Review in Advance first posted online on August 13, 2015. (Changes may still occur before final publication online and in print.)

Microbiology Meets Big Data: The Case of Gut Microbiota–Derived Trimethylamine

Gwen Falony,[1,2,*] Sara Vieira-Silva,[1,2,*] and Jeroen Raes[1,2,3]

[1]Department of Microbiology and Immunology, Rega Institute, KU Leuven, B-3000 Leuven, Belgium; email: jeroen.raes@med.kuleuven.be

[2]Center for the Biology of Disease, VIB, B-3000 Leuven, Belgium

[3]Microbiology Unit, Faculty of Sciences and Bioengineering Sciences, Vrije Universiteit Brussel, B-1050 Brussels, Belgium

Annu. Rev. Microbiol. 2015. 69:305–21

The *Annual Review of Microbiology* is online at micro.annualreviews.org

This article's doi: 10.1146/annurev-micro-091014-104422

*These authors contributed equally to this review.

Keywords

microbiota, metagenomics, gut metabolism, trimethylamine, TMA, trimethylamine *N*-oxide, TMAO, atherosclerosis, cardiovascular disease

Abstract

During the past decade, meta-omics approaches have revolutionized microbiology, allowing for a cultivation-free assessment of the composition and functional properties of entire microbial ecosystems. On the one hand, a phylogenetic and functional interpretation of such data relies on accumulated genetic, biochemical, metabolic, and phenotypic characterization of microbial variation. On the other hand, the increasing availability of extensive microbiome data sets and corresponding metadata provides a vast, underused resource for the microbiology field as a whole. To demonstrate the potential for integrating big data into a functional microbiology workflow, we review literature on trimethylamine (TMA), a microbiota-generated metabolite linked to atherosclerosis development. Translating recently elucidated microbial pathways resulting in TMA production into genomic orthologs, we demonstrate how to mine for their presence in public (meta-) genomic databases and link findings to associated metadata. Reviewing pathway abundance in public data sets shows that TMA production potential is associated with symptomatic atherosclerosis and allows identification of currently uncharacterized TMA-producing bacteria.

Contents

INTRODUCTION

The claim that microbiome research has changed how microbiology is practiced is most probably an understatement. Meta-omics approaches have gained a reputation for being the ultimate tools to unravel the full metabolic potential of any microbial community because they allow the exploration of community composition and functionality as a whole and circumvent any potential cultivation bias. Unfortunately, in the midst of the metagenomics hype, the quality of microbiome research was often estimated on the basis of sample and read quantity, which provided catalogs of genes and species, whereas true functional microbial analyses often remained rather shallow. Now that the field has matured, there is a general consensus that metagenomic microbiology should go hand in hand with bacterial culturing, phenotypical characterization, and metabolic pathway exploration efforts to interpret and validate microbiome-based observations. Additionally, genomic, biochemical, functional, and applied microbiology can benefit from big data to guide and target exploration of bacterial ecology and to assess the impact of the findings on a community level. Although the benefits of this mutualistic symbiosis are obvious, synergy is hampered by the widespread out-of-the-box application of generic analysis tools for high-level analysis in meta-omics studies and by the hesitation of traditional microbiologists to screen existing metagenomics databases to assess the full scope of their results and to study effects at the community level and in relation to environmental and/or clinical parameters. In this article, we demonstrate how to translate biochemically elucidated metabolic pathways into microbiome exploration tools, and we show their effectiveness in metagenomics (meta–) data set mining. As a case study, we focus on the microbial background of trimethylamine (TMA) production in the human colon ecosystem and its potential link to atherosclerosis development.

PEEKING INSIDE THE BLACK BOX OF THE GUT ECOSYSTEM USING META-OMICS

One of the primary target areas of microbiome research has been the human microbiota. Only 10% of the cells in our bodies are of human origin; the vast majority are microbial (43). Together, we form a human supraorganism of host and associated microbes that functions through a multitude of symbiotic interactions at various levels. The bulk of human body–associated microorganisms are located in the colon. There, the host provides its symbionts with a safe, nutrient-rich, and homeostatic environment. In turn, the microbes not only expand our digestive capacity but also produce essential nutrients, increase colonization resistance against pathogenic intruders, and assist in the detoxification of xenobiotics.

In the past decade, advances in sequencing technology enabled new approaches in microbial ecosystem research (39), including metagenomics—the random shotgun sequencing of total DNA from environmental or clinical samples. These developments provided the opportunity to study human body–associated microbial communities at an unprecedented scale. This allowed the phylogenetic assessment of the uncultivable fraction of commensal communities and simultaneously permitted functional insights into individual microorganisms and the ecosystem as a whole (15, 39). Metagenomic analyses have lifted the cover of the intestinal black box, which has allowed researchers to effectively link functional and phylogenetic alterations in the gut microbiota to pathological states such as (low-grade) inflammation, infections, autoimmune diseases, multifactorial disorders, and cancer (8, 13, 20, 27, 37). Moreover, evidence has been found that gut microbiota composition and metabolism can serve as an indicator of chronic suboptimal health and well-being, either through a direct link to suboptimal bowel functioning or through an extension to general physical and psychological health (10, 40).

Due to its population density, complexity, and proximity as well as continuous interaction with the host, the human colon microbiota represents one of the most challenging and intriguing microbial ecosystems on earth. As microbiome research continues to evolve, gut microbiota monitoring is increasingly emerging as a primary gateway to intestinal and general health assessment and management. It is clear that gut microbiome dysbiosis has a (co-)causative role in a large number of human conditions of suboptimal health and pathologies. For such conditions, comprehensive gut microbiome analysis of patient cohorts could not only reveal prognostic or diagnostic biomarkers but also enable hypotheses on pathomechanisms and target identification of novel lifestyle/nutritional/pharmaceutical microbiota–modulation strategies by providing functional insights. Apart from diseases in which a causative role has been identified, multiple human pathologies have shown a correlation with gut microbiome changes. Although there is limited use for associations that lack causative or mechanistic insight in the development of novel therapeutic approaches, the correlations identified could allow the use of microbial signature profiles (based on species, taxa, genes, and metabolic pathways) as biomarkers to predict, diagnose, or subcategorize diseases. Already, microbiome signature-based biomarker development has been applied to pathologies such as inflammatory bowel disease (IBD; e.g., Crohn's disease and ulcerative colitis) (20), diabetes (37), metabolic (27, 58) and cardiovascular diseases (CVDs) (22), rheumatism (44), and colorectal cancer (66), among others (27, 44, 58). Because of these promising developments, the excitement surrounding high-throughput microbiome research has spread throughout a large part of the life sciences community.

BIG DATA–ASSISTED, BOTTOM-UP MICROBIOLOGY

The prevailing model in contemporary microbiota research is that of big-data, big-consortium, top-down projects in which large (mostly cross-sectional) studies yield novel hypotheses for disease

mechanisms (38). The broad impact generated by such studies often gives the impression that basic microbiology and biochemistry have become obsolete. Of course, nothing could be further from the truth. Sequencing and bioinformatics, despite their potential, have their limitations. Microbial models, culturing, and molecular microbiology are indispensable for teasing out disease mechanisms and for identifying modes of action for novel therapies. At the same time, big data is shunned by basic microbiology adepts because of its mostly hypothesis-free, data-driven approach to science. To reconcile these views, we focus on how bottom-up microbiological research can benefit from using high-throughput techniques and from mining publicly available large data sets. We illustrate this process by showing how (*a*) the biochemical pathways for the production of an important microbial metabolite in CVD were elucidated by building on vast public genomic data and (*b*) 16S rDNA amplicon sequencing approaches helped to further our understanding of which organisms are associated with this process. In addition, we show how this knowledge can be further expanded by genome and metagenome mining of publicly available data and can even lead to the discovery of novel clinical hypotheses, i.e., we show how basic microbiology finds its way into big data.

TRIMETHYLAMINE: A MICROBIAL METABOLITE CONTRIBUTING TO ATHEROSCLEROSIS DEVELOPMENT

Recently, alterations in gut microbiota composition and metabolic potential have been identified as contributing factors in CVD development (46, 61). In particular, trimethylamine *N*-oxide (TMAO), the hepatic oxidation product of the colon microbial metabolite TMA, has gained a lot of attention as a potential promoter of atherosclerosis (49). Most remarkably, this research line, which undoubtedly is among the closest to identifying a smoking gun in microbiota-disease interactions, originated not from microbiome explorative efforts as described above but from untargeted metabolomic analyses of plasma samples that aimed to identify CVD risk markers (62).

Trimethylamine as an Exogenous Metabolite

Although TMA has been considered a normal constituent of human urine since the beginning of the past century, the microbial origin of the metabolite was not conclusively established until 1992 (2). Since then, the endogenous nature of TMA has been confirmed through a series of human trials (25, 48). Major dietary sources of microbial TMA production are thought to be choline and betaine. Phosphatidylcholines (PCs), common in a large variety of foods including red meat, eggs, soybeans, and peanuts, have been identified as a major source of choline, an essential nutrient in omnivores. A resource of interest for large-scale epidemiological studies linking choline consumption and microbiome fluctuations to CVD risk is the US-targeted database for the content of choline-containing metabolites in common foods (**http://www.ars.usda.gov/Services/docs.htm?docid=6232**). Interestingly, in a recent study colonization of gnotobiotic mice with choline-converting bacteria not only increased cecal TMA production but also lowered choline serum concentrations (42). This suggests that the microbial TMA production potential should be taken into account when formulating dietary choline intake recommendations for humans. Besides choline metabolites, L-carnitine, which is abundant in red meat, has been identified as a substrate for TMA-producing microorganisms (25). In contrast to choline, L-carnitine is not an essential nutrient because it can be endogenously produced in mammals from dietary lysine. Although evidence linking TMA and TMAO to atherosclerosis is accumulating, the debate concerning the nutritional implications is ongoing (60).

Trimethylamine Removal from the Colon Environment

Once TMA is produced, there are at least three nonexclusive mechanisms through which it can be removed from the colon environment: through host absorption into systemic circulation, through egestion, and through microbial cross-feeding.

Once absorbed into systemic circulation, TMA can be excreted through breath, sweat, and urine. Such unmodified excretion is particularly important in individuals suffering from trimethylaminuria or fish malodor syndrome, a metabolic disorder caused by mutations in the human flavin-containing monooxygenase isoform 3 gene (*FMO3*), which impairs N-oxygenation of TMA (56). In healthy individuals, however, TMA is efficiently converted to TMAO by hepatic FMO3 and is subsequently excreted in a 95:3 TMAO:TMA ratio (67). Of interest for studies concerning TMA production and detoxification in larger populations is the sexual dimorphism of hepatic *FMO3* expression that was observed in mice and was reflected in consistently lower TMA conversion levels in male mice when they were exposed to dietary choline supplementation (4). Variation in FMO3 production levels was mainly a result of androgenic hormonal downregulation, but estrogen was also identified as a minor promoter of *FMO3* expression. Although it is currently unclear whether sexual dimorphism in *FMO3* expression levels occurs in humans, the latter observation coincides with reports of menstruation-associated trimethylaminuria (47). The question of whether sexual dimorphism is also present in microbiota TMA production potential remains unanswered, as evidence for gender-associated microbiome variation is limited (53). Another noteworthy observation is that in mice, *FMO3* expression can be induced through supplementation of diet with the bile component cholic acid through activation of the nuclear hormone farnesoid X receptor (4). This induction reveals sexual dimorphism and was shown to be up to ten times stronger in male mice than in female mice. Although this finding requires confirmation in a human host, it could imply an effect of dietary fat content on TMA metabolism that would enhance potential TMAO production resulting from a precursor-rich diet (41).

Although egestion of TMA has been reported (11), to our knowledge, no quantitative assessments linking dietary TMA-precursor content to concentrations found in feces or flatus have been performed. TMA's potential role as an intermediate in cross-feeding interactions within the colon ecosystem's trophic chain also remains largely unexplored. Notably, TMA has been shown to be a substrate for methanogenesis both in the gastrointestinal tract of ruminants (33) and in marine sediments (16, 23). A screening of all 531 bacterial and archaeal genomes [downloaded from RefSeq (52) on March 12, 2014] of human gastrointestinal origin (29) reveals that none of these carry the genetic potential for TMA conversion to methane (using MetaCYC pathway definitions) (7). However, using the restrictions applied (human, gastrointestinal tract), the database of scanned gut reference genomes contained only methanogens of the order *Methanobacteriales*, namely *Methanobrevibacter smithii* and *Methanosphaera stadtmanae*. Recently, archaeal diversity in the human gut has been shown to be far broader than these commonly encountered species and appears to increase with age (31). The creation of a novel new order of methanogens, *Methanoplasmatales*, has been proposed (30, 35), which would include colon isolates belonging to the genera *Methanomassiliicoccus* (12, 17) and *Methanomethylophilus* (5). Because some of these strains have been shown to be able to use TMA as a substrate for methanogenesis, a probiotic application—as an archaebiotic—to reduce blood TMAO levels has been suggested (6).

Trimethylamine *N*-Oxide Promotes Atherosclerosis

In their original, hallmark study, Wang et al. (62) examined small-molecule plasma metabolite profiles of patients who underwent elective cardiac evaluation and subsequently experienced

myocardial infarction, stroke, or death over the ensuing three-year period in comparison with matched controls. A two-stage approach resulted in a short list of 18 metabolites that predict increased CVD risk, including the PC degradation products choline, betaine, and TMAO. Their fasting plasma levels were assessed in an independent $N = 1{,}876$ cohort, confirming a dose-dependent association with CVD prevalence. Through a series of follow-up experiments, the repertoire of potentially prognostic metabolites for CVD development was expanded with L-carnitine (25). Although elevated L-carnitine plasma levels were shown to be indicative of CVD prevalence in an independent $N = 2{,}595$ cohort (25), large-scale three-year follow-up studies of individuals undergoing elective coronary angiography revealed that the prognostic value of L-carnitine, choline, or betaine for cardiovascular events could be traced to accompanying elevated plasma TMAO levels (25, 48, 63). This was confirmed in a five-year follow-up study of individuals with a history of heart failure, demonstrating that after adjustment for traditional CVD risk factors and cardiorenal indexes, elevated TMAO levels remained predictive for mortality risk over a longer period of time (50).

TMAO has been proposed to affect atherosclerosis development through several mechanisms. In atherosclerosis-prone mice, TMAO has been shown to alter peritoneal macrophage phenotype. Relative to normal chow diet, macrophages of mice supplemented with TMAO or dietary precursors exhibited enhanced expression of CD36 and SR-A, two scavenger receptors implicated in the development of atherosclerosis (62). Moreover, dietary choline–induced, microbiota-dependent TMAO production was linked to the development of macrophage foam cells as well as to the occurrence of aortic root lesions. Through a series of follow-up experiments, the same research group took a closer look at the mechanisms that would lead to cholesterol accumulation within cells of the artery wall. When assessing cholesterol removal in peripheral macrophages in atherosclerosis-prone mice, reduced reversed cholesterol transport was observed in animals fed a TMAO-supplemented diet (25).

MICROBIAL METABOLIC PATHWAYS LEADING TO COLON TRIMETHYLAMINE PRODUCTION

Although bacterial conversion of, for example, choline was reported as early as 1910, the specific metabolic pathways generating TMA through degradation of dietary substrates remained uncharted for almost 100 years. Only recently, following up on the regained interest in colon TMA production, three microbial pathways were characterized, all through a combinatorial bioinformatic, genetic, and biochemical approach.

Based on mechanistic similarities to the well-studied catabolism of ethanolamine, a gene cluster responsible for anaerobic choline degradation was identified in sulfate-reducing *Desulfovibrio* (9, 54). Choline-TMA-lyase activity encoded by part of this cluster was confirmed through creation of a genetic knockout in a choline-degrading organism and through heterologous expression of the genes in a non-choline-degrading strain. Crucial genes in the cluster are *CutC* and *CutD*, encoding for choline-TMA-lyase and its activating protein, respectively.

A second microbial metabolic pathway generating TMA through L-carnitine hydroxylation was shown to be active in human isolates belonging to the genera *Acinetobacter* (45) and *Serratia* (59). This pathway, yielding TMA and malic semialdehyde, was recently characterized using a sequential approach similar to the one described above (68). Human Microbiome Project reference genomes were screened for clusters encoding for both malate and succinate production—linking L-carnitine fermentation to the central tricarboxylic acid cycle—and for L-carnitine transporters. Inspection of such clusters in *Acinetobacter* spp. led to the identification of two genes, *CntA* and *CntB*, which were demonstrated to be crucial to L-carnitine conversion through a series of knockout and heterologous

expression experiments. Carnitine monoxygenase, the key enzyme in L-carnitine conversion encoded by *CntA/B*, was identified as a two-component Rieske-type oxygenase/reductase.

Finally, in mice, a third gut microbial pathway leading to intestinal TMA production was described by Koeth et al. (24). The authors identified γ-butyrobetaine (γBB) as the dominant intermediate of L-carnitine gut microbial catabolism. Although γBB is further converted into TMA, both anatomical and species compartmentalization of the subsequent degradation processes was observed. L-carnitine degradation to γBB is initialized in the mouse ileum, whereas TMA production—the result of γBB as well as the L-carnitine conversion described above—appears restricted to the cecum and the colon. Screening for unknown enzymes in gene clusters within a reference genome while covering enzymes for both malate and succinate synthesis as well as presumed L-carnitine/betaine/choline transporters allowed the authors to identify another *YeaW/X* gene pair encoding for oxygenase and oxidoreductase activity, respectively. Purified recombinant proteins displayed L-carnitine-TMA-lyase activity with complex substrate promiscuity (γBB, L-carnitine, choline, and betaine).

Overall, it should be stressed that the elucidation of microbial TMA-generating pathways is an ongoing process. Besides microbial conversion of the dietary precursors discussed above (phosphatidylcholine, choline, and L-carnitine), plenty of other food ingredients possess a TMA moiety [including betaine, sphingomyelin, phosphocholine, and glycerophosphocholine (65)], which suggests that a substantial part of TMA-producing microbial metabolism remains uncharted territory (49).

MICROORGANISMS ASSOCIATED WITH TRIMETHYLAMINE PRODUCTION

Given the predominantly metabolomics/host focus of current TMAO/atherosclerosis research, only limited efforts have been undertaken to assess the microbial background of colon TMA production. Moreover, because the genetic basis of at least some of the metabolic pathways generating TMA has only recently been elucidated, the few attempts to correlate microbiome composition to plasma TMAO levels were unavoidably restricted to 16S rDNA amplicon sequencing studies.

To assess the impact of an L-carnitine-enriched diet on microbiota composition, Koeth et al. (25) examined the cecal microbiome of mice fed normal ($N = 10$) versus L-carnitine-supplemented ($N = 11$) chow. As expected, supplementation of the diet with L-carnitine increased blood TMA/TMAO concentrations. The microbiota composition of mice fed an L-carnitine-supplemented diet was characterized by increased abundances of *Anaeroplasma*, *Prevotella*, and *Mucispirillum*, the last of which is a mouse-specific isolate (26). During a series of follow-up experiments, Koeth et al. (24) analyzed the microbiota of mice fed normal, L-carnitine-, or γBB-supplemented chow. Species associated with L-carnitine-to-γBB conversion included jejunal *Staphylococcus* spp. as well as cecal bacteria belonging to the genera *Parasutterella*, *Prevotella*, and *Bacteroides*. Remarkably, the genus *Akkermansia* was shown to be linked to increased TMAO production in mice fed a γBB-supplemented diet. In mice, *Akkermansia muciniphila* treatment has been shown to reverse high-fat diet–induced metabolic disorders, including fat-mass gain, metabolic endotoxemia, adipose tissue inflammation, and insulin resistance (14). The species is also found in humans, where it is generally associated with a healthier phenotype (27). A recent in vitro study demonstrated the species does not convert choline to TMA (42).

Following up on the observation that human vegans and vegetarians tend to produce less TMAO during L-carnitine challenge tests, Koeth et al. (25) analyzed the microbiota composition in 23 vegetarians/vegans and 30 omnivores using 16S rDNA amplicon sequencing. Amplicon

profile analysis revealed that although the genus *Clostridium* was associated with an omnivorous diet, both *Lachnospira* and *Sporobacter* spp. were more abundant in vegetarians/vegans. However, within each dietary group, *Clostridium* and *Sporobacter* abundances correlated negatively and positively with plasma TMAO concentrations, respectively. This observation indicates that although both genera are clearly associated with long-term dietary habits, their abundance cannot be considered indicative for the microbiota TMA production potential. In contrast, *Lachnospira* spp. are not only more abundant in feces of vegetarians and vegans but also anticorrelate with plasma TMAO concentrations independent of diet. Taxa associated with TMAO concentration did not correspond with those identified in mice fed L-carnitine-supplemented chow (see above), confirming the limited overlap in microbiota composition between animal models and the human host (34).

Enterotypes have been described as densely populated regions of microbiota configurations in gut microbiome space (3). They have been linked to long-term dietary patterns (64) and/or inflammation (19). In the original publication, three enterotypes are identified and are referred to as *Prevotella* (P), *Bacteroides* (B), and *Ruminococcus* (R) based on species distribution. Recent studies have shown the existence of intermediate states, and thus the discussion of whether enterotypes are discrete or optima in a continuum is ongoing (57). However, enterotype stratification remains a powerful tool to assess global diversification in a microbiome data set. Based on plasma TMAO levels, enterotyping of a 53-sample data set indicated enhanced TMA production in the *Prevotella* subgroup (25). Although this observation has value in establishing a clear link between enterotypes and microbiotal TMA production, care needs to be taken when extrapolating or generalizing results: In this study, only four of the samples analyzed were identified as *Prevotella*, whereas the *Ruminococcus* enterotype was not detected at all.

TRIMETHYLAMINE PRODUCTION POTENTIAL IN REFERENCE GENOMES

Associations of bacterial taxa with plasma TMAO levels do not necessarily imply a causal link between the genera identified and the potentially enhanced production of TMA in the colon ecosystem. In fact, such associations can also reflect indirect changes in microbiota composition due to (*a*) bacterial interactions, (*b*) shifts caused by rather than causing changes in host metabolic status, or (*c*) the impact of independent variables that are inherent to or overlooked in study design. A more straightforward way of identifying potential colon TMA producers is through screening reference genome libraries for species carrying the genetic potential to perform the task under study. Such screens have been performed for individual pathways generating TMA. Here, we use the pathways identified in the above-mentioned studies for genome mining in public data (29, 52) to provide a general overview of potential TMA-producing bacteria in sequenced human gastrointestinal isolates.

The first step in establishing the presence and abundance of the pathway of interest in gut genomic or metagenomic data sets is to retrieve the orthologous group (51) for each essential component (e.g., enzymes or transporters) of the pathway. Such components might be included in orthology databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) (21), that are based on experimentally validated gene functions and are expanded by sequence similarity, or identified through TIGRFAMS (18), which are manually curated sets of strictly defined protein families. Alternatively, a new orthologous group can be created by clustering reciprocal best hits using the Basic Local Alignment Search Tool (BLAST) (1, 36), i.e., hits from gut reference genomes (29, 52). To reconstruct TMA production pathways, *CutC* and *CutD* TIGRFAMS profiles (TIGR04394 and TIGR04395) are available for the choline pathway, the γBB pathway genes *YeaW* and *YeaX* can be mapped by KEGG Orthology (KO)-annotation to K00470 and K00540,

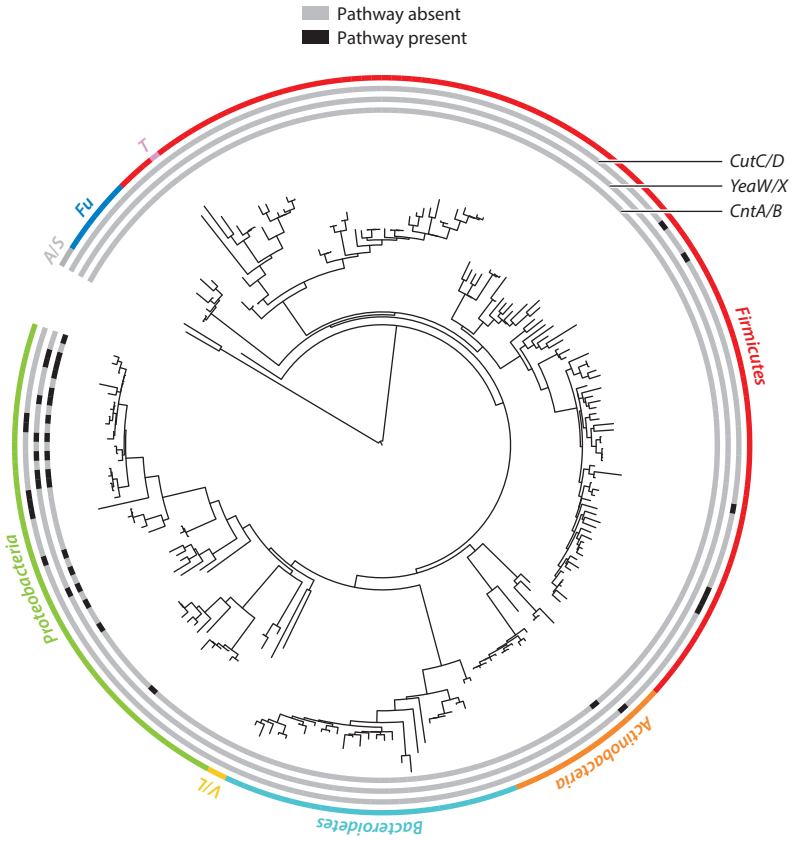

Figure 1

Human gut microbiota reference species with the potential to produce trimethylamine (TMA). Phylogenetic tree (16S rDNA) of the human gut–associated bacteria and archaea and their TMA production potential, defined by the presence of all required enzymes in the fully sequenced genome. Three alternative metabolic pathways were reviewed and are shown here: *CutC/D* (9), *YeaW/X* (24), and *CntA/B* (68). Abbreviations: *A*, *Archaea*; *Fu*, *Fusobacterium*; *L*, *Lentisphaerae*; *S*, *Synergistetes*; *T*, *Tenericutes*; *V*, *Verrucomicrobia*.

and a list of orthologs for *CntA* of the carnitine pathway is provided in the original publication (68) (*CntB* matches K00540). The oxidoreductase K00540 (*CntB/YeaX*) was excluded from the pathways used here for genome mining because of its broad definition: It maps to at least one gene in 80% of gut genera and 90% of gut phyla, whereas the ortholog groups for the other five enzymes map to 11% of gut genera and 34% of gut phyla.

Overall, TMA production potential is present in 102 reference genomes covering 36 species (**Figure 1**). Although conversion potential can be detected in *Firmicutes*, *Proteobacteria*, and *Actinobacteria*, it appears to be absent in *Bacteroidetes*, which confirms and expands the previous observations concerning the uneven distribution of the choline degradation pathway (*CutC/D*) across common gut phyla (9). Remarkably, all species encoding the *YeaW/X* pathway L-carnitine degradation are also identified as *CntA/B* positive, possibly reflecting the highly similar methodology applied for pathway discovery (**Table 1**). Other than *Clostridium* spp., which were previously associated with an L-carnitine-rich diet, no overlap with results from previously mentioned association

Table 1 Trimethylamine production potential in gut reference species

Phylum	Species	*CutC/D* (9)	*YeaW/X* (24)	*CntA/B* (68)
Actinobacteria	*Collinsella tanakaei*	+	–	–
	Corynebacterium ammoniagenes	–	–	+
Firmicutes	*Anaerococcus hydrogenalis*	+	–	–
	Bacillus smithii	–	–	+
	Bacillus sp.	–	–	+
	Clostridiales sp.	+	–	–
	Clostridium asparagiforme	+	–	–
	Clostridium citroniae	+	–	–
	Clostridium hathewayi	+	–	–
	Clostridium sporogenes	+	–	–
	Paenibacillus sp.	–	–	+
Proteobacteria	*Acinetobacter baumannii*	–	+	+
	Acinetobacter junii	–	–	+
	Arcobacter butzleri	–	–	+
	Burkholderia oklahomensis	–	–	+
	Citrobacter freundii	–	+	+
	Citrobacter sp.	–	+	+
	Citrobacter youngae	–	–	+
	Enterobacter cancerogenus	–	–	+
	Enterobacter cloacae	–	–	+
	Escherichia albertii	–	+	+
	Escherichia coli	+	+	+
	Escherichia sp.	–	+	+
	Klebsiella sp.	+	–	+
	Neisseria macacae	–	–	+
	Proteus penneri	+	–	–
	Providencia alcalifaciens	+	–	–
	Providencia rettgeri	–	+	+
	Providencia rustigianii	+	–	–
	Providencia stuartii	–	+	+
	Pseudomonas aeruginosa	–	–	+
	Pseudomonas sp.	–	–	+
	Ralstonia sp.	–	–	+
	Shigella sp.	–	+	+
	Vibrio furnissii	+	–	+
	Yersinia pseudotuberculosis	–	+	+

The + symbol indicates that the gene is present in at least one published genome of the species (Integrated Microbial Genomes Database, IMG v4.00; 29, 52). The – symbol indicates that the gene is absent from all of the genomes of the species (IMG v4.00; 29, 52).

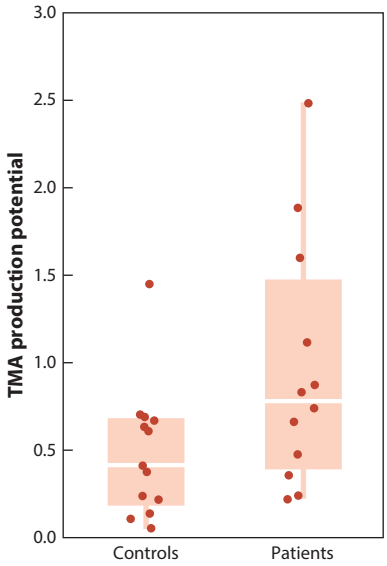

Figure 2

Microbiota trimethylamine (TMA) production potential and its association with symptomatic atherosclerosis (22). Patients ($N = 12$) had significantly greater TMA production potential than controls ($N = 13$) (Wilcoxon p-value $= 0.036$).

studies can be observed, stressing the limitations of such an approach to identify novel organisms involved in TMA production.

Only recently, Romano et al. (42) screened 79 human intestinal isolates for their capacity to produce TMA from choline in vitro. Of 8 strains identified as TMA producing, 7 also figure among the species listed in **Table 1**, demonstrating the power of the genome-based screening presented. Only 1 isolate (*Edwardsiella tarda* ATCC 23685) was shown to produce TMA in vitro without displaying the genetic potential required. Although this finding could reflect the current lack of knowledge concerning the genetic background of bacterial TMA production, it might also, as suggested by the authors, reflect a strain-specific capacity for choline conversion and hint at acquisition through lateral gene transfer.

Overall, the compilation of a catalog of potential TMA-generating colon bacteria as presented here might invite an assessment of ecosystem production capacity through an analysis of species composition based on 16S rDNA amplicon sequencing. Although this strategy has clear advantages, as it reduces both cost and analytical complexity, care needs to be taken when interpreting results because of the variation in phylogenetic inheritability of metabolic traits that can be observed within bacterial genera (32, 55).

METAGENOMIC CHARACTERIZATION OF ATHEROSCLEROSIS-ASSOCIATED DYSBIOSIS

Although large-scale projects aiming to characterize the impact of qualitative and quantitative changes in the gut microbiota on the pathogenesis of cardiometabolic diseases and their associated comorbidities are ongoing (e.g., **http://www.metacardis.eu**), to date, only one metagenomics study assessing potential microbiota dysbiosis associated with atherosclerosis has been performed.

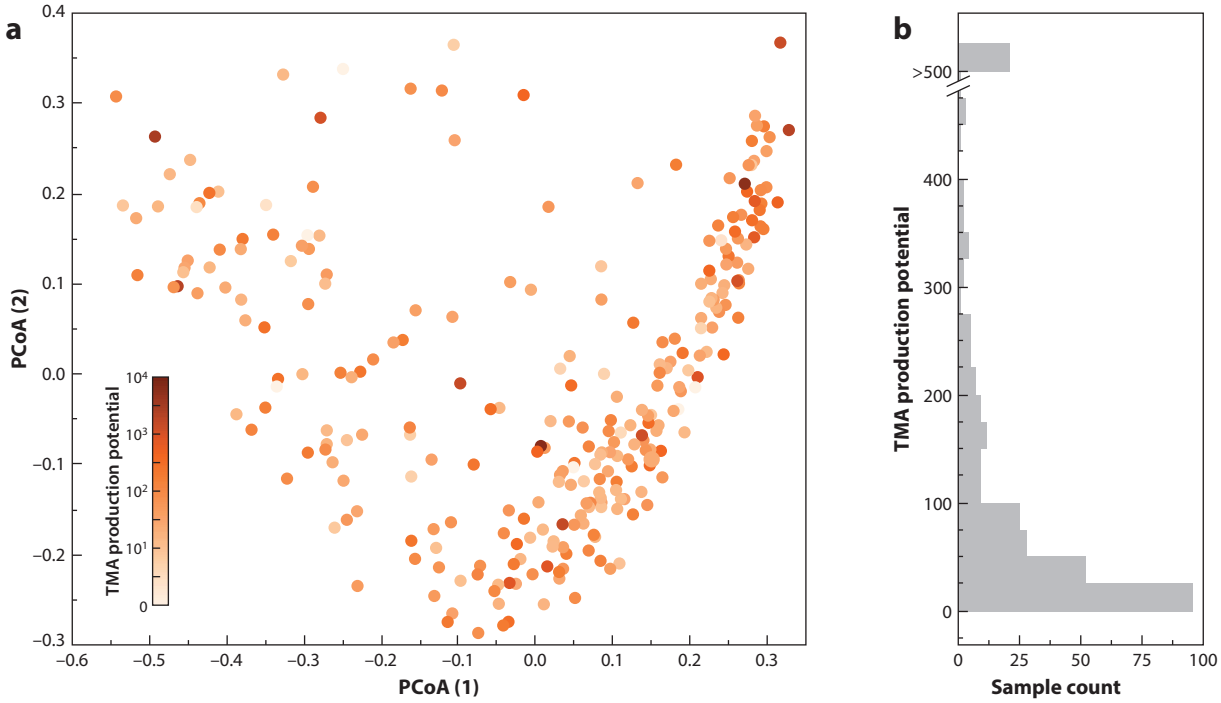

Figure 3

Trimethylamine (TMA) production potential in MetaHIT human gut microbiomes (27). (*a*) Human gut microbiota compositional diversity is represented as the principal coordinate analysis (PCoA) of the genus-level microbiome composition of the 277 individuals in the MetaHIT data set (Bray-Curtis dissimilarity). Data points are colored according to microbiome TMA production potential (sum of pathway abundances *CutC/D*, *YeaX/Y*, and *CntA/B*). (*b*) Distribution of TMA production potential across the 277 MetaHIT samples.

In a small-scale cohort consisting of 12 elderly individuals who had undergone carotid endarterectomy for minor ischemic stroke, transient ischemic attack, or amaurosis fugax and 13 gender- and age-matched controls, Karlsson et al. (22) observed an increased abundance of *Collinsella* in patient samples, whereas control subjects were enriched in *Eubacterium*, *Roseburia*, and *Bacteroides* spp.. Functional characterization revealed that atherosclerosis patients' microbiomes were enriched in genes encoding peptidoglycan biosynthesis, whereas control samples had a higher genetic potential for phytoene dehydrogenase, which is involved in the metabolism of lipid-soluble antioxidants such as the carotenoids. Accordingly, increased levels of β-carotene were detected in the serum of control individuals. Enterotype stratification (3) of fecal microbiomes revealed that patients were over- and underrepresented in the *Ruminococcus* and *Bacteroides* enterotype, respectively. To investigate a possible link between symptomatic atherosclerosis and colon TMA production potential, Karlsson et al. (22) reconstructed a metabolic pathway covering the microbial conversion of phosphatidylcholine. No significant association between KO annotated pathway gene abundances and atherosclerosis could be established. However, at the time of metabolic module construction, the genetic background of microbial TMA production remained uncharacterized.

By mapping this data set to the latest reference gene catalog (**http://www.ebi.ac.uk/ena/data/view/ERP003612;** 28) and mining out the orthologs involved in TMA production, we performed a targeted, hypothesis-driven rescreening of the Karlsson data set (**http://sra.dnanexus.com/studies/SRP016067**). Assessment of the abundance variation of a custom-generated

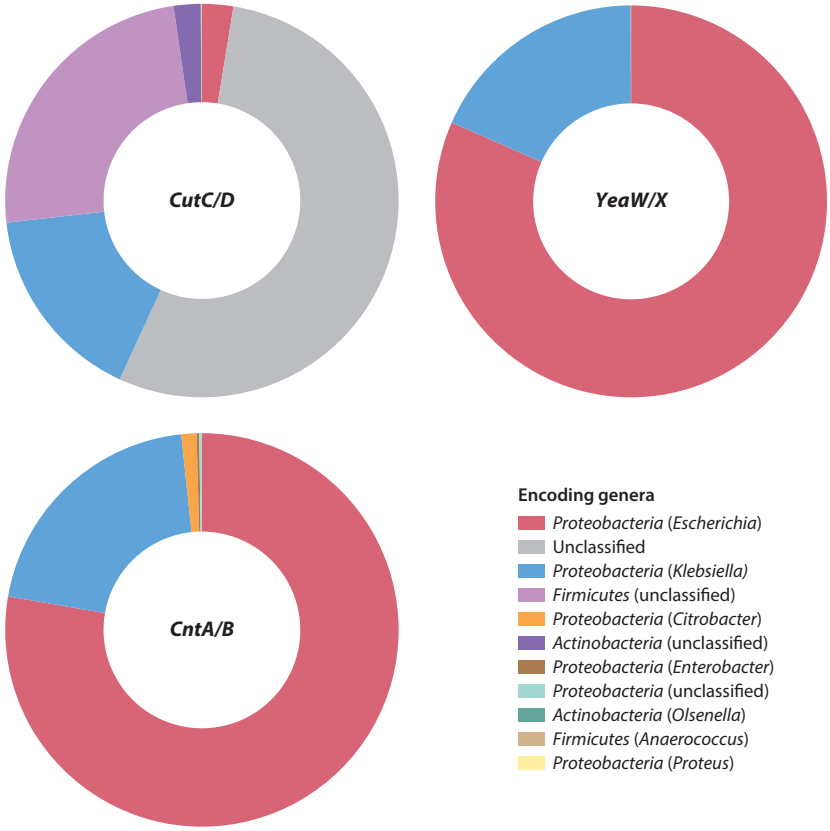

Figure 4

Bacteria carrying the genes for trimethylamine (TMA) production in the MetaHIT data set. Ring charts display the proportion of the pathways' total abundance, which is assigned to each bacterial genus (gene phylogenetic assignment based on Reference 28) for three TMA production pathways: *CutC/D* (9), *YeaW/X* (24), and *CntA/B* (68).

metabolic module set based on the three novel TMA-generating pathways described above does reveal enhanced total TMA production capacity in symptomatic atherosclerosis patients (**Figure 2**; p-value $= 0.036$).

BOTTOM-UP ASSESSMENT OF TRIMETHYLAMINE PRODUCTION POTENTIAL IN A METAGENOMICS DATA SET

The MetaHIT consortium (27) generated an impressive library of human fecal microbiomes and corresponding metadata. A large part of the consortium's sequencing effort was directed toward the assessment of obesity-associated dysbiosis in a cohort of 292 Danish volunteers. One of the main findings was the identification of two subpopulations within the data set with different microbiome gene counts and therefore different bacterial richness. Individuals with a lowered bacterial richness exhibited increased overall adiposity, insulin resistance, and dyslipidemia, and they displayed a more pronounced inflammatory phenotype.

Figure 3 shows the distribution of TMA production potential based on summed *CutC/D*, *YeaX/Y*, and *CntA/B* pathway abundances in MetaHIT samples. The figure, a classic visualization in metagenomic research, consists of a principal coordinate analysis (PCoA) of microbiome composition variation in the data set under study. Distances between data points (samples) are based on Bray-Curtis dissimilarity of genus abundance profiles. In metagenomics, PCoAs are often used to visualize microbiome diversity in data sets as an initial step in explorative analyses, allowing subsequent assessment of microbiome clusters or metadata features. Here, data points are colored according to TMA production potential.

Through linkage with the gene catalog, microbiome data sets can be used for phylogenetic characterization of detected metabolic pathways (**Figure 4**). Tracing metabolism back to microorganisms can provide inspiration for further isolation and subsequent phenotypic, genetic, and metabolic characterization of bacterial species. In this case, with the exception of *Olsenella*, most TMA-generating genera identified in the MetaHIT samples confirm those picked up in reference genomes. Hence, through our demonstration of how species-function relationships can also be retrieved from metagenomics data sets, we close the circle, translating microbiology to big data—and back again.

CONCLUSIONS

Given the technical challenges associated with handling and analyzing big data, early-day metagenomics have developed as a bioinformatics-driven branch of microbial research. In addition to these technical hurdles, the tendency to outsource actual sequencing efforts to core facilities or sequencing centers has created a gap between microbiology and its metagenomic application. Bridging this gap is one of the major challenges for microbiology in the years to come, and it will be facilitated by the advent of desktop sequencers. As various user-friendly tools and databases to analyze big data are becoming available, integrating exploration of metagenomic libraries into common lab workflows is feasible. As we demonstrated throughout this article on metabolic pathway elucidation, doing so will allow immediate assessment of the ecological relevance of novel findings, and the tools and modules developed in the process can be integrated into metabolic databases used for routine screening of metagenomes, thus increasing their effectiveness. Embracing big data as an integrative part of any microbial research program will create a synergy that has the potential to revolutionize the field once again.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This research was funded by FP7 METACARDIS HEALTH-F4-2012-305312 and IWT-SBO 100016. S.V.S. is funded by Marie Curie Actions FP7 People COFUND-Proposal 267139 (acronym OMICS@VIB). The Raes lab is supported by the Research Foundation Flanders (FWO), the Flemish agency for Innovation by Science and Technology (IWT), KU Leuven, and the Rega Institute. We thank Fredrik Karlsson and Fredrik Bäckhed for providing the metadata from their study on microbiome variation in symptomatic atherosclerosis (22).

LITERATURE CITED

1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215(3):403–10
2. al-Waiz M, Mikov M, Mitchell SC, Smith RL. 1992. The exogenous origin of trimethylamine in the mouse. *Metabolism* 41(2):135–36
3. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, et al. 2011. Enterotypes of the human gut microbiome. *Nature* 473(7346):174–80
4. Bennett BJ, de Aguiar Vallim TQ, Wang Z, Shih DM, Meng Y, et al. 2013. Trimethylamine-N-oxide, a metabolite associated with atherosclerosis, exhibits complex genetic and dietary regulation. *Cell Metab.* 17(1):49–60
5. Borrel G, Harris HMB, Tottey W, Mihajlovski A, Parisot N, et al. 2012. Genome sequence of "*Candidatus Methanomethylophilus alvus*" Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J. Bacteriol.* 194(24):6944–45
6. Brugère JF, Borrel G, Gaci N, Tottey W, O'Toole PW, Malpuech-Brugère C. 2014. Archaebiotics. *Gut Microb.* 5(1):5–10
7. Caspi R, Altman T, Billington R, Dreher K, Foerster H, et al. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 42:D459–71
8. Clemente JC, Ursell LK, Parfrey LW, Knight R. 2012. The impact of the gut microbiota on human health: an integrative view. *Cell* 148(6):1258–70
9. Craciun S, Balskus EP. 2012. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *PNAS* 109(52):21307–12
10. Cryan JF, Dinan TG. 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat. Rev. Neurosci.* 13(10):701–12
11. De Filippi F. 1906. Das Trimethylamin als normales Produkt des Stoffwechsels, nebst einer Methode für dessen Bestimmung im Harn und Kot. *Z. Physiol. Chem.* 49:433–56
12. Dridi B, Fardeau ML, Ollivier B, Raoult D, Drancourt M. 2012. *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 62(8):1902–7
13. Erickson AR, Cantarel BL, Lamendella R, Darzi Y, Mongodin EF, et al. 2012. Integrated metagenomics/metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLOS ONE* 7(11):e49138
14. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, et al. 2013. Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *PNAS* 110(22):9066–71
15. Faust K, Raes J. 2012. Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10(8):538–50
16. Fiebig K, Gottschalk G. 1983. Methanogenesis from choline by a coculture of *Desulfovibrio* sp. and *Methanosarcina barkeri*. *Appl. Environ. Microbiol.* 45(1):161–68
17. Gorlas A, Robert C, Gimenez G, Drancourt M, Raoult D. 2012. Complete genome sequence of *Methanomassiliicoccus luminyensis*, the largest genome of a human-associated archaea species. *J. Bacteriol.* 194(17):4745
18. Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2013. TIGRFAMS and genome properties in 2013. *Nucleic Acids Res.* 41:D387–95
19. Hildebrand F, Nguyen TLA, Brinkman B, Yunta RG, Cauwe B, et al. 2013. Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol.* 14(1):R4
20. Joossens M, Huys G, Cnockaert M, De Preter V, Verbeke K, et al. 2011. Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60(5):631–37
21. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42:D199–205
22. Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, et al. 2012. Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat. Commun.* 3:1245
23. King GM. 1984. Metabolism of trimethylamine, choline, and glycine betaine by sulfate-reducing and methanogenic bacteria in marine sediments. *Appl. Environ. Microbiol.* 48(4):719–25

24. Koeth RA, Levison BS, Culley MK, Buffa JA, Wang Z, et al. 2014. γ-Butyrobetaine is a proatherogenic intermediate in gut microbial metabolism of L-carnitine to TMAO. *Cell Metab.* 20(5):799–812
25. Koeth RA, Wang Z, Levison BS, Buffa JA, Org E, et al. 2013. Intestinal microbiota metabolism of L-carnitine, a nutrient in red meat, promotes atherosclerosis. *Nat. Med.* 19(5):576–85
26. Krych L, Hansen CHF, Hansen AK, van den Berg FWJ, Nielsen DS. 2013. Quantitatively different, yet qualitatively alike: a meta-analysis of the mouse core gut microbiome with a view towards the human gut microbiome. *PLOS ONE* 8(5):e62578
27. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, et al. 2013. Richness of human gut microbiome correlates with metabolic markers. *Nature* 500(7464):541–46
28. Li J, Jia H, Cai X, Zhong H, Feng Q, et al. 2014. An integrated catalog of reference genes in the human gut microbiome. *Nat. Biotechnol.* 32(8):834–41
29. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, et al. 2012. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40:D115–22
30. Mihajlovski A, Alric M, Brugère JF. 2008. A putative new order of methanogenic archaea inhabiting the human gut, as revealed by molecular analyses of the *mcrA* gene. *Res. Microbiol.* 159(7–8):516–21
31. Mihajlovski A, Doré J, Levenez F, Alric M, Brugère JF. 2010. Molecular evaluation of the human gut methanogenic archaeal microbiota reveals an age-associated increase of the diversity. *Environ. Microbiol. Rep.* 2(2):272–80
32. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD. 2013. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *PNAS* 110(50):20338–43
33. Neill AR, Grime DW, Dawson RM. 1978. Conversion of choline methyl groups through trimethylamine into methane in the rumen. *Biochem. J.* 170(3):529–35
34. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. 2015. How informative is the mouse for human gut microbiota research? *Dis. Model. Mech.* 8(1):1–16
35. Paul K, Nonoh JO, Mikulski L, Brune A. 2012. "Methanoplasmatales": *Thermoplasmatales*-related archaea in termite guts and other environments are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78(23):8245–53
36. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, et al. 2012. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res.* 40:D284–89
37. Qin J, Li Y, Cai Z, Li S, Zhu J, et al. 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490(7418):55–60
38. Raes J. 2011. Why meta-omics should be mega-omics: on experimental design and multiple testing hell. *Environ. Microbiol. Rep.* 3(1):19–20
39. Raes J, Bork P. 2008. Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 6(9):693–99
40. Rajilić-Stojanović M, Jonkers DM, Salonen A, Hanevik K, Raes J, et al. 2015. Intestinal microbiota and diet in IBS: causes, consequences or epiphenomena. *Am. J. Gastroenterol.* 110(2):278–87
41. Reddy BS. 1981. Diet and excretion of bile acids. *Cancer Res.* 41:3766–68
42. Romano KA, Vivas EI, Amador-Noguez D, Rey FE. 2015. Intestinal microbiota composition modulates choline bioavailability from diet and accumulation of the proatherogenic metabolite trimethylamine-*N*-oxide. *mBio* 6(2):e02481–14
43. Savage DC. 1977. Microbial ecology of the gastrointestinal tract. *Annu. Rev. Microbiol.* 31(70):107–33
44. Scher JU, Sczesnak A, Longman RS, Segata N, Ubeda C, et al. 2013. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* 2:e01202
45. Seim H, Löster H, Claus R, Kleber HP, Strack E. 1982. Splitting of the C–N bond in carnitine by an enzyme (trimethylamine forming) from membranes of *Acinetobacter calcoaceticus*. *FEMS Microbiol. Lett.* 15(3):165–67
46. Serino M, Blasco-Baque V, Nicolas S, Burcelin R. 2014. Far from the eyes, close to the heart: dysbiosis of gut microbiota and cardiovascular consequences. *Curr. Cardiol. Rep.* 16(11):540
47. Shimizu M, Cashman JR, Yamazaki H. 2007. Transient trimethylaminuria related to menstruation. *BMC Med. Genet.* 8:2

48. Tang WH, Wang Z, Levison BS, Koeth RA, Britt EB, et al. 2013. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* 368(17):1575–84
49. Tang WH, Hazen SL. 2014. The contributory role of gut microbiota in cardiovascular disease. *J. Clin. Investig.* 124(10):4204–11
50. Tang WH, Wang Z, Fan Y, Levison B, Hazen JE, et al. 2014. Prognostic value of elevated levels of intestinal microbe-generated metabolite trimethylamine-*N*-oxide in patients with heart failure: refining the gut hypothesis. *J. Am. Coll. Cardiol.* 64(18):1908–14
51. Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278(5338):631–37
52. Tatusova T, Ciufo S, Fedorov B, O'Neill K, Tolstoy I. 2014. RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* 42:D553–59
53. The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–14
54. Thibodeaux CJ, van der Donk WA. 2012. Converging on a mechanism for choline degradation. *PNAS* 109(52):21184–85
55. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, et al. 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLOS Genet.* 5(1):e1000344
56. Treacy EP, Akerman BR, Chow LML, Youil R, Bibeau C, et al. 1998. Mutations of the flavin-containing monooxygenase gene (*FMO3*) cause trimethylaminuria, a defect in detoxication. *Hum. Mol. Genet.* 7(5):839–45
57. Tremaroli V, Bäckhed F. 2012. Functional interactions between the gut microbiota and host metabolism. *Nature* 489(7415):242–49
58. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, et al. 2009. A core gut microbiome in obese and lean twins. *Nature* 457(7228):480–84
59. Unemoto T, Hayashi M, Miyaki K, Hayashi M. 1966. Formation of trimethylamine from DL-carnitine by *Serratia marcescens*. *Biochim. Biophys. Acta* 121:220–22
60. Ussher JR, Lopaschuk GD, Arduini A. 2013. Gut microbiota metabolism of L-carnitine and cardiovascular risk. *Atherosclerosis* 231(2):456–61
61. Vinjé S, Stroes E, Nieuwdorp M, Hazen SL. 2014. The gut microbiome as novel cardio-metabolic target: The time has come! *Eur. Heart J.* 35(14):883–87
62. Wang Z, Klipfell E, Bennett BJ, Koeth R, Levison BS, et al. 2011. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* 472(7341):57–63
63. Wang Z, Tang WH, Buffa JA, Fu X, Britt EB, et al. 2014. Prognostic value of choline and betaine depends on intestinal microbiota-generated metabolite trimethylamine-*N*-oxide. *Eur. Heart J.* 35(14):904–10
64. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, et al. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334(6052):105–8
65. Zeisel SH, Mar M, Howe JC, Holden JM. 2003. Concentrations of choline-containing compounds and betaine in common foods. *J. Nutr.* 133(5):1302–7
66. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, et al. 2014. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* 10:766
67. Zhang AQ, Mitchell S, Smith R. 1995. Fish odour syndrome: verification of carrier detection test. *J. Inherit. Metab. Dis.* 18(6):669–74
68. Zhu Y, Jameson E, Crosatti M, Schäfer H, Rajakumar K, et al. 2014. Carnitine metabolism to trimethylamine by an unusual Rieske-type oxygenase from human microbiota. *PNAS* 111(11):4268–73