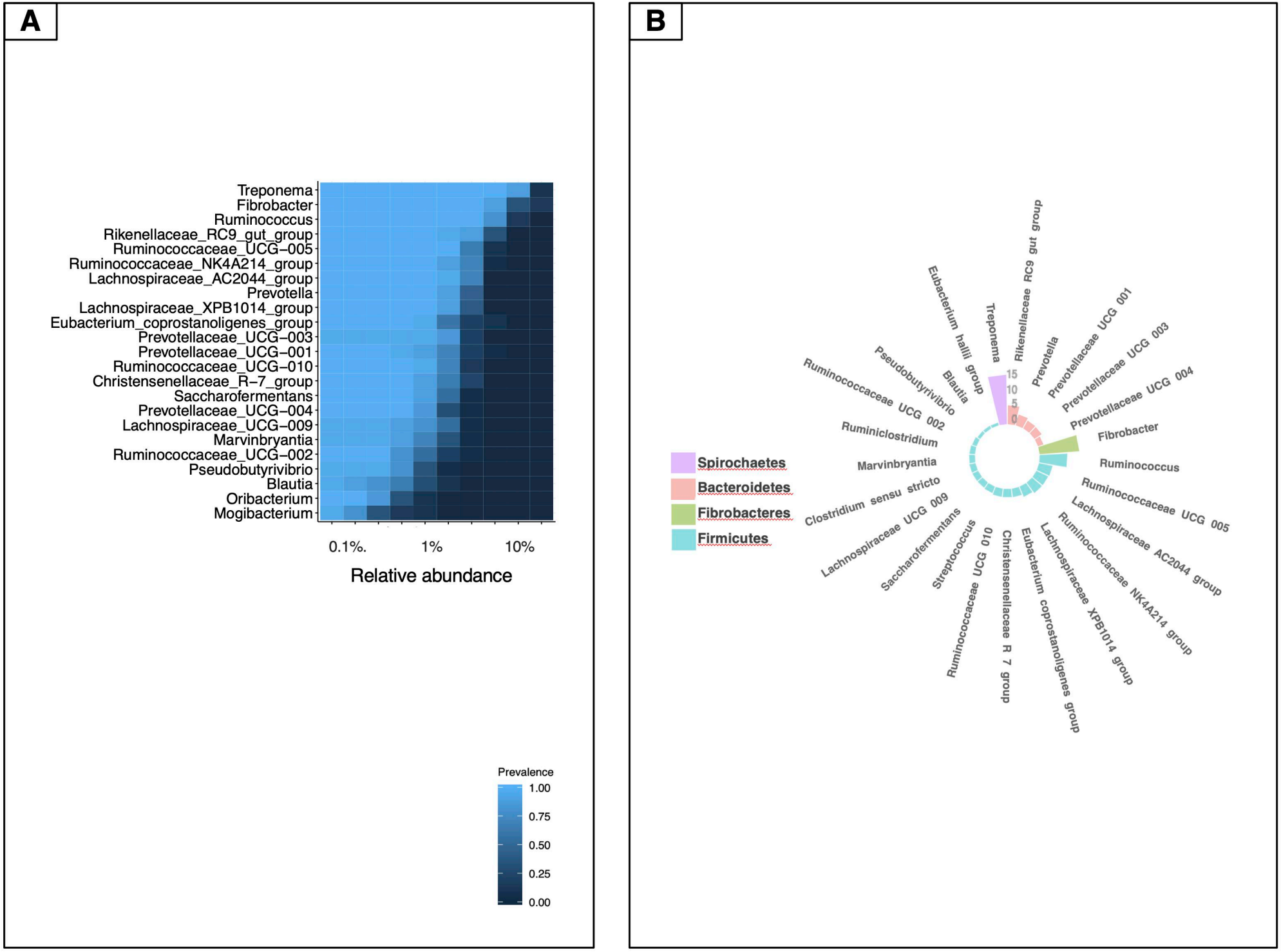
Endurance performance

- Greater aerobic metabolism
- Greater energy availability
- Reduction of blood lactate
- Glycogen preservation
- Resistance to fatigue
- Ability to mantain speed


MEdarkorange
MEwhite
MEcyan
MElightgreen
MElightcyan
MEgreen
MEdarkgreen
MEmidnightblue
MEdarkturquoise
MEsalmon
MEbrown
MEmagenta
MEdarkgrey
MEgrey60
MEgreenyellow
MEroyalblue
MEdarkred
MEorange
MEgrey
CK
ASAT
PT
Bilirubin
GGT
Creatinine
SAA
NEFA
BHB
Formate
Histidine
Histidine.1
Phenylalanine
Tyrosine
Tyrosine.1
Tyrosine.2
Glucose
Glucose.3
Lactate
Lactate.1
Creatine
Creatine.1
Glucose
Glucose.1
Glucose.2
Glucose.alpha
Glycerol
Glycerol.1
Glycerol.2
Glycerol.3
Glutamate
Glutamate.1
Glutamine
Glutamine.1
Glycerol
Glucose.10
Glucose.11
Glucose.12
Trimethylamine_N.oxine
Choline_CC
Dimethylsulfone
Citrate
Glutamine.1
Glutamine
Isovalerate
Acetate
Alanine
Methyl FA
Methyl FA.11
Hydroxyisobutyrate
Valine
Valine.1
Leucine
Isoleucine
Isoleucine.1
1
0.5
0
−0.5
−1


KEGG descriptive classification barplot
Thermogenesis (04714)
Longevity regulating pathway - worm (04212)
Neurotrophin signaling pathway (04722)
Synaptic vesicle cycle (04721)
Retrograde endocannabinoid signaling (04723)
Collecting duct acid secretion (04966)
Bile secretion (04976)
Cardiac muscle contraction (04260)
Ovarian steroidogenesis (04913)
PPAR signaling pathway (03320)
Adipocytokine signaling pathway (04920)
Glucagon signaling pathway (04922)
Insulin signaling pathway (04910)
NOD-like receptor signaling pathway (04621)
Platelet activation (04611)
Focal adhesion (04510)
Necroptosis (04217)
Ferroptosis (04216)
Apoptosis - multiple species (04215)
Apoptosis - fly (04214)
Apoptosis (04210)
Mitophagy - animal (04137)
Autophagy - animal (04140)
Peroxisome (04146)
Lysosome (04142)
Phagosome (04145)
Endocytosis (04144)
mTOR signaling pathway (04150)
AMPK signaling pathway (04152)
PI3K-Akt signaling pathway (04151)
cGMP-PKG signaling pathway (04022)
Sphingolipid signaling pathway (04071)
HIF-1 signaling pathway (04066)
Ras signaling pathway (04014)
MAPK signaling pathway (04010)
Protein processing in endoplasmic reticulum (04141)
Aminoacyl-tRNA biosynthesis (00970)
Ribosome (03010)
Drug metabolism - other enzymes (00983)
Drug metabolism - cytochrome P450 (00982)
Metabolism of xenobiotics by cytochrome P450 (00980)
Porphyrin and chlorophyll metabolism (00860)
Retinol metabolism (00830)
Nicotinate and nicotinamide metabolism (00760)
Glutathione metabolism (00480)
Beta-alanine metabolism (00410)
Tryptophan metabolism (00380)
Tyrosine metabolism (00350)
Arginine and proline metabolism (00330)
Lysine degradation (00310)
Valine leucine and isoleucine degradation (00280)
Cysteine and methionine metabolism (00270)
Glycine serine and threonine metabolism (00260)
Alanine aspartate and glutamate metabolism (00250)
Pyrimidine metabolism (00240)
Purine metabolism (00230)
Biosynthesis of unsaturated fatty acids (01040)
Arachidonic acid metabolism (00590)
Glycerophospholipid metabolism (00564)
Glycerolipid metabolism (00561)
Steroid hormone biosynthesis (00140)
Steroid biosynthesis (00100)
Fatty acid degradation (00071)
Fatty acid elongation (00062)
Fatty acid biosynthesis (00061)
Oxidative phosphorylation (00190)
Butanoate metabolism (00650)
Propanoate metabolism (00640)
Glyoxylate and dicarboxylate metabolism (00630)
Pyruvate metabolism (00620)
Ascorbate and aldarate metabolism (00053)
Pentose and glucuronate interconversions (00040)
Citrate cycle (00020)
Glycolysis / Gluconeogenesis (00010)
Biosynthesis of amino acids (01230)
Fatty acid metabolism (01212)
2-Oxocarboxylic acid metabolism (01210)
Carbon metabolism (01200)
Microbial metabolism in diverse environments (01120)
Biosynthesis of secondary metabolites (01110)
172
35
17
29
63
23
23
37
17
35
17
21
20
27
15
20
26
19
16
18
29
19
21
47
22
35
20
25
18
28
20
16
28
19
15
30
29
74
65
52
55
42
39
29
53
18
27
17
30
19
49
28
20
18
53
46
24
16
15
16
58
18
42
26
17
162
27
34
25
38
24
24
29
36
36
54
15
191
110
250
0
50
100
150
200
250
300
Organismal systems
Cellular processes
Environmental information processing
Genetic information processing
Metabolism


Oxidative phosphorylation 14.39% **
Citrate cycle (TCA cycle) 2.86% **
beta-Alanine metabolism
DNA replication 2
Base excision repair
SNARE interactions in vesicular transport
Thyroid hormone synthesis
Primary bile acid biosynthesis
Ubiquinone and other terpenoid-quinone biosynthesis
FA biosynthesis 14.29% **
Valine, leucine and isoleucine degradation 2.80% **
Glycerolipid metabolism
Glyoxylate and dicarboxylate metabolism 2.86% **
p53 signaling pathway
Propanoate metabolism 2.67% **
One carbon pool by folate 2.86% **
Butanoate metabolism 5.71%*
Pyrimidine metabolism 2.76% **
Porphyrin and chlorophyll metabolism 2.82% **
Pyruvate metabolism 5.71% **
Aminoacyl-tRNA biosynthesis 2.83% **
Tryptophan metabolism
Ribosome 2.75% **


A
Treponema
Fibrobacter
Ruminococcus
Rikenellaceae_RC9_gut_group
Ruminococcaceae_UCG−005
Ruminococcaceae_NK4A214_group
Lachnospiraceae_AC2044_group
Prevotella
Lachnospiraceae_XPB1014_group
Eubacterium_coprostanoligenes_group
Prevotellaceae_UCG−003
Prevotellaceae_UCG−001
Ruminococcaceae_UCG−010
Christensenellaceae_R−7_group
Saccharofermentans
Prevotellaceae_UCG−004
Lachnospiraceae_UCG−009
Marvinbryantia
Ruminococcaceae_UCG−002
Pseudobutyrivibrio
Blautia
Oribacterium
Mogibacterium
0.1%.
1%
10%
Relative abundance
Prevalence
1.00
0.75
0.50
0.25
0.00
B
Spirochaetes
Bacteroidetes
Fibrobacteres
Firmicutes
15
10
5
0
Treponema
Rikenellaceae RC9 gut group
Prevotella
Prevotellaceae UCG 001
Prevotellaceae UCG 003
Prevotellaceae UCG 004
Fibrobacter
Ruminococcus
Ruminococcaceae UCG 005
Lachnospiraceae AC2044 group
Ruminococcaceae NK4A214 group
Lachnospiraceae XPB1014 group
Eubacterium coprostanoligenes group
Christensenellaceae R 7 group
Ruminococcaceae UCG 010
Streptococcus
Saccharofermentans
Lachnospiraceae UCG 009
Clostridium sensu stricto
Marvinbryantia
Ruminiclostridium
Ruminococcaceae UCG 002
Pseudobutyrivibrio
Blautia
Eubacterium hallii group