



## OPEN ACCESS

## EDITED BY

Luis Miguel Rodríguez,  
University of Innsbruck,  
Austria

## REVIEWED BY

Xuesong He,  
The Forsyth Institute,  
United States  
Hugo Cesar Ramirez-Saad,  
Autonomous Metropolitan University,  
Mexico

## \*CORRESPONDENCE

Jordan E. Bisanz  
✉ Jordan.Bisanz@psu.edu

## SPECIALTY SECTION

This article was submitted to  
Systems Microbiology,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 06 December 2022

ACCEPTED 24 January 2023

PUBLISHED 17 February 2023

## CITATION

Anderson BD and Bisanz JE (2023) Challenges  
and opportunities of strain diversity in gut  
microbiome research.  
*Front. Microbiol.* 14:1117122.  
doi: 10.3389/fmicb.2023.1117122

## COPYRIGHT

© 2023 Anderson and Bisanz. This is an open-  
access article distributed under the terms of  
the [Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Challenges and opportunities of strain diversity in gut microbiome research

Benjamin D. Anderson<sup>1</sup> and Jordan E. Bisanz<sup>1,2\*</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, United States, <sup>2</sup>The Penn State Microbiome Center, Huck Institutes of the Life Sciences, University Park, PA, United States

Just because two things are related does not mean they are the same. In analyzing microbiome data, we are often limited to species-level analyses, and even with the ability to resolve strains, we lack comprehensive databases and understanding of the importance of strain-level variation outside of a limited number of model organisms. The bacterial genome is highly plastic with gene gain and loss occurring at rates comparable or higher than *de novo* mutations. As such, the conserved portion of the genome is often a fraction of the pangenome which gives rise to significant phenotypic variation, particularly in traits which are important in host microbe interactions. In this review, we discuss the mechanisms that give rise to strain variation and methods that can be used to study it. We identify that while strain diversity can act as a major barrier in interpreting and generalizing microbiome data, it can also be a powerful tool for mechanistic research. We then highlight recent examples demonstrating the importance of strain variation in colonization, virulence, and xenobiotic metabolism. Moving past taxonomy and the species concept will be crucial for future mechanistic research to understand microbiome structure and function.

## KEYWORDS

strain diversity, gut microbiome, comparative genomics, metagenomics, species concept

## Introduction

Humans may be considered a holobiont: the sum of ourselves and our microbial inhabitants (Salvucci, 2016). The gastrointestinal tract is home to an extremely diverse set of taxa that are highly unique to each individual (Human Microbiome Project Consortium, 2012). Adding to this complexity is the staggering amount of diversity that occurs among microbes of the same species which is commonly referred to as intraspecific, intraspecies, or strain diversity (Truong et al., 2017; Van Rossum et al., 2020). While these strains sometimes form coherent subpopulations termed subspecies, there is a broad spectrum of diversity within species that have led some to question the relevance of the species concept as it pertains to prokaryotes (Fraser et al., 2009). Recent estimates indicate that the gut microbiome is home to ~4,644 bacterial species that are a reservoir for ~170 million genes (Almeida et al., 2021). Based on these estimates, a simplistic calculation would suggest the average bacterial species contains ~35,000 genes, an order of magnitude more than what is observed in most bacterial genomes (Land et al., 2015). The discrepancy between these two estimates underlies a fundamental principle of bacterial genetics and challenges the relevance of the species concept itself in gut microbes.

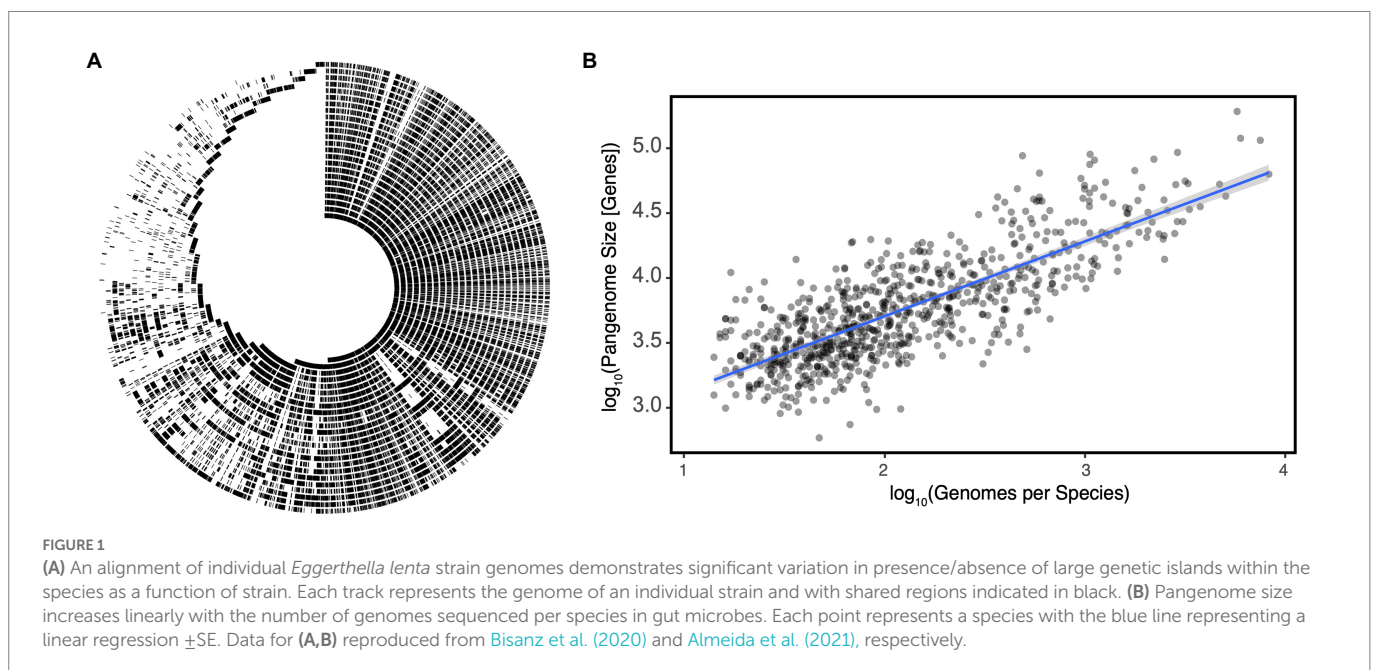
## Strain diversity and its origins

The species boundary in higher order sexually reproducing eukaryotes is relatively simple to qualify; however, the same cannot be said of prokaryotes. Early attempts to identify bacterial species were predominantly based on morphological observations and biochemical assays (Vandamme et al., 1996). Later, the molecular approach of DNA–DNA hybridization helped define species (De Ley et al., 1970) which was eventually supplanted by the more accessible sequencing of the 16S rRNA gene (Woese and Fox, 1977). These later methods provided objective quantitative thresholds for species: 70% hybridization and 97% identity, respectively. 16S rRNA gene sequencing has been a mainstay of taxonomists for the greater part of 30 years; however, rapid developments in whole genome sequencing enabled the comparison metric of whole genome average nucleotide identity (ANI) wherein  $\geq 94$ –95% has been generally accepted as the species boundary (Konstantinidis and Tiedje, 2005). Despite the widespread adoption of these quantitative boundaries, modern taxonomy is still ripe with inconsistencies such as *Shigella* spp. being a subspecies of *Escherichia coli* and the polyphyletic nature of the genus *Clostridium* (Lan and Reeves, 2002; Yutin and Galperin, 2013; The et al., 2016).

The ANI definition leads to a common misunderstanding: it is not to say that 95% of the genome is the same; rather, the portions of the genome which are shared have on average  $\geq 95\%$  nucleotide identity. This subtle difference has a not-so-subtle effect on how we interpret the meaning of a species as it masks the important observation that members of the same species share only a fraction of the genome with their closest relatives (Figure 1A). The set of genes found within a species is referred to as its pangenome, which can be split into the core genome: those conserved among all members of the species, and the accessory genome: those which are variably present (Medini et al., 2005). Pangenomes can vary significantly in size, but until the recent explosion of metagenome assembled genomes (MAGs), estimates were only known for a limited number of organisms with a strong bias toward model organisms and pathogens for which a ubiquity of strains had been sequenced. *Escherichia coli* for example has a core genome of  $\sim 2,000$

genes, but a given strain can have an additional 1,900–3,800 accessory genes resulting in a pangenome that may be as large as 75,000 genes (Denamur et al., 2021). Alternatively in *Bacillus anthracis*, a spore-forming pathogen, the pangenome is much smaller, with a much larger core genome of  $\sim 4,000$  genes, but a pangenome size of only 6,066 genes (Kim et al., 2017; McInerney et al., 2017). It is hypothesized that part of the difference in pangenome size can be attributed to differences in mutation and horizontal gene transfer (HGT) rates between species and how recently the species emerged (Segerman, 2012; Kim et al., 2017). It should however be noted that the *B. anthracis* example may illustrate another taxonomic inconsistency. *B. anthracis* may in fact be a subspecies of *B. cereus* based on ANI and a variety of other genetic approaches (Helgason et al., 2000). Indeed, re-analysis of estimates derived from MAGs indicates that a species' pangenome grows near linearly with the number of sequenced genomes within that species ( $Rho = 0.6204$ ,  $p = 2.2E-16$ , Figure 1B).

While *de novo* point mutations are common in bacterial genomes, gene gain and loss events have been estimated to occur at rates up to 4.4 fold higher (Guttman and Dykhuizen, 1994; Vos et al., 2015). These observations demonstrate that the bacterial genome is highly plastic and prone to rapid remodeling over time scales incomprehensible when thinking about genome evolution in eukaryotes. Horizontal gene transfer (HGT) is a major driver of bacterial genome evolution wherein foreign genetic material is either incorporated into the genome or maintained on mobile elements such as plasmids (Soucy et al., 2015). Horizontal gene transfer occurs through three major routes: transformation, transduction, and conjugation (Soucy et al., 2015). Transformation involves the uptake of free DNA from the environment without direct interaction between bacteria (Blokesch, 2016). Transduction is an almost accidental process by which a bacteriophage includes DNA from a donor bacterium during virion packaging, this DNA is then transferred to a recipient bacterium (Schneider, 2021). Alternatively, conjugation is a direct transfer of DNA between microbes *via pili*, which connect cells and is most commonly associated with the transfer of plasmids (Thomas and Nielsen, 2005). These traits have been well studied in the context of transferring antibiotic resistance (Lopatkin



et al., 2016; Lu et al., 2017); however, they likely drive much of the genotypic variation among closely related organisms.

Just as genes can be gained through HGT, they can be lost through reductive evolution (Batut et al., 2014) or conversion to pseudogenes (Bolotin and Hershberg, 2015). Maintaining extra genetic material is energetically costly, and if they provide no benefit, they can be quickly lost. Most often gene loss is more pronounced in pathogenic strains of bacteria which have become dependent on association with their host (Moran, 2002). Essentiality describes genes that are required for an organism to live and proliferate. Screening methods have determined that only around 10% of genes in the *E. coli* genome are essential in rich media conditions (Albalat and Cañestro, 2016). This can also lead to the domesticated lab strain phenomenon wherein organisms can acquire new mutations and lose important genetic islands after repeated passaging on rich media (Sybesma et al., 2013; Denamur et al., 2021; Monteford et al., 2021). Additionally, as strains are separated by time and space, some amount of genetic drift can occur further differentiating strains (Bolotin and Hershberg, 2016). Recent experimental models investigated *E. coli* evolution in the context of host colonization following antibiotic-mediated engraftment (Frazão et al., 2022). Even in time scales of weeks with an estimated >6,000 bacterial generations, the authors uncovered both diversifying selection supporting coexistence of strains and directional selective sweeps which were determined to predominantly arise from new mutations and prophage acquisition, respectively. As bacterial strains are separated by spatial, temporal, and environmental elements, mutations, and gene loss events can reshape their genomes.

## Measuring and manipulating strain diversity

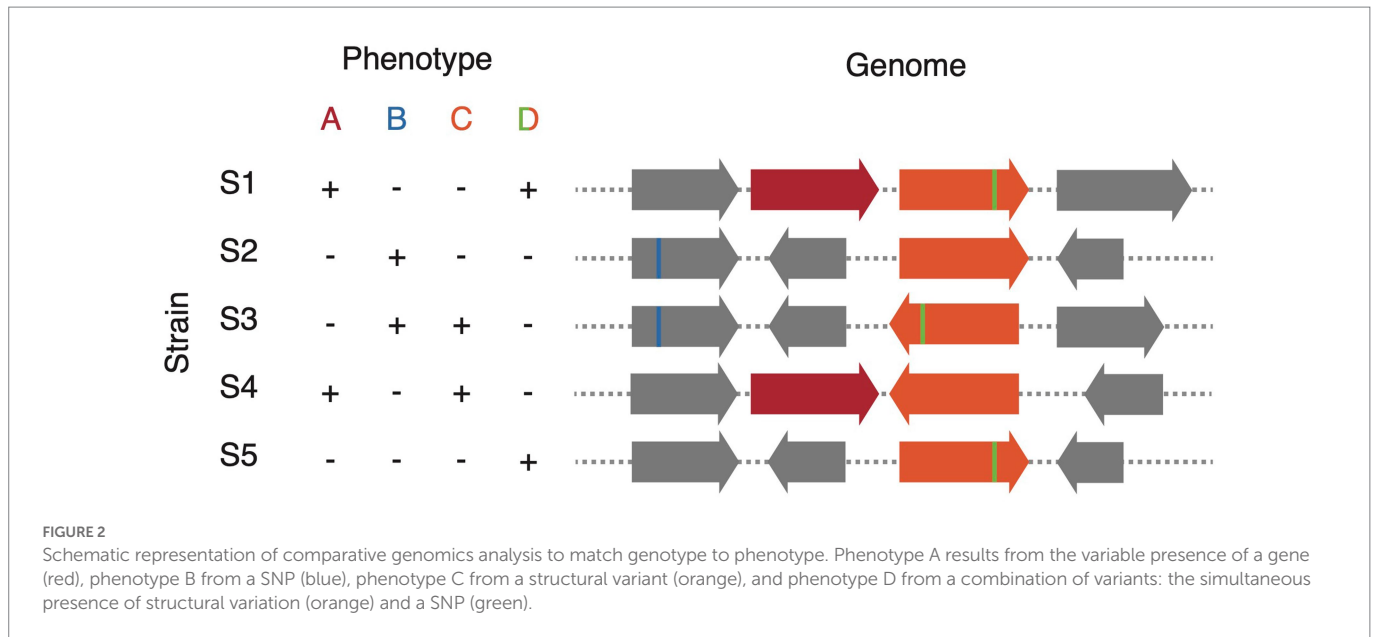
While many modern sequencing approaches are capable of resolving strains in theory, in practice, it is easier said than done. 16S rRNA gene sequencing typically lacks the ability to resolve strains and often species as well (Johnson et al., 2019). While full-length 16S rRNA sequencing may improve species resolution, it is still limited in its ability to meaningfully resolve strains due to the slow rate of evolution in the 16S rRNA gene versus the rest of the genome (Callahan et al., 2019). On the other hand, metagenomic sequencing offers the possibility to resolve strains with certain limitations. Owing to a desire for dimensional reduction and effective communication, metagenomic data are often summarized to higher taxonomic levels such as the species, genus, family, or even phylum. Instead, accessible methods are needed to meaningfully quantify strains, and perhaps more importantly, understand the biological significance of that strain variation.

Culture represents a traditional way through which strain diversity can be determined and quantified. Conventional culture methods are capable of isolating hundreds or thousands of strains with sufficient effort (Poyet et al., 2019; Hitch et al., 2021; Afrizal et al., 2022). Resulting colonies can then be dereplicated on the basis of MALDI-TOF profiles and/or fingerprinting approaches such as RAPD or ERIC PCR (Versalovic et al., 1991; Bazzicalupo and Fani, 1996). The resulting genomes of these strains can then be sequenced which will lead to assemblies almost invariably higher quality than those derived from metagenomic methods. This approach also allows for direct phenotyping and laboratory experimentation on strains. The advantages of this method are offset by the significant resources and infrastructure required to conduct these approaches at large scales and culture bias

against many of the most prevalent members of the gut microbiome: the inability to effectively culture as many as 80% of the microbes found in the gastrointestinal tract (Lagier et al., 2012).

Where strain isolates are available in culture, strain variation can be a powerful tool for comparative genomic approaches to discover genes and enzymes of interest. Traditional screening methods, such as a transposon mutagenesis screen, require a genetically tractable host to randomly inactivate genes followed by screening of thousands of clones to look for phenotypic changes (Barquist et al., 2013). Alternatively, strain variation gives rise to what could be thought of as a “natural combinatorial knockout system.” In effect, if the trait of interest is variable among members of the species, which is often the case for traits involving host–microbe interactions, screening as few as 10 strains of the same species may be sufficient to map the genetic determinant (Bisanz et al., 2020; Alexander et al., 2022). A variety of comparative genomics approaches may be used for these analyses; however, it should be noted that the genetic determinants may be driven by any combination of: gene presence/absence, single-nucleotide polymorphisms (SNPs), and structural rearrangements (Figure 2). Gene presence/absence can be inferred relatively easily through methodologies employing reciprocal blast (Li et al., 2003; Lechner et al., 2011; Page et al., 2015); however, most methodologies to call SNPs are tailored to call SNPs in the core genome rather than those in the accessory. Alternatively, we have shown that the use of tiled *k*-mers is capable of accurately detecting explanatory SNPs and other structural changes; however, their use comes at a significantly greater computational overhead (Maini Rekdal et al., 2019; Bisanz et al., 2020). Because of the potential for combinations of predictors driving phenotype: for example, gene A or gene B, or the presence of a gene A plus the absence of a negative regulator gene C, machine learning approaches provide a powerful tool. Random Forest classifier/regression models are particularly adept at this task as they have straight forward metrics for each feature’s importance/predictive value and they generally consider combinations of explanatory variables in their decision trees (Chen and Ishwaran, 2012). Indeed, we have used variants of this approach across multiple manuscripts (Koppel et al., 2018; Maini Rekdal et al., 2019, 2020; Bess et al., 2020; Bisanz et al., 2020; Pröbstel et al., 2020; Alexander et al., 2022; Kyaw et al., 2022; Noecker et al., 2022; Paik et al., 2022). These comparative genomics approaches are attractive because phenotypes can be screened at a relatively low-throughput scale and they bypass the need for genetic tools. This immediately opens up new possibilities for the vast majority of gut microbes in which genetic manipulation is not yet possible.

Metagenomics is the current state of the art approach for cataloging strain variation at massive scales and for mapping strain dynamics within samples (Qin et al., 2010; Schloissnig et al., 2013). MAGs have allowed for unprecedented analysis of genetic diversity among gut microbiomes through binning genome fragments of an individual strain from the mixed population. This is typically accomplished through examining co-occurrence, co-abundance, and sequence composition of the fragmented assembly (Wang et al., 2015; Zhernakova et al., 2016). The use of MAGs does however have some important limitations as it often struggles to bin low coverage/low abundance organisms and there is a finite probability that when two strains are present within a sample at similar abundances, they will be combined into a single MAG (Chen et al., 2020). Current approaches can detect heterogeneity within genomes on the basis of variants and duplication of single copy genes; however, careful analysis and polishing is required (Parks et al., 2015; Mineeva et al.,



2020), while new approaches to separate MAGs into separate strains are being developed (Quince et al., 2021). One challenge for the generation of high-quality MAGs from metagenomic data is limitations in sequencing technologies. Short-read sequencers are typically preferred due to lower error rates than long-read sequencers, but shorter reads are typically unable to resolve repetitive regions of genomes. This has led to the proliferation of hybrid assembly approaches combining both long-read and short-read approaches to improve metagenomic-binning and genome assembly (Bertrand et al., 2019); however, recent advances in the accuracy of long-read technologies are beginning to enable the generation of high-quality MAGs based on long reads alone (Sereika et al., 2022).

As with the comparative genomics approaches previously identified, SNPs provide a powerful tool for strain genotyping in metagenomic data (Truong et al., 2017; Shi et al., 2022). One method, StrainPhlAn, reconstructs SNPs within species-specific marker genes and uses these to infer strain-level phylogenies (Truong et al., 2017). A more recent approach, GenoTyper for Prokaryotes (GT-Pro) uses a less computationally intensive alignment-free method based on unique *k*-mers that are compared to a catalog of SNPs to more efficiently genotype strains (Shi et al., 2022).

Rapid advances in single-cell genomics technologies have brought exciting opportunities for microbiome research through physical and chemical methods to resolve single strain genomes. Flow-Assisted Cell Sorting (FACS) has been used to isolate single cells for use in single-cell sequencing (Rinke et al., 2014). More commonly, droplet-based microfluidics have been applied (Lan et al., 2017). The droplets trap individual bacterial cells from which sequencing libraries are then prepared at the single-cell scale (Hosokawa et al., 2017). Unfortunately, these methods often feature lower per-cell sequencing depth and require significant amplification to generate sufficient material for sequencing which limits the ability of these methods to resolve full genome assemblies. To aid in overcoming these limitations, single-cell sequencing and conventional metagenomic approaches may be combined to improve the quality and strain-level resolution (Arikawa et al., 2021).

Alternatively chemical methods may be employed such as high-throughput chromosome conformation capture (Hi-C; Burton et al., 2014; Du and Sun, 2022). Originally developed for analyzing chromatin structure, DNA in close proximity, i.e., belonging to the same bacterial chromosome, is ligated together before sequencing, revealing which fragments are most likely to be from the same microbial genome (Pal et al., 2019).

## Strain diversity as determinant of microbiome assembly and function

The biological impact of strain variation is widely recognized in the field of bacterial pathogenesis wherein virulence traits are known to vary significantly within species. *Escherichia coli* represents perhaps the best-known example: members of this species can be the cause of severe diarrhea and dysentery (Kopecko et al., 1985), the most common cause of urinary tract infection (Johnson, 1991), or beneficial microbes administered intentionally as a probiotic to prevent diarrheal illness (Henker et al., 2007). Indeed, the virulence properties of many well-known bacterial pathogens vary significantly as a function of gene content and lineage including *Clostridioides difficile* (Hunt and Ballard, 2013), the *Bacillus cereus* group (Ceuppens et al., 2013), *Cutibacterium [Propionibacterium] acnes* (Tomida et al., 2013), and *Bacteroides fragilis* (Pierce and Bernstein, 2016). *Bacteroides fragilis* exists as one of the most common commensals of the human gut microbiome whose colonization is established early in life and may be vertically acquired; however, conventional wisdom would suggest that this species is an opportunistic pathogen (Carrow et al., 2020). *Bacteroides fragilis* strains exhibit the variable presence of a metalloprotease toxin which disrupts barrier function and drives intestinal inflammation (Moncrief et al., 1995); however, opposing this function, some strains produce an extracellular polysaccharide (polysaccharide A) that helps promote barrier function through modulation of regulatory T cells (Round and Mazmanian, 2010).

Strain variation also drives community composition through competitive exclusion *via* a variety of mechanisms. Ecological theory dictates that strains or species that occupy the same niche

will compete resulting in exclusion from the community (Hardin, 1960). This has been observed in *Eggerthella lenta*, *Bacteroides* spp., and most recently *C. acnes* (Hecht et al., 2016; Bisanz et al., 2020; Conwill et al., 2022). While the skin can be colonized by multiple strains of *C. acnes*, individual pores contain a clonal strain population indicating that there may be a spatial component to strain variation across host-associated microbiomes (Conwill et al., 2022). These observations have implications for the gut microbiome in terms of diversity and the use of fecal microbiota transplants (FMTs) to treat diseases. One important aspect of competitive exclusion is the timing of a strain being introduced into the community as established strains and species are not as likely to be excluded from a community as newly introduced strains (Grainger et al., 2019; Munoz et al., 2022). This is particularly relevant as the gut microbiome is usually inherited from an individual's mother (Mueller et al., 2015; Asnicar et al., 2017; Duranti et al., 2017). This has been experimentally demonstrated with *Akkermansia muciniphila* strains, wherein mice colonized with one strain were resistant to colonization by a second (Munoz et al., 2022). However, strict strain homogeneity through competitive exclusion is not always the case, and some strains can co-exist as experimentally demonstrated for *Phocaeicola [Bacteroides] vulgatus* and *E. lenta* (Bisanz et al., 2020; Munoz et al., 2022). Acquisition of genes allowing for exploitation of a new nutrient source may partially alleviate this competition as has been engineered into *Bacteroides* spp. (Shepherd et al., 2018).

The ability of microbes to illicit immune responses can also be highly strain specific. Various strains of *Ruminococcus gnavus*, *A. muciniphila*, and *B. fragilis* can have wide reaching effects on modulation of the immune system (Troy and Kasper, 2010; Cassard et al., 2016; Henke et al., 2021; Liu et al., 2021). *Ruminococcus gnavus* shows strain-level differences in immune response with distinct immune responses found depending on the presence of a biosynthetic pathway for production of a capsular polysaccharide (Henke et al., 2021). Similarly, *A. muciniphila* displays variable anti-inflammatory effects through unknown mechanisms (Zhai et al., 2019; Liu et al., 2021). A range of immune responses are also detected with *Lactobacillus paracasei* strains which can have a range of inhibiting activation of mouse mast cells and human basophils (Cassard et al., 2016).

Other than virulence, perhaps, the best-known examples of strain diversity are in drug-microbe interactions. It is becoming increasingly acknowledged that there are extensive interactions between gut microbes and orally consumed drugs/xenobiotics (Maier et al., 2018; Zimmermann et al., 2019; Klünemann et al., 2021). These drug metabolism traits are particularly interesting as we are quickly gaining answers as to how microbes can metabolize drugs, but not why. In most cases, we have not determined a fitness advantage from drug metabolism which creates the perfect opportunity for these pathways to largely exist in the accessory genome. Indeed, we were inspired by early work examining strain variation in *Lactobacillus* spp. (Douillard et al., 2013; Smokvina et al., 2013) to map variation in *E. lenta*, a highly prevalent, but relatively understudied member of the gut microbiome (Koppel et al., 2018). *Eggerthella lenta* was known to metabolize the cardiac drug digoxin as early as the 1980s (Dobkin et al., 1982); however, the mechanisms were unknown until 2013 when RNA-seq revealed an operon whose expression was induced by the presence of the drug (Haiser et al., 2013). Using comparative genomics approaches, we identified that this operon was part of a gene cluster which was variably present across strains of the species

which explained why the presence of *E. lenta* was insufficient to predict digoxin metabolic activity (Koppel et al., 2018). Taking this a step further, we later determined that a single coding variant of the active enzyme CGR2 dictated its activity resulting in three phenotypes in effect: no metabolism, low metabolism, and high metabolism. Similarly, *E. lenta* operates in a meta-organismal pathway leading to the production of phytoestrogens which we mapped to the presence of a single enzyme variably present in the genome (Bess et al., 2020). *Eggerthella lenta* also cooperates with *Enterococcus faecalis* in the premature breakdown of the Parkinson's drug levodopa which is determined by a SNP affecting enzyme activity (Maini Rekdal et al., 2019). Understanding the mechanisms of strain-level variation in drug interactions opens up new possibilities for precision medicine and pharmacological therapy: i.e., rather than try to sequence microbiome composition or quantify specific microbes, we could instead design targeted assays to predict drug metabolism based on detection/quantification of specific genes or variants.

## Concluding remarks

Advancing microbiome science from a descriptive to a mechanistic science requires a detailed understanding of microbial function, but these functions are often not conserved at the species level. If we stereotype all *E. coli* as pathogens, or *A. muciniphila* as beneficial, we are likely to miss the trees for the forest. By viewing our data through a taxonomic lens, we may lose the ability to find the important determinants of microbiome structure and function. We need to be aware of strain variation in our data and carefully catalog it. Recent advances in sequencing technology are making it quickly possible to follow strains in metagenomic samples, but we then need databases incorporating functional annotations and phenotypic information to draw mechanistic insight from this data. By pairing these approaches with wet lab experimentation, we can turn strain variation from one of the major challenges in microbiome research to one of its greatest tools.

## Author contributions

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

## Conflict of interest

The authors declare 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.

## References

- Afrizal, A., Jennings, S. A. V., Hitch, T. C. A., Riedel, T., Basic, M., Panyot, A., et al. (2022). Enhanced cultured diversity of the mouse gut microbiota enables custom-made synthetic communities. *Cell Host Microbe* 30, 1630–1645.e25. doi: 10.1016/j.chom.2022.09.011
- Albalat, R., and Cañestro, C. (2016). Evolution by gene loss. *Nat. Rev. Genet.* 17, 379–391. doi: 10.1038/nrg.2016.39
- Alexander, M., Ang, Q. Y., Nayak, R. R., Bustion, A. E., Sandy, M., Zhang, B., et al. (2022). Human gut bacterial metabolism drives Th17 activation and colitis. *Cell Host Microbe* 30, 17–30.e9. doi: 10.1016/j.chom.2021.11.001
- Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z. J., et al. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat. Biotechnol.* 39, 105–114. doi: 10.1038/s41587-020-0603-3
- Arikawa, K., Ide, K., Kogawa, M., Saeki, T., Yoda, T., Endoh, T., et al. (2021). Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and metagenomics. *Microbiome* 9:202. doi: 10.1186/s40168-021-01152-4
- Asnicar, F., Manara, S., Zolfo, M., Truong, D. T., Scholz, M., Armanini, F., et al. (2017). Studying vertical microbiome transmission from mothers to infants by strain-level metagenomic profiling. *mSystems* 2, e00164–16. doi: 10.1128/mSystems.00164-16
- Barquist, L., Boinett, C. J., and Cain, A. K. (2013). Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.* 10, 1161–1169. doi: 10.4161/rna.24765
- Batut, B., Knibbe, C., Marais, G., and Daubin, V. (2014). Reductive genome evolution at both ends of the bacterial population size spectrum. *Nat. Rev. Microbiol.* 12, 841–850. doi: 10.1038/nrmicro3331
- Bazzicalupo, M., and Fani, R. (1996). The use of RAPD for generating specific DNA probes for microorganisms. *Methods Mol. Biol.* 50, 155–175.
- Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A. H. Q., Kumar, M. S., Li, C., et al. (2019). Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* 37, 937–944. doi: 10.1038/s41587-019-0191-2
- Bess, E. N., Bisanz, J. E., Yarza, F., Bustion, A., Rich, B. E., Li, X., et al. (2020). Genetic basis for the cooperative bioactivation of plant lignans by *Eggerthella lenta* and other human gut bacteria. *Nat. Microbiol.* 5, 56–66. doi: 10.1038/s41564-019-0596-1
- Bisanz, J. E., Soto-Perez, P., Noecker, C., Aksenov, A. A., Lam, K. N., Kenney, G. E., et al. (2020). A genomic toolkit for the mechanistic dissection of intractable human gut bacteria. *Cell Host Microbe* 27, 1001–1013.e9. doi: 10.1016/j.chom.2020.04.006
- Blokesch, M. (2016). Natural competence for transformation. *Curr. Biol.* 26:3255. doi: 10.1016/j.cub.2016.11.023
- Bolotin, E., and Hershberg, R. (2015). Gene loss dominates as a source of genetic variation within clonal pathogenic bacterial species. *Genome Biol. Evol.* 7, 2173–2187. doi: 10.1093/gbe/evv135
- Bolotin, E., and Hershberg, R. (2016). Bacterial intra-species gene loss occurs in a largely clocklike manner mostly within a pool of less conserved and constrained genes. *Sci. Rep.* 6. doi: 10.1038/srep35168
- Burton, J. N., Liachko, I., Dunham, M. J., and Shendure, J. (2014). Species-level deconvolution of metagenome assemblies with Hi-C–based contact probability maps. *G3* 4, 1339–1346. doi: 10.1534/g3.114.011825
- Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., et al. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* 47:e103. doi: 10.1093/nar/gkz569
- Carrow, H. C., Batachari, L. E., and Chu, H. (2020). Strain diversity in the microbiome: lessons from *Bacteroides fragilis*. *PLoS Pathog.* 16:e1009056. doi: 10.1371/journal.ppat.1009056
- Cassard, L., Lalanne, A. I., Garault, P., Cotillard, A., Chervaux, C., Wels, M., et al. (2016). Individual strains of *Lactobacillus paracasei* differentially inhibit human basophil and mouse mast cell activation. *Immun. Inflamm. Dis.* 4, 289–299. doi: 10.1002/iid3.113
- Ceuppens, S., Boon, N., and Uyttendaele, M. (2013). Diversity of *Bacillus cereus* group strains is reflected in their broad range of pathogenicity and diverse ecological lifestyles. *FEMS Microbiol. Ecol.* 84, 433–450. doi: 10.1111/1574-6941.12110
- Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A. M., and Banfield, J. F. (2020). Accurate and complete genomes from metagenomes. *Genome Res.* 30, 315–333. doi: 10.1101/gr.258640.119
- Chen, X., and Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics* 99, 323–329. doi: 10.1016/j.ygeno.2012.04.003
- Conwill, A., Kuan, A. C., Damerla, R., Poret, A. J., Baker, J. S., Tripp, A. D., et al. (2022). Anatomy promotes neutral coexistence of strains in the human skin microbiome. *Cell Host Microbe* 30, 171–182.e7. doi: 10.1016/j.chom.2021.12.007
- De Ley, J., Cattoir, H., and Reynaerts, A. (1970). The quantitative measurement of DNA hybridization from renaturation rates. *Eur. J. Biochem.* 12, 133–142. doi: 10.1111/j.1432-1033.1970.tb00830.x
- Denamur, E., Clermont, O., Bonacorsi, S., and Gordon, D. (2021). The population genetics of pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* 19, 37–54. doi: 10.1038/s41579-020-0416-x
- Dobkin, J. F., Saha, J. R., Butler, V. P., Neu, H. C., and Lindenbaum, J. (1983). Digoxin-inactivating bacteria: identification in human gut Flora. *Science* 220, 325–327. doi: 10.1126/science.6836275
- Douillard, F. P., Ribbera, A., Kant, R., Pietilä, T. E., Järvinen, H. M., Messing, M., et al. (2013). Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genet.* 9:e1003683. doi: 10.1371/journal.pgen.1003683
- Du, Y., and Sun, F. (2022). HiCBin: binning metagenomic contigs and recovering metagenome-assembled genomes using hi-C contact maps. *Genome Biol.* 23:63. doi: 10.1186/s13059-022-02626-w
- Duranti, S., Lugli, G. A., Mancabelli, L., Armanini, F., Turroni, F., James, K., et al. (2017). Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* 5:66. doi: 10.1186/s40168-017-0282-6
- Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., and Hanage, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science* 323, 741–746. doi: 10.1126/science.1159388
- Frazão, N., Konrad, A., Amicone, M., Seixas, E., Güleresi, D., Lässig, M., et al. (2022). Two modes of evolution shape bacterial strain diversity in the mammalian gut for thousands of generations. *Nat. Commun.* 13:5604. doi: 10.1038/s41467-022-33412-8
- Grainger, T. N., Letten, A. D., Gilbert, B., and Fukami, T. (2019). Applying modern coexistence theory to priority effects. *Proc. Natl. Acad. Sci. U. S. A.* 116, 6205–6210. doi: 10.1073/pnas.1803122116
- Guttman, D. S., and Dykhuizen, D. E. (1994). Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* 266, 1380–1383. doi: 10.1126/science.7973728
- Haiser, H. J., Gootenberg, D. B., Chatman, K., Sirasani, G., Balskus, E. P., and Turnbaugh, P. J. (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* 341, 295–298. doi: 10.1126/science.1235872
- Hardin, G. (1960). The competitive exclusion principle. *Science* 131, 1292–1297. doi: 10.1126/science.131.3409.1292
- Hecht, A. L., Casterline, B. W., Earley, Z. M., Goo, Y. A., Goodlett, D. R., and Bubeck Wardenburg, J. (2016). Strain competition restricts colonization of an enteric pathogen and prevents colitis. *EMBO Rep.* 17, 1281–1291. doi: 10.15252/embr.201642282
- Helgason, E., Okstad, O. A., Caugant, D. A., Johansen, H. A., Fouet, A., Mock, M., et al. (2000). *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. *Appl. Environ. Microbiol.* 66, 2627–2630. doi: 10.1128/AEM.66.6.2627-2630.2000
- Henke, M. T., Brown, E. M., Cassilly, C. D., Vlamakis, H., Xavier, R. J., and Clardy, J. (2021). Capsular polysaccharide correlates with immune response to the human gut microbe *Ruminococcus gnavus*. *Proc. Natl. Acad. Sci. U. S. A.* 118, e35168. doi: 10.1073/pnas.2007595118
- Henker, J., Laass, M., Blokhin, B. M., Bolbot, Y. K., Maydannik, V. G., Elze, M., et al. (2007). The probiotic *Escherichia coli* strain Nissle 1917 (EcN) stops acute diarrhoea in infants and toddlers. *Eur. J. Pediatr.* 166, 311–318. doi: 10.1007/s00431-007-0419-x
- Hitch, T. C. A., Afrizal, A., Riedel, T., Kioukis, A., Haller, D., Lagkouvardos, I., et al. (2021). Recent advances in culture-based gut microbiome research. *Int. J. Med. Microbiol.* 311:151485. doi: 10.1016/j.ijmm.2021.151485
- Hosokawa, M., Nishikawa, Y., Kogawa, M., and Takeyama, H. (2017). Massively parallel whole genome amplification for single-cell sequencing using droplet microfluidics. *Sci. Rep.* 7:5199. doi: 10.1038/s41598-017-05436-4
- Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi: 10.1038/nature11234
- Hunt, J. J., and Ballard, J. D. (2013). Variations in virulence and molecular biology among emerging strains of *Clostridium difficile*. *Microbiol. Mol. Biol. Rev.* 77, 567–581. doi: 10.1128/MMBR.00017-13
- Johnson, J. R. (1991). Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* 4, 80–128. doi: 10.1128/CMR.4.1.80
- Johnson, J. S., Spakowicz, D. J., Hong, B.-Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10:5029. doi: 10.1038/s41467-019-13036-1
- Kim, Y., Koh, I., Young Lim, M., Chung, W.-H., and Rho, M. (2017). Pan-genome analysis of *Bacillus* for microbiome profiling. *Sci. Rep.* 7:10984. doi: 10.1038/s41598-017-11385-9
- Klünemann, M., Andrejev, S., Blasche, S., Mateus, A., Phapale, P., Devendran, S., et al. (2021). Bioaccumulation of therapeutic drugs by human gut bacteria. *Nature* 597, 533–538. doi: 10.1038/s41586-021-03891-8
- Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102
- Kopecko, D. J., Baron, L. S., and Buysse, J. (1985). Genetic determinants of virulence in *Shigella* and dysenteric strains of *Escherichia coli*: their involvement in the pathogenesis of dysentery. *Curr. Top. Microbiol. Immunol.* 118, 71–95.
- Koppel, N., Bisanz, J. E., Pandelia, M.-E., Turnbaugh, P. J., and Balskus, E. P. (2018). Discovery and characterization of a prevalent human gut bacterial enzyme sufficient for the inactivation of a family of plant toxins. *elife* 7: e33953. doi: 10.7554/eLife.33953

- Kyaw, T. S., Sandy, M., Trepka, K., Goh, J. J. N., Yu, K., Dimassa, V., et al. (2022). Human gut Actinobacteria boost drug absorption by secreting P-glycoprotein ATPase inhibitors. *bioRxiv* [Preprint]. doi: 10.1101/2022.10.13.512142
- Lagier, J.-C., Armougom, F., Million, M., Hugon, P., Pagnier, I., Robert, C., et al. (2012). Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin. Microbiol. Infect.* 18, 1185–1193. doi: 10.1111/1469-0691.12023
- Lan, F., Demaree, B., Ahmed, N., and Abate, A. R. (2017). Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nat. Biotechnol.* 35, 640–646. doi: 10.1038/nbt.3880
- Lan, R., and Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect.* 4, 1125–1132. doi: 10.1016/S1286-4579(02)01637-4
- Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M. R., Ahn, T.-H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15, 141–161. doi: 10.1007/s10142-015-0433-4
- Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P. F., and Prohaska, S. J. (2011). Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12:124. doi: 10.1186/1471-2105-12-124
- Li, L., Stoekert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503
- Liu, Q., Lu, W., Tian, F., Zhao, J., Zhang, H., Hong, K., et al. (2021). *Akkermansia mucimiphila* exerts strain-specific effects on DSS-induced ulcerative colitis in mice. *Front. Cell. Infect. Microbiol.* 11:698914. doi: 10.3389/fcimb.2021.698914
- Lopatkin, A. J., Huang, S., Smith, R. P., Srimani, J. K., Sysoeva, T. A., Bewick, S., et al. (2016). Antibiotics as a selective driver for conjugation dynamics. *Nat. Microbiol.* 1:16044. doi: 10.1038/nmicrobiol.2016.44
- Lu, Y., Zeng, J., Wang, L., Lan, K. E. S., Wang, L., Xiao, Q., et al. (2017). Antibiotics promote *Escherichia coli*-*Pseudomonas aeruginosa* conjugation through inhibiting quorum sensing. *Antimicrob. Agents Chemother.* 61, e01284–17. doi: 10.1128/AAC.01284-17
- Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E. E., et al. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature* 555, 623–628. doi: 10.1038/nature25979
- Maini Rekdal, V., Bess, E. N., Bisanz, J. E., Turnbaugh, P. J., and Balskus, E. P. (2019). Discovery and inhibition of an interspecies gut bacterial pathway for levodopa metabolism. *Science* 364: eaau6323. doi: 10.1126/science.aau6323
- Maini Rekdal, V., Nol Bernadino, P., Luescher, M. U., Kiamehr, S., Le, C., Bisanz, J. E., et al. (2020). A widely distributed metalloenzyme class enables gut microbial metabolism of host- and diet-derived catechols. *elife* 9:e50845. doi: 10.7554/eLife.50845
- McInerney, J. O., McNally, A., and O'Connell, M. J. (2017). Why prokaryotes have pangenes. *Nat. Microbiol.* 2:17040. doi: 10.1038/nmicrobiol.2017.40
- Medini, D., Donati, C., Tettelin, H., Massignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006
- Mineeva, O., Rojas-Carulla, M., Ley, R. E., Schölkopf, B., and Youngblut, N. D. (2020). DeepMAsED: evaluating the quality of metagenomic assemblies. *Bioinformatics* 36, 3011–3017. doi: 10.1093/bioinformatics/btaa124
- Moncrief, J. S., Obiso, R. Jr., Barroso, L. A., Kling, J. J., Wright, R. L., Van Tassel, R. L., et al. (1995). The enterotoxin of *Bacteroides fragilis* is a metalloprotease. *Infect. Immun.* 63, 175–181. doi: 10.1128/iai.63.1.175-181.1995
- Monteford, J., Bilverstone, T. W., Ingle, P., Philip, S., Kuehne, S. A., and Minton, N. P. (2021). What's a SNP between friends: The lineage of *Clostridioides difficile* R20291 can effect research outcomes. *Anaerobe* 71:102422. doi: 10.1016/j.anaerobe.2021.102422
- Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cells* 108, 583–586. doi: 10.1016/S0092-8674(02)00665-7
- Mueller, N. T., Bakacs, E., Combellick, J., Grigoryan, Z., and Dominguez-Bello, M. G. (2015). The infant microbiome development: mom matters. *Trends Mol. Med.* 21, 109–117. doi: 10.1016/j.molmed.2014.12.002
- Munoz, R. R. S., Segura Munoz, R. R., Mantz, S., Martinez, I., Schmaltz, R. J., Walter, J., et al. (2022). Experimental evaluation of ecological principles to understand and modulate the outcome of bacterial strain competition in gut microbiomes. *ISME J.* 16, 1594–1604. doi: 10.1038/s41396-022-01208-9
- Noecker, C., Sanchez, J., Bisanz, J. E., Escalante, V., Alexander, M., Trepka, K., et al. (2022). Systems biology illuminates alternative metabolic niches in the human gut microbiome. *bioRxiv* [Preprint]. doi: 10.1101/2022.09.19.508335
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421
- Paik, D., Yao, L., Zhang, Y., Bae, S., D'Agostino, G. D., Zhang, M., et al. (2022). Human gut bacteria produce TH17-modulating bile acid metabolites. *Nature* 603, 907–912. doi: 10.1038/s41586-022-04480-z
- Pal, K., Forcato, M., and Ferrari, F. (2019). Hi-C analysis: from data generation to integration. *Biophys. Rev.* 11, 67–78. doi: 10.1007/s12551-018-0489-1
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pierce, J. V., and Bernstein, H. D. (2016). Genomic diversity of Enterotoxigenic strains of *Bacteroides fragilis*. *PLoS One* 11:e0158171. doi: 10.1371/journal.pone.0158171
- Poyet, M., Groussin, M., Gibbons, S. M., Avila-Pacheco, J., Jiang, X., Kearney, S. M., et al. (2019). A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research. *Nat. Med.* 25, 1442–1452. doi: 10.1038/s41591-019-0559-3
- Pröbstel, A.-K., Zhou, X., Baumann, R., Wischniewski, S., Kutza, M., Rojas, O. L., et al. (2020). Gut microbiota-specific IgA+ B cells traffic to the CNS in active multiple sclerosis. *Sci. Immunol.* 5: eabc7191. doi: 10.1126/sciimmunol.abc7191
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821
- Quince, C., Nurk, S., Raguideau, S., James, R., Soyer, O. S., Kimberly Summers, J., et al. (2021). STRONG: metagenomics strain resolution on assembly graphs. *Genome Biol.* 22:214. doi: 10.1186/s13059-021-02419-7
- Rinke, C., Lee, J., Nath, N., Goudeau, D., Thompson, B., Poulton, N., et al. (2014). Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. *Nat. Protoc.* 9, 1038–1048. doi: 10.1038/nprot.2014.067
- Round, J. L., and Mazmanian, S. K. (2010). Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12204–12209. doi: 10.1073/pnas.0909122107
- Salvucci, E. (2016). Microbiome, holobiont and the net of life. *Crit. Rev. Microbiol.* 42, 485–494. doi: 10.3109/1040841X.2014.962478
- Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., et al. (2013). Genomic variation landscape of the human gut microbiome. *Nature* 493, 45–50. doi: 10.1038/nature11711
- Schneider, C. L. (2021). “Bacteriophage-mediated horizontal gene transfer: transduction” in *Bacteriophages: Biology, Technology, Therapy*. eds. D. R. Harper, S. T. Abedon, B. H. Burrows and M. L. McConville (Cham: Springer International Publishing), 151–192.
- Segerman, B. (2012). The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* 2:116. doi: 10.3389/fcimb.2012.00116
- Sereika, M., Kirkegaard, R. H., Karst, S. M., Michaelsen, T. Y., Sørensen, E. A., Wollenberg, R. D., et al. (2022). Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat. Methods* 19, 823–826. doi: 10.1038/s41592-022-01539-7
- Shepherd, E. S., DeLoache, W. C., Pruss, K. M., Whitaker, W. R., and Sonnenburg, J. L. (2018). An exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* 557, 434–438. doi: 10.1038/s41586-018-0092-4
- Shi, Z. J., Dimitrov, B., Zhao, C., Nayfach, S., and Pollard, K. S. (2022). Fast and accurate metagenotyping of the human gut microbiome with GT-pro. *Nat. Biotechnol.* 40, 507–516. doi: 10.1038/s41587-021-01102-3
- Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., et al. (2013). *Lactobacillus paracasei* comparative genomics: towards species pan-genome definition and exploitation of diversity. *PLoS One* 8:e68731. doi: 10.1371/journal.pone.0068731
- Soucy, S. M., Huang, J., and Gogarten, J. P. (2015). Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* 16, 472–482. doi: 10.1038/nrg3962
- Sybesma, W., Molenaar, D., van Ijcken, W., Venema, K., and Kort, R. (2013). Genome instability in *Lactobacillus rhamnosus* GG. *Appl. Environ. Microbiol.* 79, 2233–2239. doi: 10.1128/AEM.03566-12
- The, H. C., Thanh, D. P., Holt, K. E., Thomson, N. R., and Baker, S. (2016). The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat. Rev. Microbiol.* 14, 235–250. doi: 10.1038/nrmicro.2016.10
- Thomas, C. M., and Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi: 10.1038/nrmicro1234
- Tomida, S., Nguyen, L., Chiu, B.-H., Liu, J., Sodergren, E., Weinstock, G. M., et al. (2013). Pan-genome and comparative genome analyses of *Propionibacterium acnes* reveal its genomic diversity in the healthy and diseased human skin microbiome. *MBio* 4, e00003–e00013. doi: 10.1128/mBio.00003-13
- Troy, E. B., and Kasper, D. L. (2010). Beneficial effects of *Bacteroides fragilis* polysaccharides on the immune system. *Front. Biosci.* 15, 25–34. doi: 10.2741/3603
- Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res.* 27, 626–638. doi: 10.1101/gr.216242.116
- Van Rossum, T., Ferretti, P., Maistrenko, O. M., and Bork, P. (2020). Diversity within species: interpreting strains in microbiomes. *Nat. Rev. Microbiol.* 18, 491–506. doi: 10.1038/s41579-020-0368-1
- Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* 60, 407–438. doi: 10.1128/mr.60.2.407-438.1996
- Versalovic, J., Koeuth, T., and Lupski, R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res.* 19, 6823–6831. doi: 10.1093/nar/19.24.6823
- Vos, M., Hesselman, M. C., Te Beek, T. A., van Pessel, M. W. J., and Eyre-Walker, A. (2015). Rates of lateral gene transfer in prokaryotes: high but why? *Trends Microbiol.* 23, 598–605. doi: 10.1016/j.tim.2015.07.006

- Wang, W.-L., Xu, S.-Y., Ren, Z.-G., Tao, L., Jiang, J.-W., and Zheng, S.-S. (2015). Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* 21, 803–814. doi: 10.3748/wjg.v21.i3.803
- Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci.* 74, 5088–5090. doi: 10.1073/pnas.74.11.5088
- Yutin, N., and Galperin, M. Y. (2013). A genomic update on clostridial phylogeny: gram-negative spore formers and other misplaced *clostridia*. *Environ. Microbiol.* 15, 2631–2641. doi: 10.1111/1462-2920.12173
- Zhai, R., Xue, X., Zhang, L., Yang, X., Zhao, L., and Zhang, C. (2019). Strain-specific anti-inflammatory properties of two *Akkermansia muciniphila* strains on chronic colitis in mice. *Front. Cell. Infect. Microbiol.* 239:9. doi: 10.3389/fcimb.2019.00239
- Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., et al. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 352, 565–569. doi: 10.1126/science.aad3369
- Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A. L. (2019). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* 570, 462–467. doi: 10.1038/s41586-019-1291-3