Cell Host & Microbe

CellPress

Commentary

The human microbiome: A coming of age story

Aleksandar D. Kostic[1,2,*]
[1]Joslin Diabetes Center, Boston, MA, USA
[2]Department of Microbiology, Harvard Medical School, Boston, MA, USA
*Correspondence: aleksandar.kostic@joslin.harvard.edu
https://doi.org/10.1016/j.chom.2022.03.020

The human microbiome field is coming of age, but it is still defining itself. I can say the same as an investigator who started his career in the early days of this expanding field. This commentary reflects on my *Cell Host & Microbe* papers along this journey that captured the field's progress.

> Nothing in Biology Makes Sense Except in the Light of Evolution—Theodosius Dobzhansky (*The American Biology Teacher*)

A respected senior colleague at the Broad Institute where I did my training once asked me why I was so "gung-ho" on the microbiome. He asked this question with sincerity, commenting that as an outsider it seemed a kind of "peripheral system" to human health, and why not rather focus my efforts on something more central and therefore with higher potential impact, in his estimation. I remember fumbling through my answer—something about how microbiome composition is correlated with so many human diseases or some other boring fact along those lines. My microbiome colleagues have likely been asked some form of this question at some point in their careers. Different researchers have different answers, but for me, after this question made me sit down and ponder it awhile, I realized my motivation is rather simple: we understand what life is like without a microbiome—it's not good. *But* we hardly have a clue as to *how* the microbiome supports our health or *why* mammals and their microbiomes have settled on this mutualistic arrangement. The microbiome field was borne in part by the extraordinary efforts of James Reyniers and colleagues who delivered the world's first fully sterile mammals in the 1940s. Although these animals survived, they had remarkably few leukocytes, antibodies, and lymphatic tissues and an astoundingly large cecum. Today, we understand that germ-free mice not only have many gross anatomical abnormalities, but every major physiologic system is affected: they must eat twice as many calories as their conventional counterparts just to maintain their body weight; they are more prone to mortality from infectious disease and autoimmune diseases; and they have extreme neurologic and—as anyone has tried to handle them knows—behavioral abnormalities. What's even more fascinating is that upon colonization with a mouse microbiome at appropriate developmental time windows, almost all these phenotypes can be rescued (in the classical genetic sense). In Figure 1, I summarize just a small subset of published phenotypes that can be rescued by colonizing germ-free mice with typical lab mouse gut microbiomes. We have barely scratched the surface on a mechanistic understanding of how the microbiome accomplishes these incredible feats and why, from an evolutionary perspective, this microbe-mammal symbiosis has come to be.

So what?! Unearthing the biology behind microbiome-disease correlations

The first time I glimpsed the ecology of the human colon—and I do realize how silly this sounds—I felt like I had stumbled upon a direct channel to nature's secrets. I was a graduate student in Matthew Meyerson's lab, and we were searching for novel cancer-associated pathogens by "computationally subtracting" human DNA from random shotgun sequencing of resected human tumors and adjacent normal tissues. We were conducting this on some of the first whole genomes of cancer tissues ever sequenced. This strategy successfully discovered new associations between cancers and known and novel microbes. Though most of the resected tissues were dotted with the rare microbe here and there (mostly viruses), colorectal tissue was in a different league. Not only were these tissues teeming with microbes, but there was also an incredible order to the ecosystem that to my naive mind should have been a chaotic concoction of more-or-less random bacteria. Of course, this is something Jeff Gordon and colleagues at Washington University had been describing for several years before I even touched microbiome data. But nonetheless, there's something exciting about rediscovering it for yourself in your own data. The normal colon was dominated by the Bacteroidetes and/or Firmicutes phyla, depending on the person, and there was a large but seemingly finite list of species present across individuals. The tumors and adjacent normal colonic tissues from the same person were just about identical in their composition statistically—with one major exception. The genus *Fusobacterium* was several-fold more abundant in tumors relative to matched normal tissue, and it was present in the tumors of *most* people. By microbiome standards, these were screaming signals. Rob Holt's team at the University of British Columbia independently made the same discovery using tumor transcriptomics data, and we published our stories back-to-back (Castellarin et al., 2012; Kostic et al., 2012). In my naiveté, what we had discovered felt as significant as Barry Marshall and Robin Warren's discovery that *Helicobacter pylori* caused gastric ulcers, which won them a Nobel Prize. I remember presenting a short talk at the Cold Spring Harbor Laboratory Cancer Genomics conference before our paper was published, and the feeling of intimidation as the larger-than-life portrait of James Watson looked down at me as I was presenting. Despite

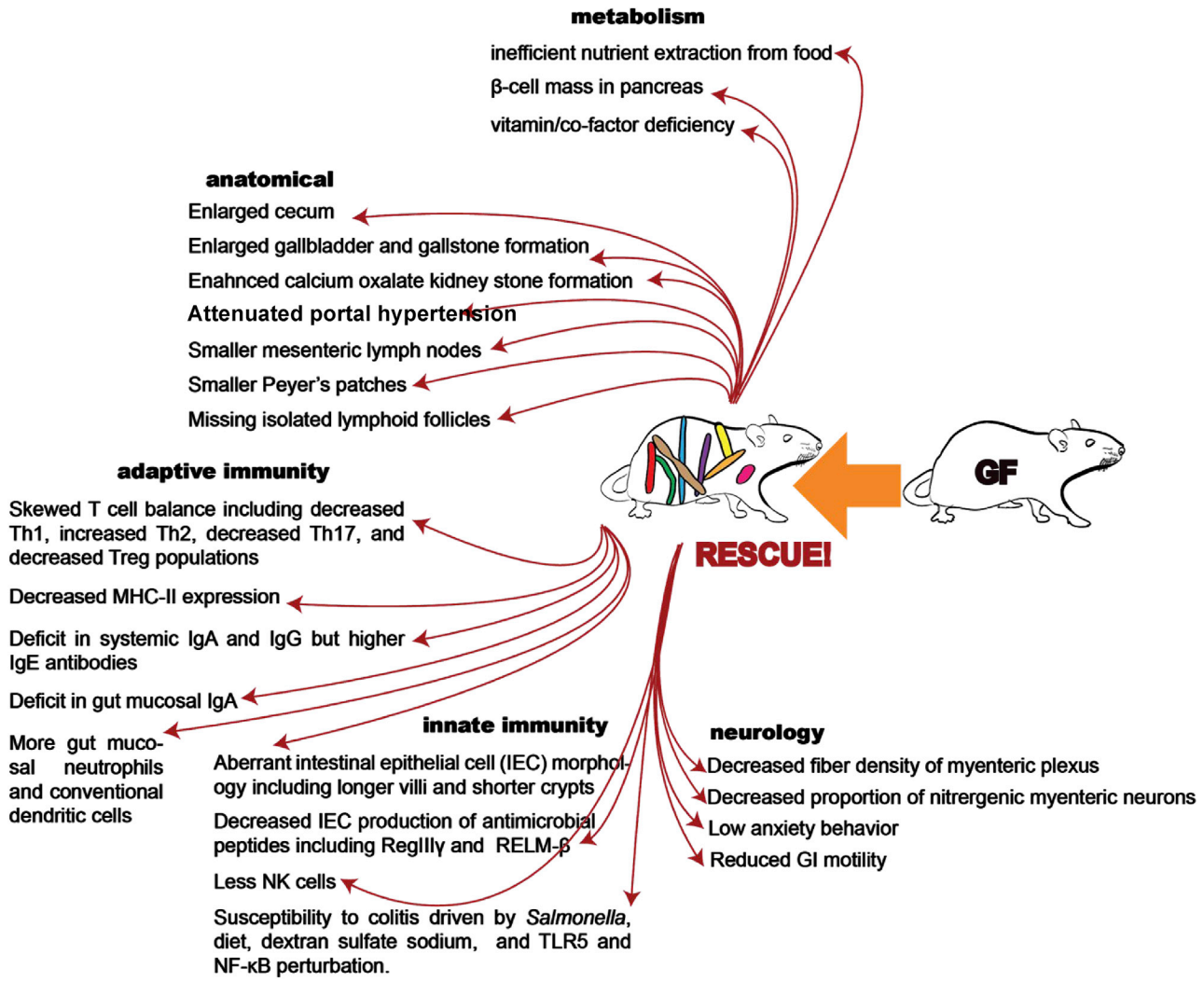

Figure 1. Select germ-free phenotypes that can be at least partially rescued upon colonization with a mouse microbiome

my nerves, I thought the talk went rather well, until receiving skepticism in the questions from the audience and later learning that a member of the audience was not-so-quietly exclaiming: "So what??!!"

My heckler was right, of course. There is a well-trodden process to determining whether a microbial agent causes disease: Koch's postulates. Essentially, upon isolation the agent should recapitulate the disease when administered to a naive animal model. The major difference, as I saw it then, between our *Fusobacterium* work and the *H. pylori*-garstric ulcer work was that Barry Marshall actually *did* the causality experiment—with full fanfare—by swallowing a flask of *H. pylori* and demonstrating that he developed stomach ulcers. I was dedicated to my science, but not that dedicated. I was incredibly fortunate to have been welcomed into the lab of Wendy Garrett, where I completed my last several years of graduate school trying to answer the causality question. Rather than drink the stuff, we decided we would colonize a mouse model of intestinal cancer with human isolates of *Fusobacterium nucleatum* and track their intestinal tumors in reference to mice colonized with another colorectal cancer-associated microbe, *Streptococcus sanguinis*. This required daily inoculations with densely grown cultures of these bacteria for 8 weeks each—an experiment that took nearly 2 years to complete. And now my own students enjoy being regaled with tales of how I would trudge through snowstorm or hurricane, weekend, and holiday to see my experiments to completion. *Almost* as dedicated as Barry Marshall. We were excited to find that the *F. nucleatum*-colonized mice relative to controls had significantly accelerated tumorigenesis and larger, but not more numerous, intestinal tumors, which was consistent with the hypothesis of a role for Fusobacteria in the progression, though perhaps not initiation, of colorectal tumorigenesis (Kostic et al., 2013). This was an important step toward closing the causal loop. We also showed that both human and mouse *F. nucleatum*-associated tumors had a much stronger inflammatory signature relative to Fuso-negative tumors, which we believe is important to tumor progression.

It has been among the most fulfilling aspects of my scientific career to observe the incredible work that has been done by so many to further understand the mechanisms and the clinical consequences of the role of *F. nucleatum* in colorectal cancer (reviewed in Brennan and Garrett, 2019). I must admit to expecting that as we study the microbiome of other diseases, we would similarly find a "smoking gun" microbe like *F. nucleatum* and colon cancer. I would be proved wrong.

So long, Koch's postulates: Complex diseases have a complex etiology

To determine causality, temporality matters—the abundance of a microbe associated with a diseased condition does not imply causation. For example, there is a good chance that the microbe simply thrives in the microenvironment of diseased tissue. Finding the microbe or microbial feature (i.e., taxon, metabolic pathway, gene family, etc.) *prior* to disease onset is critical in attempting to find a "trigger event." Unfortunately, prospective, longitudinal disease cohorts that include stool samples are incredibly rare. So, I was extremely fortunate, having begun my postdoctoral position co-mentored by Ramnik Xavier and Curtis Huttenhower, that they had exactly such a collection ready and waiting for analysis, the product of a collaboration forged with Mikael Knip who directs the DIABIMMUNE Consortium. This effort collected and banked fresh stool samples, from birth to 3 years of age at approximately monthly intervals, from thousands of infants at high-risk for type 1 diabetes (T1D) from Finland, which has the highest incidence of T1D of any country. This was a microbiome cohort without parallels in any disease and by far the best chance at finding a "trigger" for T1D in the microbiome, if one existed. We conducted metagenomic and metabolomics analysis on a subset of children that did and did not go on to develop T1D within a 5-year time frame. We observed that children who did develop T1D exhibited a significantly lower microbiome diversity—even more fascinating, we observed it more than 1 year *prior* to clinical diagnosis (Kostic et al., 2015). This was an important clue that microbiome data might have great utility in predicting complex diseases like T1D. We further identified specific inflammation-associated metabolites and bacteria, such as *Ruminococcus gnavus*, to not only be of higher abundance in the T1D group but also to bloom in abundance just prior to clinical diagnosis, which might explain the overall drop in microbiome diversity in these patients. This work painted a picture of the T1D-associated microbiome as one that "adjusts the thermostat" on gut mucosal inflammation, such that in children with genetic risk for the disease, all other things being equal, it further increases their odds of autoreactivity going full-throttle into T1D. Around this time, similar conclusions were being reached in other complex diseases such as obesity and inflammatory bowel diseases. So much for the one-bug, one-disease model and Koch's postulates!

In a subsequent paper we compared the high-risk Finnish children to those in neighboring Russian Karelia, who have a 6-fold lower incidence in T1D. National T1D incidence, as with incidence of autoimmune and allergic diseases in general, is correlated with gross domestic product. This is thought to be explained by the "hygiene hypothesis," postulating that a more industrialized and urbanized lifestyle hinders early life immune development because it limits exposures to natural microbial compounds (von Mutius, 2007). Although the gut microbiome of children in the Finnish versus Karelian cohort were not remarkably different after 1 year of age, they looked just about worlds apart in the first 12 months. In particular, we found that *Bifidobacterium* species, a key human milk oligosaccharide (HMO) utilizer, was much more abundant in the Karelian children (Vatanen et al., 2016), whereas the Finnish children were largely colonized with a different class of HMO utilizing bacteria, the *Bacteroides*, and specifically *B. dorei*. In contrast, the abundance of another key Gram-negative microbe, *E. coli*, was roughly the same in both country cohorts. When we looked at the ratio of *Bacteroides* to *E. coli*, a greater than 10-fold ratio in the Finnish cohort led us to hypothesize that overwhelming abundance of *Bacteroides* in the Finns dampened early-life immune education. The Gram-negative cell membrane component lipopolysaccharide (LPS) from *E. coli* acts as an inflammatory molecule (via its recognition on cells by TLR4), but other Gram-negatives such as the *Bacteroides* do not trigger inflammation. Thus, we hypothesized and demonstrated that the lipid A moiety of LPS from *Bacteroides* species actually serves as a TLR4 antagonist and prevents TLR4 activation by lipid A from *E. coli*. Furthermore, administration of purified *E. coli* lipid A to the non-obese diabetic murine model of T1D significantly delayed diabetes incidence relative to mice that received *B. dorei* lipid A (Vatanen et al., 2016). Conceptually, this work was an important demonstration of how the functional role of the human microbiome in a complex human disease, with the help of appropriate comparison groups, large cohorts, and sample sizes, could be distilled into discrete, reductionist hypotheses that push our mechanistic understanding forward.

So…what *is* the human microbiome, exactly?

Where did it come from?

Shortly after completing these studies I began my independent lab in the incredible environments at Joslin Diabetes Center and the Microbiology Department at Harvard Medical School. The question foremost on my mind, central to both host-microbiome symbiosis and its role in diabetes following my previous work, was: "what *is* the human microbiome"? How is it possible that the gut microbiome of two groups of people living just a few hundred miles apart were so different, particularly in infancy? Can it be said that the microbiome of more industrialized and urbanized people is "less adept" at promoting human health? I felt the importance of these questions because I've seen the stark microbiome differences in my own data, but I would not have been able to even pose the question were it not for the incredible work of Martin Blaser and his theory of "Missing Microbes." This theory poses that members of the human microbiome are "disappearing" as humans become more industrialized (Blaser, 2018). Indeed, around this time several studies demonstrated that the gut microbiomes of people living a subsistence farming or hunter-gatherer lifestyle had a distinct and much more diverse gut microbiome than industrialized Westerners (Brito et al., 2016; Smits et al., 2017). Identifying which gut microbes were part of human evolutionary history (or "ancestral") and

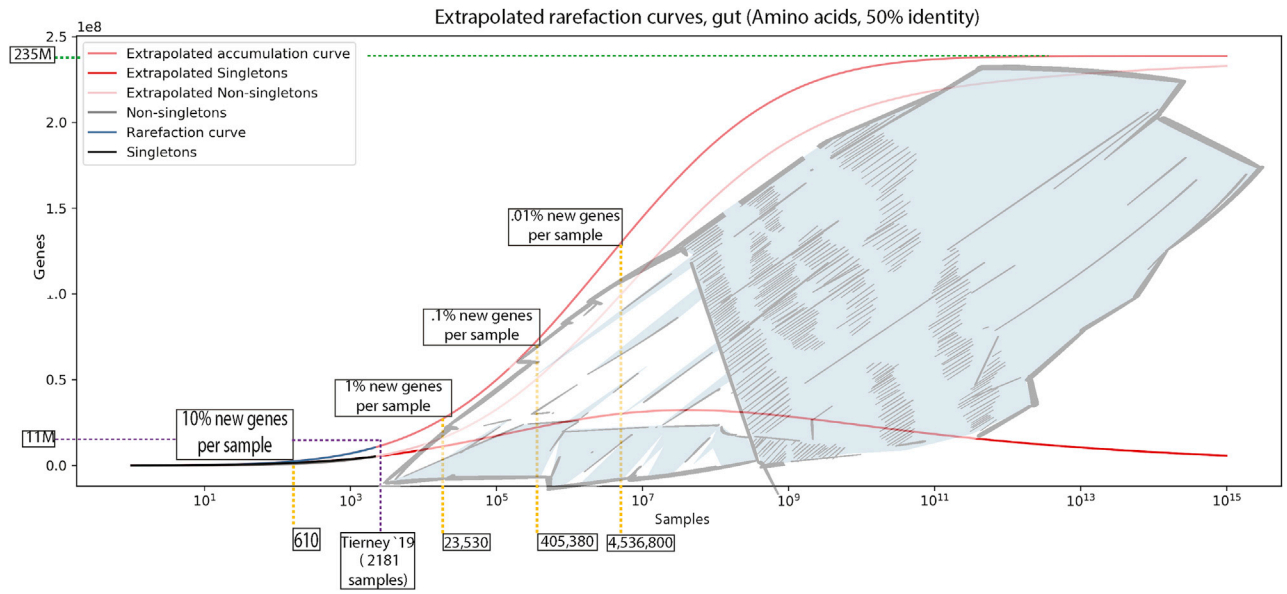

Figure 2. Extrapolating the gene content of the human microbiome
The purple dashed line marks the size of the cited study (Tierney et al., 2019). The green dashed line is the asymptotic number of genes in the human gut microbiome, by extrapolation.
Iceberg designed by rawpixel.com / Freepik.

which have become lost may be instrumental in understanding human-microbiome relationships in health and disease. Also around this time, I had been contacted by Dr. Steven LeBlanc, the retired director of collections at the Peabody Museum of Archaeology and Ethnology, who was eager to help coordinate an effort to sequence the ancient human gut microbiome. He supplied a remarkably preserved collection of paleofeces (i.e., ancient human stool samples) from different regions of North America and are 1,000 to 2,000 years old. With a highly dedicated collaborative team, we were able to extract not only high-quality DNA from these samples, but DNA fragments sufficient in size (~300bp) to enable *de novo* assembly into nearly 500 metagenome-assembled genomes (MAGs) (Wibowo et al., 2021). This was a time machine into human history that revealed nearly 40% of MAGs are novel species not seen in the modern human microbiome, in support of the theory of disappearing human microbes. In the remaining 60% of previously observed microbial species, the ancient gut microbiome is much more similar to that of modern, non-industrialized people than that of industrialized people. In other words, these ancient microbiomes share features (i.e., species and genes) with the gut microbiomes of subsistence farmers in Fiji, hunter-gatherers in Tanzania, and 2,000-year-old Aztecs and are conspicuously absent or very rare in modern Americans. Therefore, these studies support the existence of a shared, "ancestral" human gut microbiome from which our industrialized Westerner gut microbiome has substantially diverged.

What is its genetic "parts-list"?

A major turning point in human genetics was the completion of the human genome in 2003. The "complete parts-list" of human genetic elements put efforts to characterize gene functions into warp speed. It's possible that a "complete gene catalog" of the human microbiome could do the same for our understanding of host-microbiome interactions. As metagenomic sequencing was expanding, datasets were being deposited into public collections, and MAGs began to emerge, it seemed that there was an opportunity to build a comprehensive database of genes from the human microbiome that might near "completeness." But given inter-individual variability in microbiome composition and the dynamic nature of prokaryotic genomes, how many samples would it take to achieve this? In our 2019 *Cell Host & Microbe* paper, we performed *de novo* gene discovery on nearly 4,000 human gut and oral metagenomes that resulted in a catalog of 46 million non-redundant (nr) genes (clustered at 95% nucleotide identity) (Tierney et al., 2019). The precise number of nr genes is almost irrelevant because it changes with the clustering threshold used to define a "gene." What we discovered to be the most insightful metric was not a number but a ratio. Across all samples we found that many genes were "singletons"; in other words, they were only observed in a single sample. The ratio of singletons-to-non-singletons (*S/NS*) was approximately one in our sample set. This means that of 46 million genes, *half* were observed only *once*. And this ratio did not change significantly with the percent identity threshold used to define a gene (e.g., 50% amino acid identity). Because *S/NS* changes as a function of the number of samples, we reasoned that were it possible to sequence enough samples, we would observe all genes in the human microbiome when *S/NS* reaches 0 (that is assuming a finite number of microbiome genes, of course!). The surprising result was that although we had thousands of samples, a significant fraction of all publicly available metagenomes at the time, we were nowhere close to the asymptote and could only make a very approximate estimate of how many genes might exist (Figure 2).

This work revealed that there is a fundamental genetic diversity to the human microbiome that is all but unexplored; the genes we have observed are only the tip of the iceberg of what is encoded by the human microbiome. And it would be a mistake to assume that the small fraction of genes we have observed are somehow representative or more important or relevant than the majority we have not. This may sound counterintuitive or even outright wrong; however it begins to make sense in light of a few simple observations: (1) all bacterial species have not a single genome but a pan-genome, and for many human-associated microbes the pan-genome is open (Tettelin et al., 2005); (2) we continue to add new species to the human microbiome the more samples we sequence (Almeida et al., 2020); and (3) even the unusually high sequencing depth of 12Gb/sample in the human gut microbiome yields as little as $^{2}/_{3}$ of the total genes present because of a long tail of lowly abundant genomes (Tierney et al., 2019).

Conclusions

Looking back at these works, which are representative of the broad themes of the human microbiome field if only a meager sampling, one major conclusion jumps out. In trying to make sense of the human microbiome, it's very hard to see the forest for the trees. What we've discovered by sequencing and experimenting with the human microbiome is that it is a system so vast and so complex, it is without parallel in human biology (at least human genetics can be summed up in 46 neat chromosomes). Current human microbiome efforts have been phenomenally successful, and they should be continued and accelerated, but at the same time we are in need of an overarching theory to put all the pieces together. Key new directions include investigations into the origins of the ancestral human microbiome, how and from where fragile obligate anaerobes come to colonize humans, the degree to which fitness of microbial symbionts is pegged to the fitness of their human hosts, and whether and the extent to which co-evolution even exists in the human microbiome (current state of knowledge well reviewed in Foster et al., 2017 and Groussin et al., 2020). By beginning to answer these questions, we may be able to look back at all we have learned about the human microbiome and see it in an entirely new light.

ACKNOWLEDGMENTS

This work was supported by the American Diabetes Association Pathway to Stop Diabetes Initiator Award 1-17-INI-13 and NIH grant P30-DK036836-30. I plead forgiveness from the many colleagues whose work could not be cited because of space and reference limitations. I also thank my wife D.D. Shao, who studies the 46 chromosomes, for editing the manuscript.

DECLARATION OF INTERESTS

The author holds a patent related to the colon cancer microbiome (US11060148B2).

REFERENCES

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., et al. (2020). A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat. Biotechnol. *39*, 105–114.

Blaser, M.J. (2018). The Past and Future Biology of the Human Microbiome in an Age of Extinctions. Cell *172*, 1173–1177.

Brennan, C.A., and Garrett, W.S. (2019). Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. Nat. Rev. Microbiol. *17*, 156–166.

Brito, I.L., Yilmaz, S., Huang, K., Xu, L., Jupiter, S.D., Jenkins, A.P., Naisilisili, W., Tamminen, M., Smillie, C.S., Wortman, J.R., et al. (2016). Mobile genes in the human microbiome are structured from global to individual scales. Nature *535*, 435–439.

Castellarin, M., Warren, R.L., Freeman, J.D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R.A., and Holt, R.A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. *22*, 299–306.

Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. Nature *548*, 43–51.

Groussin, M., Mazel, F., and Alm, E.J. (2020). Co-evolution and Co-speciation of Host-Gut Bacteria Systems. Cell Host Microbe *28*, 12–22.

Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., et al. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. *22*, 292–298.

Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., Clancy, T.E., Chung, D.C., Lochhead, P., Hold, G.L., et al. (2013). Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe *14*, 207–215.

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al.; DIABIMMUNE Study Group (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *17*, 260–273.

Smits, S.A., Leach, J., Sonnenburg, E.D., Gonzalez, C.G., Lichtman, J.S., Reid, G., Knight, R., Manjurano, A., Changalucha, J., Elias, J.E., et al. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science *357*, 802–806.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., et al. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. USA *102*, 13950–13955.

Tierney, B.T., Yang, Z., Luber, J.M., Beaudin, M., Wibowo, M.C., Baek, C., Mehlenbacher, E., Patel, C.J., and Kostic, A.D. (2019). The Landscape of Genetic Content in the Gut and Oral Human Microbiome. Cell Host Microbe *26*, 283–295.e8.

Vatanen, T., Kostic, A.D., d'Hennezel, E., Siljander, H., Franzosa, E.A., Yassour, M., Kolde, R., Vlamakis, H., Arthur, T.D., Hämäläinen, A.-M., et al.; DIABIMMUNE Study Group (2016). Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell *165*, 842–853.

von Mutius, E. (2007). Allergies, infections and the hygiene hypothesis–the epidemiological evidence. Immunobiology *212*, 433–439.

Wibowo, M.C., Yang, Z., Borry, M., Hübner, A., Huang, K.D., Tierney, B.T., Zimmerman, S., Barajas-Olmos, F., Contreras-Cubas, C., García-Ortiz, H., et al. (2021). Reconstruction of ancient microbial genomes from the human gut. Nature *594*, 234–239.