Next Generation Microbiome Research: Identification of Keystone Species in the Metabolic Regulation of Host-Gut Microbiota Interplay

Héloïse Tudela¹², Sandrine P. Claus¹* and Maya Saleh²³*

¹YSOPIA Bioscience, Bordeaux, France, ²ImmunoConcEpT, CNRS UMR 5164, University of Bordeaux, Bordeaux, France, ³Department of Medicine, McGill University, Montreal, QC, Canada

The community of the diverse microorganisms residing in the gastrointestinal tract, known as the gut microbiota, is exceedingly being studied for its impact on health and disease. This community plays a major role in nutrient metabolism, maintenance of the intestinal epithelial barrier but also in local and systemic immunomodulation. A dysbiosis of the gut microbiota, characterized by an unbalanced microbial ecology, often leads to a loss of essential functions that may be associated with proinflammatory conditions. Specifically, some key microbes that are depleted in dysbiotic ecosystems, called keystone species, carry unique functions that are essential for the balance of the microbiota. In this review, we discuss current understanding of reported keystone species and their proposed functions in health. We also elaborate on current and future bioinformatics tools needed to identify missing functions in the gut carried by keystone species. We propose that the identification of such keystone species functions is a major step for the understanding of microbiome dynamics in disease and toward the development of microbiome-based therapeutics.

Keywords: microbiome, dysbiosis, keystone, metagenomics, bioinformatics, inflammation, immunity, metabolism

INTRODUCTION

Animals are superorganisms composed of eukaryotic and prokaryotic cells in a similar proportion along an even larger number of viruses (Sender et al., 2016). The reason for this intricate mixture of organisms spanning all kingdoms of life is that every living animal is the result of a long co-evolution between all of these organisms. Hence, within every gut lies a complex community of microorganisms composed of trillions of prokaryotic and eukaryotic microbial cells, including bacteria, fungi and archaea along a multitude of viruses (Hillman et al., 2017; Moissl-Eichinger et al., 2018). As a result, we humans carry within our microbiomes an immense reservoir of genes that perform numerous functions for our own benefit, many of which are still unknown.

There is currently no consensus about the definition of a healthy gut microbiome because of high inter-individual variability, which is influenced by numerous external factors (Rinninella et al., 2019). Nevertheless, it is generally considered that a healthy gut microbiome is a rich and diverse ecosystem acting in symbiosis with its host (van de Guchte et al., 2018). Even if there is a lack of evidence to identify a robust universal set of core healthy microbial taxa, there is a remarkable
stability of microbial functions that maintain symbiosis with the host (Human Microbiome Project Consortium, 2012). Conversely, we have learned over the past 20 years that some chronic disorders are consistently associated with a shift in microbial patterns, often referred to as “dysbiosis” (Hooks and O’Malley, 2017). For example, obesity has been reported to be associated with a low Bacteroidetes/Firmicutes ratio (Turnbaugh et al., 2009). Even if the Firmicutes phylum regroups a large number of potentially beneficial bacteria, this ratio has progressively been established as a hallmark of the obese dysbiotic gut microbiome (Crovesy et al., 2020). Another feature of obesity-associated dysbiosis is a reduced microbiome diversity, as illustrated by the high proportion of fecal samples from obese individuals that fall within the “low gene count” category (Le Chatelier et al., 2013). Similarly, several chronic diseases have been associated with reduced gut microbiome diversity, such as Crohn’s disease (Manichanh et al., 2006), hypertension (Li et al., 2017) and non-alcoholic steatohepatitis (NASH) (Astbury et al., 2020).

THE KEYSTONE SPECIES CONCEPT AS A DRIVER OF MICROBIAL DIVERSITY

An important ecological concept is that every complex ecosystem is structured by a few important species dubbed “keystones.” This term was coined in 1966 by the American ecologist Paine (1966) who identified specific sea stars as important predators that regulate the biodiversity of seashores. Since, this term has been used in various ways and with different meanings. For the purpose of this review, we adopt the definition of keystone taxa proposed by Banerjee et al. (2018): “where microbial keystone taxa are highly connected taxa that individually or in a guild exert a considerable influence on microbiome structure and functioning irrespective of their abundance across space and time. These taxa have a unique and crucial role in microbial communities, and their removal can cause a dramatic shift in microbiome structure and functioning” (Banerjee et al., 2018). This is a crucial concept as it shapes our understanding of the regulation of complex ecosystems, how they establish, how they remain stable over long periods of time and how they adapt to environmental changes.

Translated to the gut environment, we must first appreciate that mammals harbor not just one gut ecosystem, but a variety of ecosystems, each roughly corresponding to a different section of the gastrointestinal tract from the mouth to the rectum. Each ecosystem is regulated by a set of environmental factors such as pH, bile acid concentration and peristalsis, which have long been thought of as a barrier that segregates ecosystems from one another. However, this view has been recently challenged, as it is now proposed that oral bacteria act as a reservoir of microorganisms that pass through the gut to replenish the downstream ecosystems (Schmidt et al., 2019). Within each ecosystem, microbes interact with each other through numerous mechanisms, such as secretion of quorum sensing molecules, cross-feeding and synthesis of antimicrobial compounds. Of particular interest, quorum sensing is a cell–cell interaction mechanism used by bacteria to regulate their own population. Usually in biofilms, some bacterial cells stimulate their own growth and those of their neighboring kin, through secretion of autoinducer molecules (Mukherjee and Bassler, 2019). In the gut microbiota, it has been shown that Firmicutes use this strategy to maintain their population level (Thompson et al., 2015). Interestingly, there is emerging evidence that host cells can interfere with these bacterial signals to shape the microbial community (Mukherjee and Bassler, 2019). Yet, most of our knowledge of quorum sensing is derived from the study of pathogens and there remains numerous gaps in our understanding of its use by commensal bacteria. Inter-species syntrophy or cross-feeding occurs when a species depends on the availability of nutrients (e.g., sugars, amino acids, and vitamins) that are produced by other species. For instance, this typically involves degradation of complex molecules such as carbohydrates by specialized species that release monosaccharides in the environment. The latter are then taken up by non-degrading species for their own benefit. These mechanisms have been recently thoroughly reviewed by D’Souza et al. (2018). Inter-species cross-feeding interactions within an ecosystem cause reliance on specific microbes that carry essential functions for other species. Hence many keystone species have been described based on the identification of enzymes involved in cross-feeding interactions (Centanni et al., 2018; Table 1). These are only a few examples of the diversity of possible microbial interaction routes. For a thoughtful review of the topic, we refer the reader to the review by Pacheco and Segré (2019).

Together, these mechanisms depict a high level of inter-dependencies between bacterial species within an ecosystem. These interactions lead the ecosystem to structure around clusters of microbes that co-develop into a guild of co-abundant species. This concept was well illustrated in a recent study aimed at identifying gut bacterial species involved in post-antibiotic recovery in human cohorts. In a metagenome-wide association study, Chng et al. (2020) demonstrated that a succession of primary colonizers set the stage for late dominant species, which feed on the breakdown products of the pioneer species (Gibbons, 2020). This study identified 7 bacterial species acting as primary colonizers, with a metabolic capacity to extract carbon and energy from mucin and complex dietary carbohydrates, thus acting at the bottom of the food chain. Even if in this example most of the identified primary colonizers were abundant species, low abundance bacteria (<0.1% relative abundance) should not be neglected as they may carry essential functions that support growth of other dominant species. This concept has been very well illustrated in a study of “Candidatus Desulfosporosinus infrequens,” a sulfate-reducing organism found in wetlands (Hausmann et al., 2019). Although it remained in a seemingly dormant state at zero-growth over more than 7 weeks, it was reported to be in fact highly metabolically active, contributing to regulate methane production and therefore to sustain a diverse ecosystem (Hausmann et al., 2019).

In view of the intricate interplay between the gut microbiome function and its host metabolism, it is now established that loosing part of these functions is associated with a number of modern non-communicable diseases. As a consequence, techniques designed to manipulate gut dwelling microbiomes

Frontiers in Cell and Developmental Biology | www.frontiersin.org 2 September 2021 | Volume 9 | Article 719072
are gaining increasing attention, and several microbiome-based biotherapies are currently in development (Doré et al., 2017; Valencia et al., 2017). Hence, a deep understanding of gut microbiome ecology and how it can be durably restored is crucial for effective clinical translation. In this regard, a recent study evaluated bacterial dispersal strategies of human gut-associated microbes and classified them in five categories that may provide a guide for appropriate restoration strategies: (i) “tenacious” strains that are highly persistent among human communities, (ii) “spatiopersistent” strains that tend to be associated with specific geographical locations but colonize at a later developmental stage (i.e., not in infants), (iii) “heredipersistent” strains that tend to persist within closely related individuals such as within families and have broad geographical presence, (iv) average persistent strains, and (v) non-persistent strains (Hildebrand et al., 2021). Interestingly, the authors propose that fecal microbial transplantation may be most efficient to target tenacious and spatiopersistent taxa while heredipersistent taxa may require regular reinforcements, which may therefore be best targeted through chronic single strain exposure.

Table 1: Non-exhaustive list of prominent keystone taxa of the human gut microbiome.

| Keystone species (in alphabetical order) | Function | Method of identification | Example of reported disease association in humans |
|------------------------------------------|----------|--------------------------|-------------------------------------------------|
| Akkermansia muciniphila                  | Mucin degrader | Empirical (Belzer et al., 2017) | Intestinal inflammation, obesity and metabolic diseases (Yassour et al., 2016) |
| Bacteroides thetaotaomicron              | Degradation of complex carbohydrates (arabinogalactan); selective BSH activity (Yao et al., 2018) | Empirical (Cartmell et al., 2018) | Unclear – Controversial association with IBD (Sitkin and Pokrotnieks, 2019) |
| Bifidobacterium longum                   | Degradation of complex carbohydrates, particularly Human Milk Oligosaccharides; BSH activity (Tanaka et al., 2000) | Empirical (Yú et al., 2013; Gotoh et al., 2019) | Highly prevalent in healthy newborns (Favier et al., 2003) |
| Bifidobacterium pseudolongum             | Degradation of complex carbohydrates | Empirical (Centanni et al., 2018) | Highly prevalent human breast milk (Lugi et al., 2020) |
| Christensenella minuta                   | Stimulate ecosystem diversity (Mazier et al., 2021); acetate producer (Morotomi et al., 2012); BSH activity (Déjean et al., 2021) | Co-occurrence networks (Goodrich et al., 2014; Kumpitsch et al., 2020 ahead of publication) | Obesity and metabolic diseases (Goodrich et al., 2014); Crohn's disease (Pascal et al., 2017) |
| Faecalibacterium prausnitzii             | Butyrate producer | Presence/absence (Leylabadlo et al., 2020) | Crohn's disease (Sokol et al., 2008); Ulcerative Colitis (Varela et al., 2013) |
| Methanobrevibacter smithii               | Produces methane from H₂ and acetate | Empirical and co-occurrence networks (Goodrich et al., 2014; Kumpitsch et al., 2020 ahead of publication) | Obesity (Goodrich et al., 2014); Crohn's disease (Pascal et al., 2017) |
| Ruminococcus bromii                      | Resistant starch degrader; Butyrate producer | Empirical (Ze et al., 2012) | Highly prevalent microbe in healthy individuals (Beighn et al., 2021) |

**IMPORTANT METABOLIC PATHWAYS UNDER THE GUT MICROBIOME INFLUENCE**

Beyond microbiota classification, insights on the functional impact of the microbiome are emerging from metagenomic analyses, integration with omics data sets, particularly metabolomics and in vivo validation studies. In this section, we review recent studies highlighting the key contribution of microbial metabolites and associated pathways in controlling host physiology. Among the multitude of metabolic activities harbored by human gut microbiomes, we focus on the roles of short chain fatty acids (SCFA), tryptophan- and cholesterol-derived metabolites, and their crosstalk with host factors, e.g., histone modifying enzymes, G proteins-coupled receptors, aryl hydrocarbon receptor (AhR), indoleamine 2,3-dioxygenase 1 (IDO1) and tryptophan hydroxylase 1 (Tph1) in barrier maintenance, immune regulation, and the gut-brain axis.

**SCFAs**

Short chain fatty acids are the primary end products of bacterial fermentation of dietary fibers (but can also be derived from proteins and peptides in a lesser extent) and are important regulators of gut microbial ecology as well as host physiology. The main SCFAs are acetate, propionate, and butyrate (Cummins et al., 1987). Fiber-derived monosaccharides, such as hexoses, deoxyhexoses, and pentoses are converted by several bacterial metabolic enzymes to pyruvate which is then further metabolized to acetyl-CoA, succinate or lactate that primarily feed SCFA production. Acetate is derived from acetyl-CoA generated from pyruvate directly or through the Wood-Ljungdahl pathway. Butyrate is also produced from acetyl-CoA, but through the condensation of two acetyl-CoA molecules into acetooacetyl-CoA that is metabolized to butyryl-CoA and then butyrate. Some gut
bacteria can also convert lactate to butyrate. Propionate is derived from lactate or succinate in the acrylate and succinate pathways, respectively. It can also be produced by the propanoiod pathway that converts deoxyhexoses to propionyl-CoA. The concentration of SCFAs is highest in the proximal colon reaching \(\sim 130 \, \text{mmol/kg of luminal content} \) (Cummins et al., 1987). However, the effective concentration reaching the intestinal cells is presumably lower due to the thick mucus layer and intestinal peristalsis. Among the SCFAs, butyrate constitutes an important energy source for colonocytes and is mostly consumed in the colon. Propionate and acetate are further metabolized in the liver, but taken the high concentration of acetate in the gut, it is the main SCFA that remains in the systemic circulation. Nonetheless, butyrate and propionate can also impact host systemic physiology indirectly through hormonal and nervous system signals. SCFAs can enter the cells through diffusion or via the transporter SLC5A8 and exert their effects through three reported mechanisms: (a) epigenetic control of gene expression via inhibition of histone deacetylases (HDAC), e.g., in intestinal epithelial cells (IECs) and immune cells, (b) by acting as ligands of G-protein coupled receptors (GPCRs), primarily GPR43 and GPR41, also called free fatty acid receptors 2 (FFAR2) and FFAR3, respectively, and GPR109A, also known as niacin receptor 1 or Hydroxycarboxylic Acid Receptor 2 (HCA2), and/or (c) by acting as an AhR agonist, as has been shown for butyrate in IECs (Marinelli et al., 2019).

Because of the energetic reliance of colonocytes on butyrate, it is not surprising that this SCFA is a critical regulator of intestinal barrier integrity and mucosal immune homeostasis. Butyrate confers a protective role in experimental mouse models of colitis [e.g., with dextran sodium sulfate (DSS)], IL10\(^{-/-}\) mice (Wang et al., 2016) or Clostridium difficile infection (Fachi et al., 2019). These effects were also noted in ulcerative colitis patients, as shown early on by Scheppach et al. (1992). Butyrate also protects against colitis-associated colorectal cancer (CRC) as has been reported using the \(Apc^{min/+}\) mice (Singh et al., 2014) or with the azoxymethane (AOM)-DSS model (Singh et al., 2014). In contrast, in the \(Apc^{min/+}\)/\(Msh2^{-/-}\) mouse model with stem-like CRC characteristics, butyrate was shown to promote tumorigenesis (Belcheva et al., 2014), presumably through enhancing stem cell regeneration. Mechanistically, butyrate signals through GPR43 and GPR109A on IECs to stimulate inflammasome-dependent IL-18 production (Arpaia et al., 2013; Furusawa et al., 2015). Similarly, colonic infusion of SCFA mixtures in overweight/obese men, at concentrations comparable to those reached after dietary fibers intake, increased fat oxidation and energy expenditure (Canfora et al., 2017). Mechanistically, SCFAs may act by stimulating the production of the anorexigenic gut hormones peptide YY (PYY) and glucagon-like peptide-1 (GLP-1) (Figure 1), as has been shown in humans with acetate (Freeland and Wolever, 2010) and propionate (Venter et al., 1990) or through intestinal gluconeogenesis (IGN), where both propionate and butyrate were shown to enhance IEC de novo synthesis of glucose, stimulating increased insulin sensitivity through gut-brain communication (De Vadder et al., 2015). GPR41 mediates improved energy metabolism through its expression on neurons of the enteric nervous system (ENS) and on sympathetic neurons that promote enhanced energy expenditure (Figure 1). Consistently, \(Gpr41^{-/-}\) mice were shown to be leaner than wild-type controls (Samuel et al., 2008). On the other hand, the results with \(Gpr43^{-/-}\) mice are controversial as these mice were described to be obese even on normal diet in one study (Kimura et al., 2013), but lean with improved metabolic parameters in another (Bjursell et al., 2011). Nonetheless, these mouse phenotypes were lost under germ-free conditions or with antibiotic treatment, demonstrating microbiota-mediated metabolic effects.

**Tryptophan Metabolism**

Microbial metabolism of dietary tryptophan (enriched in cruciferous green leaf vegetables, e.g., parsley, cauliflower, kale, broccoli, etc.) has recently emerged as an important pathway by which the microbiota regulates intestinal homeostasis, particularly through AhR activation. Since its cloning in 1992 (Burbach et al., 1992), AhR has gained much interest for
its role as an environmental sensor. Beyond its activation by xenobiotics, tryptophan-derived ligands catabolized by the microbiota, including indole, indolo[3,2-b]carbazole, indole acetic acid (IAA), 3-methylindole and tryptamine, to name a few, have been demonstrated to act as potent high-affinity AhR ligands (Zelante et al., 2013; Hubbard et al., 2015). AhR is expressed in different intestinal cell types, including IECs, immune cells, particularly intraepithelial lymphocytes (IELs), innate lymphoid cells (ILC3) (Gomez de Agüero et al., 2016), more recently ILC2 (Li et al., 2018), Th17 and Treg (Quintana et al., 2008; Veldhoen et al., 2008), and neurons of the ENS (Obata et al., 2020). Through this collective expression, AhR exerts physiologically crucial roles in barrier integrity and intestinal and immune homeostasis, notably through regulation of IEC tight junctions (Yu et al., 2018; Singh et al., 2019), generation and survival of IELs (Cervantes-Barragan et al., 2017), production of IL-22 (Zelante et al., 2013) and IL-10 (Aoki et al., 2018; Powell et al., 2020), and regulation of peristalsis and microbiota density (Figure 1). In IECs, AhR has been recently implicated in the regulation of goblet cells differentiation, particularly in preventing goblet cell depletion in the colon during aging (Powell et al., 2020). This process is mediated by IL-10 and induced by AhR in response to microbiota-derived indoles. IL-22 or type I IFN, which are involved in intestinal tissue repair following acute injury, do not seem to be required in this case (Powell et al., 2020). In parallel to IECs, AhR activation in CD4+ T cells was shown to regulate their differentiation into CD4+CD8αα double-positive immunoregulatory IELs. These cells are absent in germ-free mice but restored with Lactobacillus reuteri, a species with tryptophan catabolizing capacity (Cervantes-Barragan et al., 2017). AhR activity is similarly required for the generation and maintenance of ILC3 (Gomez de Agüero et al., 2016), and patients with Crohn’s disease exhibit decreased AhR expression in their inflamed ileum accompanied by a conversion of ILC3

**FIGURE 1** | Schematic representation of the effects of select microbial metabolites on intestinal homeostasis, mucosal immune regulation, and metabolic health of the host.
to ILC1 (Li et al., 2016). Last, AhR expression is elevated in intestinal T\textsubscript{Reg} and seems to be required for their gut homing as well as for suppression of Th1 inflammatory gene expression (Ye et al., 2017). AhR is also expressed in colonic neurons, in a microbiota-dependent manner, and this neuronal-specific expression is central in the control of intestinal peristalsis, positioning AhR as a nexus of intestinal neural circuitry (Obata et al., 2020). Collectively, AhR protects the epithelial barrier, promotes intestinal immune tolerance and protects from intestinal inflammation. Consequently, deregulation of its activity is associated with inflammatory and metabolic diseases, and microbiome stimulation of this pathway, particularly through tryptophan metabolism, has been demonstrated as an “environmental” mean to counter these pathologies. For instance, individuals with inflammatory bowel diseases (IBD) (Lamas et al., 2016), the metabolic syndrome (Natividad et al., 2018) or celiac disease (Lamas et al., 2020) have decreased fecal concentrations of AhR ligands and reduced AhR activity. Interestingly, supplementation of experimental mice modeling these pathologies with a high-tryptophan diet, AhR ligands or with bacterial species that metabolize tryptophan such as L. reuteri, improved the mice health status (Marafini et al., 2019; Chen et al., 2020). In a randomized controlled clinical trial, administration of AhR ligands in the form of traditional Chinese medicine indigo naturalis to ulcerative colitis patients for 8 weeks showed clinical benefit, including a decrease in the Mayo score, mucosal healing and remission in some cases (Naganuma et al., 2018). Nonetheless, caution is warranted prior to considering the development of AhR ligands for therapeutics taken toxicity issues with deregulated AhR responses.

Besides AhR ligands, tryptophan is additionally metabolized into kynurenine and serotonin (Clarke et al., 2012; Yano et al., 2015). In the kynurenine pathway (KP), IDO1 is mainly responsible to convert tryptophan to kynurenine and downstream end products including niacin, nicotinamide adenine dinucleotide (NAD), quinolinic acid and kynurenic acid (Cervenka et al., 2017; Kennedy et al., 2017). The latter exerts immunoregulatory effects and protects the intestine by signaling through GPR35, expressed on IECs and immune cells (Gao et al., 2018). The serotonin pathway converts tryptophan into the neurotransmitter 5-hydroxytryptamine (5-HT), i.e., serotonin, via TpH1 expressed in a specialized IEC type known as enterochromaffin cells in the gut, and TpH2 in the brain. While peripheral 5-HT, which constitutes 90% of all serotonin produced in the body, does not cross the blood-brain barrier (BBB), it regulates several intestinal processes including stimulation of ENS neurons, peristalsis and nutrient absorption, to name a few (Mawe and Hoffman, 2013). Moreover, both tryptophan and 5-HT precursor (5-HTP) cross the BBB impacting central serotonin effects on host physiology. The commensal microbiota plays important roles in tryptophan metabolism to kynurenine and serotonin as demonstrated in GF or antibiotics-treated mice [reviewed in Agus et al. (2018)]. Several commensal bacteria express enzymes related to KP enzymes and can thus produce kynurenine metabolites, e.g., the neurotoxic 3-hydroxyanthranilic acid (O’Farrell and Harkin, 2017). Further, through SCFA and BA metabolism, the microbiota can induce TpH1 and stimulate 5-HT biosynthesis (Reigstad et al., 2015; Yano et al., 2015). Gut-derived kynurenines and 5-HT are additionally implicated in the pathogenesis of chronic inflammatory, metabolic and neuropsychiatric diseases. For instance, IDO1\textsuperscript{-/-} mice are more susceptible to 2,4,6-trinitrobenzene sulfate-induced colitis (Takamatsu et al., 2013) whereas TpH1\textsuperscript{-/-} mice show enhanced protection in response to DSS- or dinitrobenzene sulfonic acid-induced colitis (Ghia et al., 2009), suggesting that while kynurenine is protective in the gut, 5-HT might be deleterious. However, more recent findings indicate that 5-HT could exert pro- or anti-inflammatory effects in the gut depending on the respective engagement of 5-HT\textsubscript{7} versus 5-HT4 receptors (Spohn and Mawe, 2017). Kynurenines and 5-HT play contrasting roles in obesity and insulin resistance. The KP is overactivated in obesity and its activity correlates with BMI and the metabolic syndrome, presumably through the action of kynurenine derivatives such as xanthurenic acid (Oxenkrug, 2013). In contrast, 5-HT levels are decreased in obese individuals, which is consistent with the role of 5-HT in promoting satiety (Voigt and Fink, 2013), lipolysis in white adipose tissue and hepatic gluconeogenesis (Sumara et al., 2012).

### Cholesterol and Lipid Metabolism

Cholesterol and lipid metabolism by the intestinal microbiota is an additional facet by which the microbiota influences host physiology. On one hand, microbial components, specifically P. anaerobius, have been identified as inducers of cholesterol biosynthesis in colonocytes, mediated by SREBP2 activation downstream of TLR signaling, which supports dysplasia and CRC development in a mouse model, and is consistent with elevated levels of this bacteria in the stool of CRC patients (Tsoi et al., 2017). On the other hand, microbial metabolism of cholesterol into coprostanol, a poorly absorbed sterol, has been recently demonstrated as a mechanism reducing host serum cholesterol levels (Kenny et al., 2020). Notably, this function was attributed to a clade of uncultured bacteria harboring ismA genes encoding cholesterol dehydrogenases (Kenny et al., 2020). Besides cholesterol, bacterial metabolism of sphingolipids (SL), lipids with a long-chain amino alcohol backbone, also contributes to immune homeostasis in the gut and to the host metabolic health. Bacteria of the Bacteroidetes phylum, which constitutes ∼30-40% of the healthy human intestinal microbiota, have the capacity to synthesize sphingolipids (SL), owing to their expression of the enzyme serine palmitoyltransferase (Heaver et al., 2018). Bacterially derived SL promote immune homeostasis in the gut, acting early in life to tame invariant natural killer T (iNKT) cells proliferation (An et al., 2014). Consistently, in a model of oxazolone-induced colitis, treatment of mice with B. fragilis SL lessened the colitis phenotype by reducing iNKT cell numbers (An et al., 2014). Notably, the stools of IBD patients contain an elevated signature of host SL including ceramides, but are depleted of bacterially derived SL which protect the intestine. Indeed, colonization of germ-free mice with B. thetaiotaomicron deficient in SL elicited intestinal inflammation (Brown et al., 2019). To address how the gut microbiota influence host metabolic pathways and ceramide levels, a recent study explored this question in a...
model of diet-induced obesity (Johnson et al., 2020). The authors showed that bacterially derived SL can enter colonocytes and reach the liver through the portal vein circulation, impacting metabolic parameters, e.g., insulin resistance, primarily through liver ceramides (Johnson et al., 2020). Together, these indicate that cholesterol and lipid metabolism by gut bacteria significantly influence host metabolism and physiology. However, these pathways are still poorly understood and require to be fully investigated before further therapeutic exploitation.

**Bile Acid Metabolism**

Bile acid metabolism by the gut microbiota has gained considerable interest in the recent years as they are being recognized as crucial microbiome-derived metabolites that regulate multiple important functions involved in health and disease (Hylemon et al., 2018). Primary bile acids are mostly synthesized by the liver hepatocytes from cholesterol following irreversible 7alpha-hydroxylation by the cytochrome P450 CYP7A1 (Chávez-Talavera et al., 2017). In humans, these are cholic acid (CA) and chenodoxycholic acid (CDCA), while in mice, CDCA is further metabolized into bismuricholic acid (betaMCA) (Pandak and Kakiyama, 2019). These hydrophobic primary bile acids are then made amphipathic through conjugation with glycine and taurine before being secreted in the gall bladder along phospholipids to make up the bile that will primarily serve as detergent upon release into the duodenum during digestion (Chávez-Talavera et al., 2017). In the small intestine, conjugated bile acids are deconjugated by microbial bile salt hydrolases (BSH) that release the hydrophobic moieties that are reabsorbed mostly through passive diffusion along the epithelium and through active reabsorption in the terminal ileum. In total, 95% of the initial bile acid pool is reabsorbed through this enterohydratory cycle. Yet, 5% of bile acids escape reabsorption and travel down the colon where they undergo further metabolism by gut bacteria such as 7alpha-dehydroxylation, which leads to formation of secondary bile acids such as deoxycholic acid and lithocholic acid from the metabolism of CA and CDCA, respectively, (Ridlon et al., 2016). This is a brief overview of the complex metabolism of bile acids. For an exhaustive review of the role of gut microbes on bile acid metabolism, we refer the reader to the work by Ridlon et al. (2016). Gut microbial taxa with documented BSH activity include *Lactobacillus* spp. (Foley et al., 2021), *Bifidobacterium longum* (Tanaka et al., 2000), *Enterococcus faecalis* (Chand et al., 2018), *B. thetaiotaomicron* (Yao et al., 2018), and *Christensenella minuta* (Déjean et al., 2021), some of which have been classified as keystone species (Table 1).

Beyond their role in lipid absorption, bile acids have systemic functions as they also regulate hepatic energy metabolism, adipocyte differentiation and dampen immune activation through their interaction with bile acid receptors Farnesoid X Receptor (FXR) and Takeda G protein Receptor 5 (TGR-5) (Foley et al., 2021). Because of these multiple effects, they were recently suggested to form gut microbiota-derived hormones (Kliewer and Mangelsdorf, 2015; Hylemon et al., 2018). Both conjugated and unconjugated bile acids participate in this host-microbiota crosstalk. Interestingly, new bile acid conjugates specifically produced in the small intestine, were recently discovered. These involve bacterial conjugation with phenylalanine, tyrosine, and leucine, three hydrophobic amino acids, which had never been described associated with these molecules. Unsurprisingly, we still ignore the physiological role of these novel microbially derived compounds (Quinn et al., 2020).

Bile acids have been associated with a number of chronic disorders including obesity, NASH, IBD (Schirmer et al., 2019), Primary Biliary Cholangitis (formerly known as Primary Biliary Cirrhosis) and Primary Sclerosing Cholangitis. Interestingly, some disorders have been specifically associated with a defect in gut bacterial metabolism of bile acids such as *Clostridioides difficile* infections, which have been shown to be corrected by restoring gut microbial BSH activity (Mullish et al., 2019). Indeed, bile acid deconjugation releases free unconjugated bile salts that are toxic to most bacteria and thus act as a regulator of the microbial ecosystem. Hence, this is one of the key functions carried by gut bacteria that impacts significantly on the composition of the gut microbiome community.

**BIOINFORMATIC TOOLS TO IDENTIFY KEYSTONE SPECIES AND FUNCTIONS OF THE GUT MICROBIOME**

In light of these findings, restoring key metabolic pathways carried by gut microbiota could open a wide range of therapeutic perspectives. In particular, the identification of keystone species carrying these functions in the gut microbiome appears as an essential step for the development of future biotherapies targeting the gut microbial ecosystem. Keystone species of the human microbiome have often been identified using empirical evidence (Banerjee et al., 2018). Nevertheless, bioinformatics is increasingly used to identify keystone species and several methods have been developed to exploit next-generation sequencing (NGS) data.

**“Presence or Absence” Associated With Health and Disease**

One of the most common methods to characterize the human gut microbiome has been the use of amplification and sequencing of marker genes in stool samples. The most used marker gene is the 16S rRNA gene for the detection of bacteria, but other housekeeping genes are occasionally used to capture bacterial diversity (Case et al., 2007; Ogier et al., 2019). 16S rRNA can be used to compare the abundance of microorganisms at the genus or species level in different states (e.g., healthy versus diseased, different food sources, etc.) by assigning the reads to clusters of organisms grouped by taxonomic marker gene similarity, called Operational Taxonomic Units (OTUs). A widely used technique to then assess variations between microbial communities is to use UniFrac beta-diversity metric coupled with unsupervised multivariate statistics using Principal Component Analysis (PCA) or its derivatives (Lozupone and Knight, 2005; Ramette, 2007). However, the main hurdle associated with the calculation of the UniFrac method is the high computer resources.
power required, although this has been largely improved in the latest Striped UniFrac algorithm (McDonald et al., 2018). More complex probabilistic methods, such as Dirichlet multinomial mixtures, have been developed to improve the analysis of metagenomics samples by clustering and classifying microbial communities based on a probability distribution. This method especially considers the discrete nature, the sparsity and the variability of the sequencing depth, and has been applied to the analysis of samples from obese and lean twins and to IBD patients (Holmes et al., 2012; Ding and Schloss, 2014).

Even if evaluating the presence or absence of a genus or species between two states using marker gene sequencing in stool samples seems promising, a key part of the keystone species definition is the interspecies interaction (Banerjee et al., 2018). Hierarchical clustering of bacterial communities correlated with disease association have been extensively used to attempt identification of important bacterial taxa (Jackson et al., 2018). Correlation with taxon presence or absence is often confirmed using metrics such as Jaccard’s index: between two species, this index is defined by the ratio between the number of samples where both species are present out of the number of samples tested. The values of this index range between 0 and 1, from no correlation to a strong correlation, respectively, (Mainali et al., 2017). Nevertheless, an extensive study on microbial community modeling showed that indexes selected to evaluate correlations between species should be adapted to the specific dataset being studied. It is noteworthy that Jaccard index has been reported to have a relatively low sensitivity compared to other metrics such as the Pearson index when applied to co-occurrence networks (Berry and Widder, 2014). Therefore, even if the use of Jaccard’s index is reliable to identify the correlations using taxon presence or absence, the use of Pearson or Spearman indexes should be preferred to assess correlations when using network-based methods.

To overcome the issues associated with taxon-based correlation analysis that may lead to conflicting results due to spurious associations, it is noteworthy that Wu and co-authors have recently proposed to exploit the concept of bacterial guild to reduce metagenomic data dimensionality into ecologically meaningful functional units (Wu et al., 2021). Although this approach is still limited by the use of relative abundance data, it proposes an interesting approach to refine data analysis of 16S-based datasets to identify relevant disease associations.

**Prediction of Interspecies Interactions by Network-Based Methods**

The need to consider interspecies interactions to identify keystone species in a community led to the development of new network-based methods. The most common methods are co-occurrence or co-abundance networks applied to 16S rRNA gene (or other marker gene) sequencing or metagenomic data. These networks are often produced by calculating a pairwise correlation coefficient between each pair of OTUs but other methods are being developed to build interaction networks (Berry and Widder, 2014).

Co-occurrence networks have been extensively used to identify keystone species. An exhaustive study by Berry and Widder (2014), evaluated the performance of these networks in interaction with several correlation metrics. They used generalized Lotka-Volterra modeling (gLVM) to simulate competitive and cooperative interactions between species (Berry and Widder, 2014). This study revealed that high mean degree (the average number of edges a node has in the network), high closeness centrality (the average distance of a node to any other node), high transitivity and low betweenness centrality (the betweenness centrality of the node A is the number of shortest paths between a pair of nodes B and C passing through the node A) can predict the keystone nature of species with 85% accuracy (Berry and Widder, 2014). These parameters produce highly interconnected nodes (i.e., keystone species) called “hub” patterns corresponding to a number of species interacting together directly and indirectly (Berry and Widder, 2014; Layeghifard et al., 2017; Banerjee et al., 2018). This enables identification of the guilds of bacteria that depend on the presence of keystone species. For example, this method was used by Zhang et al. (2014) on 454-pyrosequencing 16S rRNA sequencing data from human intestinal biopsy samples. It allowed to identify Ruminococcus gnavus, Faecalibacterium prausnitzii, Prevotella copri, and Anoxybacillus flavithermus as potential keystones of a healthy human intestinal mucosal microbiota because they displayed the highest number of linkers (Zhang et al., 2014). As already mentioned, only F. prausnitzii has been so far empirically validated as a keystone species for its ability to produce butyrate (Table 1). Its loss has been associated with the development of IBD in several studies (Sokol et al., 2008; Varela et al., 2013).

Another study from Fisher and Mehta (2014) used discrete-time Lotka-Volterra models to simulate the abundance variations of 10 species for 1000 timesteps and 100 initial conditions. After demonstrating that correlations in species abundance were not predictive of interactions between species, they used this simulated dataset to prove that LIMITS, an algorithm that uses sparse linear regression corrected by bootstrap aggregation, can be relevant to identify keystone species in two time-series samples from the gut microbiome (Fisher and Mehta, 2014). They concluded that B. stercoris and B. fragilis showed more interactions than other bacteria and that these interactions were mainly beneficial, since the growth of these two bacteria has been correlated with an increased abundance of B. thetaiotaomicron and E. rectale, the latter being a well-known butyrate producer (Fisher and Mehta, 2014). These results are coherent with other studies identifying Bacteroides sp. as key members of the gut microbiome (Loftus et al., 2021) involved in the degradation of complex carbohydrates (Cartmell et al., 2018). Nevertheless, the major drawback of this method is the need of time-series samples which are difficult and time-consuming to obtain. Furthermore, only a few abundant species are studied, thus the identification of low abundant keystone species is fairly unlikely using this method.

Other network-based methods, such as association rule-mining, are being applied to microbiome sequencing data in order to identify keystone species. Briefly, this method allows to estimate whether the presence of a species is required for
the presence of other species. In a study by Chng et al. (2020), using the "efficient_apriori" Python package, they inferred binary association rules between species on 782 microbiome profiles. This method allowed to identify “primary” species, which presence is required for other species to thrive. Furthermore, they showed that some of these "primary" species, such as *Bacillus uniforinmis*, *F. praunntzii* or *Ruminococcus torques*, were associated with the recovery of the gut microbiota after antibiotic exposure and carried specific metabolic functions such as mucin and carbohydrates degradation (Chng et al., 2020). As mentioned in Table 1, other members of the *Ruminococcus* genus have been identified as resistant-starch degraders (Ze et al., 2012) essential to maintain a healthy gut microbiome (Beghini et al., 2021).

Co-occurrence networks between members of the microbiota can also be applied using generalized boosted linear models (GBLMs), as exemplified in a study where it was implemented to investigate the Human Microbiome Project cohort (Faust et al., 2012). More recently, a co-occurrence network using the Random Matrix Theory (RMT) method was applied to murine gut microbiome data in order to identify putative keystone species. In this study, 32 were identified as highly connected species linked to other OTUs (Liu et al., 2019).

Although these methods are useful to predict putative keystone species, the identification of the keystone functions carried by these species is essential to understand the interactions between the keystone bacteria and its guild. However, the use of amplicon sequencing does not allow the precise identification of the strains or the metabolic functions they carry within the gut microbiome because the ascertainment of OTUs only allows the reconstruction of metabolic pathways based on reference genomes, thus inducing a possible loss of strain-specific metabolic functions (Frioux et al., 2020). Therefore, a few recent methods have been developed to overcome these issues using metagenomic sequencing that reconstruct strain-specific metabolic pathways.

**Reconstruction of Metabolic Pathways at Ecosystem Level**

The recent development of metagenomics has provided a clearer view of the diversity of the gut microbiota, especially by allowing access to yet-uncultured bacteria (Almeida et al., 2019). The reconstruction of genome-scale metabolic networks and models (GSMNs) using Metagenome-Assembled Genomes (MAGs) or the inference of functional categories to single genes allows to precisely predict the metabolic capabilities of the gut microbiome (Frioux et al., 2020). A good example of such tools is the Meta2G2Metabo algorithm developed by Belcour et al. (2019) that enables the analysis of metabolic pathways at the ecosystem level using both reference genomes and MAGs. As keystone species carry essential metabolic functions in the gut ecosystem, this tool was used to detect putative keystone bacteria. In order to accurately predict the metabolic capabilities of the communities, both available nutrients and genome information (from reference databases or metagenomic samples) are combined to predict minimal communities of bacteria that can produce target metabolic compounds. The bacteria encountered in all predicted minimal communities are considered essential symbionts, whereas bacteria encountered in at least one of the predicted minimal communities, but not necessarily all of them, are considered alternative symbionts (Belcour et al., 2019). This method has been applied to a set of over 1,500 reference genomes from the human gut, allowing the identification of 11 essential symbionts, namely *Propionibacterium* sp. KPL2009, *Paenibacillus polymyxa*, *Bacillus licheniformis*, *Lactococcus lactis*, *Enterococcus casseliflavus*, *E. faecalis*, *Hungatella hathewayi*, *Dorea longicatena*, *R. torques*, *Burkholderiales bacterium*, and 194 alternative symbionts (Belcour et al., 2019). As the reference genomes were assembled from 155 fecal samples, bacterial communities from each individual were mixed, maybe explaining the large amount of alternative symbionts (Zou et al., 2019). In addition, the fairly reduced number of predicted keystone species compared to the input dataset of genomes could be due to the non-exhaustive list of initial nutrients and target metabolic compounds that need to be provided to the software in order to predict minimal communities. Nonetheless, this tool is the first tool to the best of our knowledge that is specifically designed to identify both putative keystone species and their associated metabolic functions in complex microbial communities.

**FINAL CONSIDERATIONS**

Modern sequencing technologies enable broad mapping of virtually any microbial ecosystem. Nevertheless, it is important to keep in mind that the quality of microbiome data and the information we derive from them highly rely on the quality of the original sampling. Indeed, environmental factors strongly influence bacterial communities that adapt to any variation, being the time of day, temperature, pH, food supply (i.e., diet for gut microbiomes), and the age of their host for host-associated microbiomes. In this regard, the keystone species *C. minuta* has been reported to be increased with aging (Waters and Ley, 2019). Hence, repeated sampling can be recommended in order to account for temporal variations and obtain more accurate pictures of bacterial ecosystem compositions (Ji et al., 2019)

Another important consideration about sampling is that most large human studies focus on easily accessible locations with non-invasive tools, analyzing oral and fecal samples for instance. As a consequence, the analyzed microbiomes poorly represent the inner ecosystems located deep within the gut. Typically, fecal samples reflect the microbiome of the distal colon, largely dominated by Clostridia, while it has been shown that the gut microbiota follows a specific spatial distribution along the gastrointestinal tract, which is even further complicated by regional specificities as illustrated by the differences observed between luminal and mucosa-associated species (Donaldson et al., 2016). Hence, most studies of human cohorts are limited to the study of the distal gastrointestinal tract. Despite this limitation, NGS methods and bioinformatics have been effectively supporting the search for keystone species in the gut microbiome. The rapid improvement of NGS techniques that generates increasingly large datasets allowing for deep
characterization of the gut microbiome community also calls for new bioinformatics tools to analyze NGS datasets in a more effective and complete way.

For metagenomics studies and as summarized in Figure 2, two methodology decisions drastically influence the results: (i) the DNA sequencing technology and (ii) the bioinformatic methods that will be applied to analyze these datasets.

The choice of a DNA sequencing strategy determines the information retrieved from a sample. Amplicon sequencing, especially 16S rRNA gene sequencing, allows to have an overview of the bacterial content of the samples by assigning the obtained reads to taxonomies. Metabolic networks can then be inferred using the reference genomes of the species or genus identified in the samples (Frioux et al., 2020). Although this approach has some merits, several technological biases can result in partial taxonomic assignation, thus reducing the completeness of the analysis. For instance, 16S rRNA gene sequencing is often partial because it mostly targets a couple of hypervariable regions, although it is technically possible to target a nearly complete 16S rRNA sequence, depending on chosen primers and sequencing technology. Indeed, an extensive study by Johnson et al., revealed that a full 16S rRNA gene sequencing significantly improves the taxonomic resolution. Out of all of the partial sequencing tested, the V4 region performed worst, and the relative number of OTUs produced using the different sub-regions was not consistent depending on the identity threshold (Johnson et al., 2019). It has also been noted that the development of long-read sequencing platforms enhanced the accuracy of the sequencing thus improving the detection of single nucleotide polymorphisms (SNPs) in the complete 16S rRNA gene. Multiple copies of the 16S rRNA genes carrying unique SNPs can even be detected using this technique, allowing a strain-level identification (Johnson et al., 2019).

The thrive of metagenomic sequencing also benefited from the use of NGS to perform shotgun sequencing. A study by Ranjan et al. (2016) concluded that the use of shotgun sequencing compared to 16S sequencing significantly improved the diversity of bacterial species detected, thus allowing a finer prediction of the genes carried by a bacterial community. For instance, their study showed that more than 1,000 species of proteobacteria were only detected by shotgun sequencing performed on a stool sample and twice the amount of genes were predicted on average using one of the shotgun short-reads metagenomic technologies. The comparison of different short-read sequencing
technologies, namely MiSeq and HiSeq, showed that the extra length produced by MiSeq sequencing (150–300 bp compared to 100 bp) enables a more specific identification of species and improves accuracy and completeness of MAGs that are essential to reliably predict the metabolic potential of the ecosystem (Ranjan et al., 2016; Frioux et al., 2020). A more recent study showed that the use of long-read technologies (e.g., PacBio) improved drastically the completeness of the MAGs and associated predicted genes: more than 98% of the genes predicted from PacBio-assembled MAGs were complete compared to only 40% using HiSeq-assembled MAGs (Xie et al., 2020). Indeed, the use of long reads enables the detection of repeated elements often found in ribosomal RNA genes or bacteriophage-related insertions that are frequently missed with short read sequencing (Derakhshani et al., 2020).

Once sequencing data have been acquired, the identification of species in the gut microbiome requires the assignment of the reads to taxonomies. At this step, not only is the quality of the reads important, but the accurate assignment to taxonomies also depends on the reference catalog used. For the assignment of 16S rRNA genes, two factors strongly influence results. First, there is high reliance on the diversity contained in the reference catalog and second, the amplified regions can skew the identification of OTUs. Indeed, 16S rRNA gene catalogs are often built from complete 16S sequences whereas the amplification and sequencing of the 16S rRNA gene usually targets hypervariable regions (V-regions), leading to biased performances and challenges when comparing different studies. To overcome this hurdle, a new tool called OTUX has been proposed, that combines several custom datasets of OTUs defined by 16S rRNA V-regions retrieved from full-length 16S rRNA sequences (Yadav et al., 2019). This new method was challenged against conventional full-length 16S databases and revealed an improved assignment of reads for all of the V-regions targeted, except for the V1–V3 region (Yadav et al., 2019). For the assignment of shotgun reads, using MAGs could significantly improve identification accuracy of bacterial species. Recent research efforts produced a comprehensive catalog containing more than 200,000 reference genomes from the human gut microbiome referenced as the UHGG database (Almeida et al., 2020). This work also revealed that 81% of the species in the catalog lacked a cultured representative, indicating the huge potential for discovering new keystone species in the future (Almeida et al., 2020).

Another critical parameter in the analysis of microbial communities using NGS is the sequencing depth. Several keystone species from the human gut microbiome, such as *C. minuta*, which presence in stool samples is significantly associated with BMI, are low abundance microbes that can be missed when using a reduced sequencing depth (Waters and Ley, 2019). Thus, an increased depth is required to ensure that the whole diversity of species from the gut microbiome is identified (Berry and Widder, 2014).

Considering data processing using bioinformatic tools, one of the main issues regarding the use of sequencing data to identify correlations between taxa within the gut microbiota is the compositional nature of these datasets. The analysis of read counts begins with their normalization by the total number of sequenced reads. As explained by Friedman and Alm, the estimation of correlation between parameters (e.g., species) is biased by the relationship between the fractions: because they must sum to 1, they are not independent (Friedman and Alm, 2012). Thus, a study can be artificially biased by the disappearance of highly abundant species due to technical or assignment issues: the relative abundance of low-abundant species will increase as the variables are dependent even if their absolute abundance is constant. Quantitative PCR (qPCR) is the predominant method to quantify biomass using 16S rRNA gene amplification, however, this method could significantly influence the models. Li et al. (2019) noticed a high variation between replicates when quantifying biomass from stool samples. Another flaw of this technique is due to the variable number of 16S rRNA gene copies in several microbial phyla such Firmicutes and Bacteroidetes, which results in an over-representation of such species. Friedman and Alm (2012) demonstrated that standard Pearson correlation estimation can falsely predict negative correlations between one dominant and several low abundant species because of the dependence between abundances. This issue has also been pointed by Berry and Widder (2014): they noted a loss of specificity in the co-occurrence networks when relative abundances were used. As a consequence, the SparCC method was developed to estimate the linear Pearson correlation between transformed variables: the log transformation of the ratio of abundances between a pair of OTUs because the ratio between abundances is independent from other OTUs included. Using SparCC, it was demonstrated that the bias observed in standard correlation studies that is induced by dominant species, is greatly reduced in simulated data of varying diversity. Applied to the HMP dataset, SparCC revealed new positive correlations between highly abundant and low abundant species, instead of the spurious negative correlations usually observed when using standard correlations (Friedman and Alm, 2012).

Recently, another method was developed by Li et al., to overcome the compositional bias when using generalized Lotka-Volterra (gLV) models. As explained above, gLV models are one of the most common approaches for modeling microbial interactions. These models also suffer from the compositional bias induced by the relative abundances. Indeed, absolute biomass values are needed in order to accurately fit the gLV differential equations model of each organisms’ growth rate to the data (Li et al., 2019). Therefore, Li et al. (2019) developed an algorithm called BEEM that estimates biomass *in silico* before inferring interactions when total biomass cannot be estimated experimentally. This algorithm introduces relative abundances in the equation modeling the growth rate of each species, resulting in two parameters that can be estimated using longitudinal datasets. This method was applied to diverse synthetic or existing datasets and accurately estimated biomass and gLV parameters. It also allowed the identification of *F. prausnitzii* and *B. uniformis* as putative keystone species sharing numerous positive interactions with other bacteria (Li et al., 2019). Yet, this method is limited by available growth information for common species. Hence, only common dominant bacteria for which the information exists, may be accurately predicted.
Finally, recent developments have proposed to apply inferential statistics to network-based models in order to gain confidence in result interpretation. A study by Röttjers et al. (2021) recently proposed the use of null models to identify network properties that can be used to compare networks. They showed that among 20 networks built from time-series stool samples collected from 20 women, the new tool called ANuran, could identify patterns that were found in 20–25% of the networks but only 3 associations were found in 10 networks or more, suggesting that associations between species or taxa may greatly vary from one individual to another (Röttjers et al., 2021).

Although this is a proof-of-concept work applied to a limited population, this is an interesting approach to robustly identify keystone species based on stable interaction networks.

CONCLUSION

In this review, we discussed the role of key gut microbes-derived metabolites in intestinal homeostasis, metabolic health, immune regulation, and gut-brain interactions. We next described current bioinformatic tools used in microbiome studies and highlighted weaknesses associated with some of these approaches for the identification of keystone species and their missing functions. We point for instance that commonly used bioinformatic tools that provide “presence or absence” information are not suitable in this regard as they fail to provide a view of species interactions. Network-based methods that calculate co-occurrence or co-abundance of species are thus more suited to predict keystoneness. However, these methods need to be complemented with approaches that reconstruct species-specific metabolic pathways, such as the Metage2Metabo algorithm, but in a system where nutrient, genomic information and metabolites output are simultaneously analyzed.

Our review has provided a snapshot of recent discoveries on gut microbiome metabolic activities and the current state of the field with respect to bioinformatics analysis of the microbiome. We propose that in the quest for keystone species, future studies should consider harmonization of sample and data processing and the integration of additional variables including age, gender, nutrition, and other environmental cues. It is also particularly important to provide mechanistic evidence supporting the functions of keystone species in modulating the microbiome ecosystem for instance through quorum sensing, cross-feeding, bacteriocins or through other as of yet unknown mechanisms. Such studies will set the stage to design microbiota-based therapeutic interventions to counter chronic diseases, by restoring keystone species and their beneficial effects on microbiome balance to support a healthy symbiotic interaction with their host.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

MS’s research is supported by grants from the ARC Foundation, New Aquitaine Region, IDEX Bordeaux, and SIRIC BRIO.

REFERENCES

Agus, A., Planchas, J., and Sokol, H. (2018). Gut microbiota regulation of tryptophan metabolism in health and disease. *Cell Host Microbe* 23, 716–724. doi: 10.1016/j.chom.2018.05.003

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., et al. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568, 499–504. doi: 10.1038/s41586-019-0965-1

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., et al. (2020). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnol.* 39, 105–114. doi: 10.1038/s41587-020-0603-3

An, D., Oh, S. F., Olzak, T., Neves, J. F., Avci, F. Y., Erturk-Hasdemir, D., et al. (2014). Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* 156, 123–133. doi: 10.1016/j.cell.2013.11.042

Aoki, R., Aoki-Yoshida, A., Suzuki, C., and Takayama, Y. (2018). Indole-3-pyruvic acid, an aryl hydrocarbon receptor agonist, suppresses experimental colitis in mice. *J. Immunol. Baltim.* 201, 3683–3693. doi: 10.4049/jimmunol.1701734

Arpaia, N., Campbell, C., Fan, X., Dikty, S., van der Veeken, J., deRoo, P., et al. (2013). Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation. *Nature* 504, 451–455. doi: 10.1038/nature12726

Asbury, S., Atallah, E., Vijay, A., Aithal, G. P., Grove, J. I., and Valdes, A. M. (2020). Lower gut microbiome diversity and higher abundance of proinflammatory genus Collinsella are associated with biopsy-proven nonalcoholic steatohepatitis. *Gut Microb.* 11, 569–580. doi: 10.1080/19490976.2019.1681861

Banerjee, S., Schlaeppe, K., and van der Heijden, M. G. A. (2018). Keystone taxa as drivers of microbiome structure and functioning. *Nat. Rev. Microb.* 16, 567–576. doi: 10.1038/s41579-018-0024-1

Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Aznicar, F., Maharjan, S., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* 10:e65088. doi: 10.7554/elife.65088

Belcheva, A., Irrazabal, T., Robertson, S. J., Streutker, C., Maughan, H., Rubino, S., et al. (2014). Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells. *Cell* 158, 288–299. doi: 10.1016/j.cell.2014.04.051

Belcour, A., Frioux, C., Aite, M., Bretaudeau, A., and Siegel, A. (2019). Metagene2Metao: metabolic complementarity applied to genomes of large-scale microbiota for the identification of keystone species. *Bioinformatics* 35, 1969–1975. doi: 10.1093/bioinformatics/btz840

Belzer, C., Chia, L. W., Aalvink, S., Chamlagain, B., Pironen, V., Knol, J., et al. (2017). Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *mBio* 8:e00770-17. doi: 10.1128/mBio.00770-17

Berry, D., and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5:219. doi: 10.3389/fmicb.2014.00219

Bjursell, M., Admyre, T., Göransson, M., Marley, A. E., Smith, D. M., Oscarsson, J., deRoo, P., et al. (2019). Microbial metabolic networks at the mucus layer lead to diet-independent butyrate and vitamin B12 production by intestinal symbionts. *mBio* 8:e00770-17. doi: 10.1128/mBio.00770-17

Brown, E. M., Ke, X., Hitchcock, D., Jeafavre, S., Avila-Pacheco, J., Nakata, T., et al. (2019). *Bacteroides*-derived sphingolipids are critical for maintaining intestinal homeostasis and symbiosis. *Cell Host Microb.* 25, 668.e7–680.e7. doi: 10.1016/j.chom.2019.04.002

Burbach, K. M., Poland, A., and Bradfield, C. A. (1992). Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor.
Chand, D., Panigrahi, P., Varshney, N., Ramasamy, S., and Suresh, C. G. (2012). Microbial co-occurrence relationships in the human microbiome. PLoS Comput. Biol. 8:e1002687. doi: 10.1371/journal.pcbi.1002687

Chen, J., Haller, C. A., Jernigan, F. E., Koerner, S. K., Wong, D. J., Wang, Y., et al. (2020). Modulation of lymphocyte-mediated tissue repair by rational design of heterocyclic aryl hydrocarbon receptor agonists. Sci. Adv. 6:eaya8230.

Cheng, K. R., Ghosh, T. S., Tan, H. Y., Nandi, T., Lee, I. R., Ng, A. H. Q., et al. (2020). Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. Nat. Ecol. Evol. 4, 1256–1267. doi: 10.1038/s41559-020-1236-0

Clarke, G., McKeown, D. P., Gasczner, G., Quigley, E. M., Cryan, J. F., and Morrison, D. J., Murphy, K. G., Zachariae, C., et al. (2014). Human genetics shape the gut microbiome. Science 346, 1744–1754. doi: 10.1126/science.aah5825

Cummings, J. H., Pomare, E. W., Branch, W. J., Naylor, C. P., and Macfarlane, G. T. (1991). Short chain fatty acids in human large intestine, portal, hepatic and venous blood. Gut 32, 1221–1227. doi: 10.1136/gut.32.10.1221

Ding, T., and Schloss, P. D. (2014). Dynamics and associations of microbial community types across the human body. Nature 509, 357–360. doi: 10.1038/nature13178

Donaldson, G. P., Lee, S. M., and Mazmanian, S. K. (2016). Gut biogeography of the bacterial microbiota. Nat. Rev. Microbiol. 14, 20–32. doi: 10.1038/nrmicro3552

Decottignies, A., Collins, L. B., Wali, A., Bigler, R., Sun, W., and Bullman, S. J. (2012). The Warburg effect dictates the mechanism of butyrate-mediated histone acetylation and cell proliferation. Mol. Cell 48, 612–626. doi: 10.1016/j.molcel.2012.08.033

D’Souza, G., Shitut, S., Preussger, D., Youssif, G., Waschina, S., and Kost, C. (2018). Ecology and evolution of metabolic cross-feeding interactions in bacteria. Nat. Prod. Rep. 35, 455–488. doi: 10.1039/c8np00096c

Dupaul-Chicoine, J., Yeretzian, G., Doiron, K., Bergstrom, K. S. R., McNtclere, C. R., LeBlanc, P. M., et al. (2010). Control of intestinal homeostasis, colitis, and colitis-associated colorectal cancer by the inflammatory caspases. Immunity 32, 367–378. doi: 10.1016/j.immuni.2010.02.012

Doré, J., Mullanot, M.-C., Béhier, J.-M., and participants of Giens Xxii, Round Table No. 2. (2017). The human gut microbiome as source of innovation for health: which physiological and therapeutic outcomes could we expect? Therapie 72, 21–38. doi: 10.1016/j.therap.2016.12.007

Faci, J. L., Felipe, J., de, S., Pral, L. P., da Silva, B. K., Corrêa, R. O., et al. (2019). Butyrate protects mice from clostridium difficile-induced colitis through an HIF-1-dependent mechanism. Cell Rep. 27, 750.e7–761.e7. doi: 10.1016/j.celrep.2019.03.034

Falk, K.-S., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., et al. (2012). Microbial co-occurrence relationships in the human microbiome. PLoS Comput. Biol. 8:e1002660. doi: 10.1371/journal.pcbi.1002660

Favier, C. F., de Vos, W. M., and Akkermans, A. D. L. (2003). Development of bacterial and bifidobacterial communities in feces of newborn babies. Anaerobe 9, 219–229. doi: 10.1016/j.anaerobe.2003.07.001

Fisher, C. K., and Mehta, P. (2014). Identifying keystone species in the human gut microbiome from metagenomic timeseries using sparse linear regression. PLoS One 9:e102451. doi: 10.1371/journal.pone.0102451

Foley, M. H., O’Flaherty, S., Allen, G., Rivera, A. J., Stewart, A. K., Barrangou, R., et al. (2021). Lactate/sodium salt hydrolase substrate specificity governs bacterial fitness and host colonization. Proc. Natl. Acad. Sci. U.S.A. 118:2017799118. doi: 10.1073/pnas.2017799118

Freeland, K. R., and Wolfever, T. M. S. (2010). Acute effects of intravenous and rectal acetate on glucagon-like-peptide-1, peptide YY, ghrelin, adiponectin and tumour necrosis factor-alpha. Br. J. Nutr. 103, 460–466. doi: 10.1017/S0007114509991863

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. PLoS Comput. Biol. 8:e1002687. doi: 10.1371/journal.pcbi.1002687

 Frioux, C., Singh, D., Korcsmaros, T., and Hildebrand, F. (2020). From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes. Comput. Struct. Biotechnol. J. 18, 1722–1734. doi: 10.1016/j.csbj.2020.06.028

Furusawa, Y., Obata, Y., Fukuda, S., Endo, T. A., Nakato, G., Takahashi, D., et al. (2013). Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. Nature 504, 446–450. doi: 10.1038/nature12721

Gao, J., Xu, K., Liu, H., Liu, G., Bai, M., Peng, C., et al. (2018). Impact of the gut microbiota on intestinal immunity mediated by tryptophan metabolism. Front. Cell. Infect. Microbiol. 8:13. doi: 10.3389/fcimb.2018.00013

Ghia, J.-E., Li, N., Wang, H., Collins, M., Deng, Y., El-Sharkawy, R. T., et al. (2009). Serotonin has a key role in pathogenesis of experimental colitis. Gastroenterology 137, 1649–1660. doi: 10.1053/j.gastro.2008.09.041

Gibbons, S. M. (2020). Ketoacids indispensable for microbiome recovery. Nat. Microbiol. 5, 1067–1068. doi: 10.1038/s41564-020-0783-0

Gomez de Agüero, M., Ganal-Vonarburg, S. C., Fuhrer, T., Rupp, S., Uchimura, Y., Li, H., et al. (2016). The maternal microbiota drives early postnatal innate immune development. Science 351, 1296–1302. doi: 10.1126/science.aad2571

Goodchild, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., et al. (2014). Human genetics shape the gut microbiome. Cell 159, 789–799. doi: 10.1016/j.cell.2014.09.053

Gotoh, A., Ojima, M. N., and Katayama, T. (2019). Minority species influences microbiota formation: the role of Bifidobacterium with extracellular glycosidases in bifidus flora formation in breastfed infant guts. Microbiol. Biotechnol. 12, 259–264. doi: 10.1111/1751-7915.13366
Lamas, B., Hernandez-Galan, L., Galipeau, H. J., Constante, M., Clarizio, A., Jury, J., et al. (2020). Aryl hydrocarbon receptor ligand production by the gut microbiota is decreased in celiac disease leading to intestinal inflammation. *Sci. Transl. Med.* 12, eabk06824. doi: 10.1126/scitranslmed.abk06824

Lamas, B., Richard, M. L., Leducq, V., Pham, H.-P., Michel, M.-L., Da Costa, G., et al. (2016). CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat. Med.* 22, 598–605. doi: 10.1038/nm.4102

Layeghifard, M., Hwang, D. M., and Guttman, D. S. (2017). Disentangling interactions in the microbiome: a network perspective. *Trends Microbiol.* 25, 217–228. doi: 10.1016/j.tim.2016.11.008

Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* 490, 45–48. doi: 10.1038/nature11234

Leylabado, H. E., Ghotaslou, F., Reizabadi, M. M., Farajnia, S., Moaddab, S. Y., Ganbarov, K., et al. (2020). The critical role of Fecalibacterium prausnitzii in human health: an overview. *Microb. Pathog.* 149:104344. doi: 10.1016/j.micpath.2020.104344

Li, C., Ching, K. R., Kwah, J. S., Av-Shalom, T. V., Tucker-Kellogg, L., and Nagarajan, N. (2019). An expectation-maximization algorithm enables accurate ecological modeling using longitudinal microbiome sequencing data. *Microbiome* 7:118. doi: 10.1186/s40168-019-0729-9

Li, J., Doty, A., and Glover, S. C. (2016). Aryl hydrocarbon receptor signaling involves in the human intestinal ILC3/IIF1C1 conversion in the inflamed terminal ileum of Crohn’s disease patients. *Inflamm. Cell. Signal.* 3:e1404. doi: 10.14800/ics.1404

Li, J., Zhao, F., Wang, Y., Chen, J., Tao, J., Tian, G., et al. (2017). Gut microbiota dysbiosis contributes to the development of hypertension. *Microbiome* 5:14. doi: 10.1186/s40168-016-0022-x

Li, S., Bostick, J. W., Ye, J., Qiu, J., Zhang, B., Urban, J., F. et al. (2018). Aryl hydrocarbon receptor signaling cell intrinsically inhibits intestinal group 2 innate lymphoid cell function. *Immunity* 49, 915.e5–928.e5. doi: 10.1016/j.immuni.2018.09.015

Liu, F., Li, Z., Wang, X., Xue, C., Tang, Q., and Li, R. W. (2019). Microbial co-occurrence patterns and keystone species in the gut microbiota community of mice in response to stress and chondroitin sulfate disaccharide. *Int. J. Mol. Sci.* 20:43. doi: 10.3390/ijms20092130

Lofus, M., Hassouneh, S. A., and Yoseph, S. (2021). Bacterial associations in the healthy human gut microbiome across populations. *Sci. Rep.* 11:2828. doi: 10.1038/s41595-021-82449-0

Lousi, P., Young, P., Holtrop, G., and Flint, H. J. (2010). Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl-CoA:acetate CoA-transferase gene. *Environ. Microbiol.* 12, 304–314. doi: 10.1111/j.1462-2920.2009.02066.x

Lozupone, C., and Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005

Lugli, G. A., Duranti, S., Milani, M., Mancabelli, L., Turroni, F., Alessandri, G., et al. (2020). Investigating bifidobacteria and human milk oligosaccharide composition of lactating mothers. *FEMS Microbiol. Ecol.* 96:faa049. doi: 10.1093/femsec/fiaa049

Macia, L., Tan, J., Vieira, A. T., Leach, K., Stanley, D., Luong, S., et al. (2015). Metabolite-sensing receptors GPR43 and GPR109A facilitate dietary fibre-induced gut homeostasis through regulation of the inflammasome. *Nat. Commun.* 6:6734. doi: 10.1038/ncomms6734

Mainali, K. P., Bewick, S., Thielen, P., Mehoke, T., Breitwieser, F. P., Paudel, S., et al. (2017). Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. *PLoS One* 12:e0181732. doi: 10.1371/journal.pone.0181732

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn’s disease revealed by a metagenomic approach. *Gut* 55, 205–211. doi: 10.1136/gut.2005.073817

Marafini, I., Di Fusco, D., Dinallo, V., Franzè, E., Stolfi, C., Sica, G., et al. (2019). NPD-0414-2 and NPD-0414-24, two chemical entities designed as aryl hydrocarbon receptor (AhR) ligands, inhibit gut inflammatory signals. *Front. Pharmacol.* 10:2130. doi: 10.3389/fphar.2019.002130

Marfani, I., Di Fusco, D., Dinallo, V., Franzè, E., Stolfi, C., Sica, G., et al. (2019). NPD-0414-2 and NPD-0414-24, two chemical entities designed as aryl hydrocarbon receptor (AhR) ligands, inhibit gut inflammatory signals. *Front. Pharmacol.* 10:2130. doi: 10.3389/fphar.2019.002130

Mainali, K. P., Bewick, S., Thielen, P., Mehoke, T., Breitwieser, F. P., Paudel, S., et al. (2017). Statistical analysis of co-occurrence patterns in microbial presence-absence datasets. *PLoS One* 12:e0181732. doi: 10.1371/journal.pone.0181732

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn’s disease revealed by a metagenomic approach. *Gut* 55, 205–211. doi: 10.1136/gut.2005.073817

Marfani, I., Di Fusco, D., Dinallo, V., Franzè, E., Stolfi, C., Sica, G., et al. (2019). NPD-0414-2 and NPD-0414-24, two chemical entities designed as aryl hydrocarbon receptor (AhR) ligands, inhibit gut inflammatory signals. *Front. Pharmacol.* 10:2130. doi: 10.3389/fphar.2019.002130

Mawé, G. M., and Hofman, J. M. (2013). Serotonin signalling in the gut—functions, dysfunctions and therapeutic targets. *Nat. Rev. Gastroenterol. Hepatol.* 10, 473–486. doi: 10.1038/nrgastro.2013.105

Mazier, W., Le Corf, K., Martinez, C., Tudela, H., Kissi, D., Kropp, C., et al. (2021). A new strain of *Christensenella minuta* as a potential biotherapy for obesity and associated metabolic diseases. *Cells* 10:823. doi: 10.3390/cells10040823
McDonald, D., Vázquez-Baeza, Y., Koslicki, D., McClelland, J., Reeve, N., Xu, Z., et al. (2018). Striped UniFrac: enabling microbiome analysis at unprecedented scale. Nat. Methods 15, 847–848. doi: 10.1038/s41592-018-0187-8

Millard, A. L., Mertes, P. M., Ittelet, D., Villard, F., Jeannesson, P., and Bernard, J. (2002). Butyrate affects differentiation, maturation and function of human monocyte-derived dendritic cells and macrophages. Clin. Exp. Immunol. 130, 245–255. doi: 10.1046/j.1090-9940.2002.01977.x

Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, B., Bang, C., and Schmitz, R. A. (2018). Archaea are interactive components of complex microbiomes. Trends Microbiol. 26, 70–85. doi: 10.1016/j.tim.2017.07.004

Morotomi, M., Nagai, F., and Watanabe, Y. (2012). Description of Christensenellae minuta gen. nov., sp. nov., isolated from human faeces, which forms a distinct branch in the order Clostridiales, and proposal of Christensenellaceae fam. nov. Int. J. Syst. Evol. Microbiol. 62, 144–149. doi: 10.1099/ijs.0.026999-0

Muñoz-Tamayo, R., and Bassler, B. L. (2019). Bacterial quorum sensing in complex and dynamically changing environments. Nat. Rev. Microbiol. 17, 371–382. doi: 10.1038/s41579-018-0186-5

Mullish, B. H., McDonald, J. A. K., Peclhivianis, A., Allegretti, J. R., Kao, D., Barker, G. F., et al. (2019). Microbial bile salt hydrolyses mediate the efficacy of faecal microbiota transplant in the treatment of recurrent Clostridiodes difficile infection. Gut 68, 1791–1800. doi: 10.1136/gutjnl-2018-317842

Muñoz-Tamayo, R., Laroche, B., Walter, E., Doré, J., Duncan, S. H., Flint, H. J., et al. (2019). rpoB, a promising marker for analyzing the diversity of bacterial communities by multivariate analyses in microbial ecology. FEMS Microbiol. Ecol. 947. doi: 10.1093/fmedef/fmz043

Ogier, J.-C., Pagès, S., Galan, M., Barret, M., and Gaudriault, S. (2019). rpoB, a gene of the bacterial rpoB operon, is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of mice. Gut Microb. 7, 284–291. doi: 10.1080/19490976.2015.1127483

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: advantages of whole genome shotgun versus 16S ampiclon sequencing. Biochem. Biophys. Res. Commun. 469, 967–977. doi: 10.1016/j.bbrc.2015.12.083

Rao, A. S., Mann, T. S., and Sexton, D. (2017). The gut microbiota as a host defender and a foe: Consequences of bile salt biotransformations by intestinal bacteria. Gut. 72, 22–39. doi: 10.1136/gutjnl-2015-311270

Rigaud, C., S. Almendros, C., Ramírez, J. F., Szurzowski, J. H., Linden, D. R., Sonnenburg, J. L., et al. (2015). Gut microbes promote colonic serotonin production through an effect of short-chain fatty acids on enterochromaffin cells. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 29, 1395–1403. doi: 10.1096/fj.14-25998

Ridlon, J. M., Harris, S. C., Bhowmik, S., Kang, D.-J., and Hylemon, P. B. (2016). Consequences of bile salt biotransformations by intestinal bacteria. Gut Microb. 7, 22–39. doi: 10.1080/19490976.2015.1127483

Röyttä, L., Vandeputte, D., Raes, J., and Faust, K. (2021). Null-model-based network comparison reveals core associations. ISME J. 15, 360–370. doi: 10.1038/s41396-020-0668-2

Rosenberg, M. D., Bickerton, A. S., Dennis, A. L., Vidal, H., and Frayn, K. N. (2005). Insulin-sensitizing effects of dietary resistant starch and effects on skeletal muscle and adipose tissue metabolism. Am. J. Clin. Nutr. 82, 559–567. doi: 10.1093/ajcn/82.3.559

Sandor, J., Fodor, N., Schell, K., Kolesar, S., Muller, I., Kav, S., et al. (2020). Syntrophy via interspecies H2 transfer between christensenellae and methanobrevibacter underlies their global cooccurrence in the human gut. mbio 11, e03235–19. doi: 10.1128/mbio.03235-19

Samuel, B. S., Shaito, A., Motoi, T., Rey, F. E., Backhed, F., Manchester, J. K., et al. (2018). Effects of the gut microbiota on host adiposity are modulated by the short-chain fatty-acid binding G protein-coupled receptor, Gpr41. Proc. Natl. Acad. Sci. U.S.A. 105, 16767–16772. doi: 10.1073/pnas.1804057105

Schepbach, W., Sommer, H., Kirchner, T., Paganelli, G. M., Bartram, P., Christl, S., et al. (2019). Effect of butyrate enemas on the colonic mucosa in distal ulcerative colitis. Gastroenterology 103, 51–56. doi: 10.1016/j.gastro.2019.02.005

Schirmer, M., Garner, A., Vlamakis, H., and Xavier, R. J. (2019). Microbial genes and pathways in inflammatory bowel disease. Nat. Rev. Microbiol. 17, 497–511. doi: 10.1038/s41579-019-0213-6

Schmidt, T. S., Hayward, M. R., Coelho, L. P., Li, S. S., Costea, P. I., Voigt, A. Y., et al. (2019). Extensive transmission of microbes along the gastrointestinal tract. eLife 8:e42693. doi: 10.7554/eLife.42693

Sender, R., Fuchs, S., and Milo, R. (2016). Revised estimates for the number of human and bacteria cells in the body. PLoS Biol. 14, e1002533. doi: 10.1371/journal.pbio.1002533

Singh, N., Gurav, A., Swapprakasam, S., Brady, E., Padia, R., Shi, H., et al. (2014). Activation of Gpr109a receptor for niacin and the commensal metabolite butyrate suppresses colonic inflammation and carcinogenesis. Immunity 40, 128–139. doi: 10.1016/j.immuni.2013.12.007

Singh, N., Thangaraju, M., Prasad, P. D., Martin, P. M., Lambert, N. A., Boettger, T., et al. (2010). Blockade of dendritic cell development by bacterial fermentation products butyrate and propionate through a transporter (Slc5a8)-dependent inhibition of histone deacetylases. J. Biol. Chem. 285, 27601–27606. doi: 10.1074/jbc.M110.110294

Singh, P., Chandrasekharappa, S., Bodduluri, S. R., Hegde, B., Kotla, N. G., et al. (2019). Enhancement of the gut barrier integrity by a microbial metabolite through the Nr2f2 pathway. Nat. Commun. 10:89. doi: 10.1038/s41467-018-07859-7

Spohn, S. N., and Mawe, G. M. (2017). Non-conventional features of peripheral serotonin signalling - the gut and beyond. Nat. Rev. Gastroenterol. Hepatol. 14, 412–420. doi: 10.1038/nrgastro.2017.51

Tudela et al. Identifying Keystone Species Missing Functions
Sumara, G., Sumara, O., Kim, J. K., and Karsenty, G. (2012). Gut-derived serotonin is a multifunctional determinant to fasting adaptation. Cell Metab. 16, 588–600. doi: 10.1016/j.cmet.2012.09.014

Yadav, D., Dutta, A., and Mande, S. S. (2019). OTUX: V-region specific OTU picking and efficient cross-study taxonomic comparison of microbiomes. DNA Res. Int. J. Rapid Publ. Rep. Genes Genomes 26, 147–156. doi: 10.1093/dnares/dsy045

Wang, Y., Yu, T., Huang, X., Bilotta, A. J., Xu, L., Lu, Y., et al. (2020). Intestinal microbiota-derived short-chain fatty acids regulation of immune cell IL-22 production and gut immunity. Nat. Commun. 11:4457. doi: 10.1038/s41467-020-18262-6

Yano, J. M., Yu, K., Donaldson, G. P., Shastri, G. G., Ann, P., Ma, L., et al. (2015). Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. Cell 161, 264–276. doi: 10.1016/j.cell.2015.02.047

Yao, L., Seaton, S. C., Noudsse-Fetter, S., Adhikari, A. A., DiBenedetto, N., Mina, A. I., et al. (2018). A selective gut bacterial bile salt hydrolase alters host metabolism. eLife 7:e37182. doi: 10.7554/eLife.37182

Yassour, M., Lim, M. Y., Yun, H. S.,Tickle, T. L., Sung,J., Song, Y.-M., et al. (2016). Sub-clinical detection of gut microbial biomarkers of obesity and type 2 diabetes. Genome Med. 8:17. doi: 10.1186/s13073-016-0271-6

Yu, J., Qiu, J., Bostick, J. W., Ueda, A., Schijven, H., Li, S., et al. (2017). The aryl hydrocarbon receptor preferentially marks and promotes gut regulatory T cells. Cell Rep. 21, 2277–2290. doi: 10.1016/j.celrep.2017.10.114

Yu, M., Wang, Q., Ma, Y., Li, L., Yu, K., Zhang, Z., et al. (2018). Aryl hydrocarbon receptor activation modulates intestinal epithelial barrier function by maintaining tight junction integrity. Int. J. Biol. Sci. 14, 69–77. doi: 10.7150/ijbs.22259

Yu, Z.-T., Chen, C., Kling, D. E., Liu, B., McCoy, J. M., Merighi, M., et al. (2013). The principal fusocylated oligosaccharides of human milk exhibit prebiotic properties on cultured infant microbiota. Glycobiology 23, 169–177. doi: 10.1093/glycob/cws138

Ze, X., Duncan, S. H., Louis, P., and Flint, H. I. (2012). Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. ISME J. 6, 1535–1543. doi: 10.1038/ismej.2012.4

Zelante, T., Iannitti, R. G., Cunha, C., De Luca, A., Giovannini, G., Pieraccini, G., et al. (2013). Tryptophol catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. Immunity 39, 372–385. doi: 10.1016/j.immuni.2013.08.003

Zhang, Z., Geng, J., Tang, X., Fan, H., Xu, J., Wen, X., et al. (2014). Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. ISME J. 8, 881–893. doi: 10.1038/ismej.2013.185

Zou, Y., Xue, W., Luo, G., Deng, Z., Qin, P., Ruijin, G., et al. (2019). 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses. Nat. Biotechnol. 37:179. doi: 10.1038/s41587-018-0008-8

Conflict of Interest: SC and HT are employees of YSOPIA Bioscience.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Tudela, Claus and Saleh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.