



OPEN

The Edible Plant Microbiome represents a diverse genetic reservoir with functional potential in the human host

Maria J. Soto-Giron¹, Ji-Nu Kim¹, Eric Schott¹, Claudine Tahmin¹, Thomas Ishoey¹, Tracy J. Mincer^{2,3}, Jillian DeWalt¹ & Gerardo Toledo¹✉

Plant microbiomes have been extensively studied for their agricultural relevance on growth promotion and pathogenesis, but little is known about their role as part of the diet when fresh fruits and vegetables are consumed raw. Most studies describing these communities are based on 16S rRNA gene amplicon surveys, limiting our understanding of the taxonomic resolution at the species level and functional capabilities. In this study, we characterized microbes colonizing tomatoes, spinach, brined olives, and dried figs using shotgun metagenomics. We recovered metagenome-assembled genomes of novel lactic acid bacteria from green olives and identified high intra- and inter-specific diversity of *Pseudomonas* in tomatoes. All samples were colonized by *Pseudomonas*, consistent with other reports with distinct community structure. Functional characterization showed the presence of enzymes involved in vitamin and short chain fatty acid metabolism and degradation of diverse carbohydrate substrates including plant fibers. The dominant bacterial members were isolated, sequenced, and mapped to its metagenome confirming their identity and indicating the microbiota is culturable. Our results reveal high genetic diversity, previously uncultured genera, and specific functions reflecting a likely plant host association. This study highlights the potential that plant microbes can play when consumed as part of our diet and proposes these as transient contributors to the gut microbiome.

Advancements in molecular methods over the past 30 years have expanded our knowledge of the vast extent of biological diversity on earth. The use of 16S rRNA gene sequences has enabled characterization of microbial communities in a wide range of habitats including the human body and in fluids including stool, prompting establishment of the microbiome field and industry. Diet has been identified as one of the main variables shaping the gut microbiome, as dietary intervention studies have demonstrated that the gut microbiota composition can reflect variations in food consumption despite immense inter-individual variation and the heterogeneous nature of stool that complicates representative sample collection¹⁻³.

In contrast to the efforts aimed to describe the microbiota in stool, very little exploration has been conducted in our foods and the microbes consumed in a typical diet⁴. Most of what we know about food microbiomes, especially as it relates to fresh fruits and vegetables, is related to agricultural pathogens and toxins. However, the recognition that our diets contain potentially beneficial bacteria and fungi is a new concept. Berg et al.⁵ coined the term *Edible Plant Microbiome* to identify the types of tissues, compartments, and plant groups that carry live microbes in our diets, especially when eaten raw. The overall view is that microbes colonizing internal and external plant tissues can play a role in human nutrition and health and provide insights of co-adaptation processes between plants, animals, and their microbiomes. For example, Wasserman et al.⁶ studied the microbiomes of cruciferous vegetables and showed they provide protection against fungal plant pathogens and anticancer molecules to the human host, indicating an important relationship between the plant host and its microbiota. Plant-associated bacteria play crucial roles in their host, including beneficial effects on the production of secondary plant metabolites, protection against pathogen colonization, plant growth promotion, among others^{7,8}.

¹Solarea Bio, 100 Edwin H Land Blvd, Cambridge, MA 02142, USA. ²Department of Biology, Wilkes Honors College, Jupiter, FL 33458, USA. ³Present address: Harbor Branch Oceanographic Institute, Florida Atlantic University, Fort Pierce, FL 34946, USA. ✉email: gtoledo@solareabio.com

However, some metabolites such as plant hormones produced in concert with the plant microbiome can also be relevant in the human gut⁹, suggesting a co-adaptation to colonize both the plant and animal host.

The microbial diversity of fresh fruits and vegetables commonly eaten raw has mainly been studied using both 16S rRNA gene sequences and cultivation approaches. For example, baby spinach and romaine, green leaf, iceberg, and red leaf lettuce^{10–13}, cruciferous vegetables^{6,13}, tomatoes^{14,15}, and fruits such as peaches, strawberries, grapes¹³, bananas¹⁶, and apples¹⁷ have all been studied by this approach. These studies have in most cases reported a microbial community dominated by Gammaproteobacteria, followed by Bacteroidetes, Actinobacteria, and Firmicutes. Kõiv et al.¹⁸ analyzed five tubers (potato, carrot, beet, neep, and Jerusalem artichoke) by separating the peels from the pulp and found the peels were enriched in Alphaproteobacteria and Actinobacteria, while the pulps were enriched in Gammaproteobacteria and Firmicutes. At the genus level, *Pseudomonas* seems to be the most common and abundant taxon across samples in most studies, indicating it is consumed in raw items that are considered part of a healthy diet. However, the extent of intraspecific genetic diversity and possible roles in its passage through the gastrointestinal (GI) tract are largely unknown. The abundance of viable bacteria and fungi living in leafy salad vegetables ranges from 8.0×10^3 to 5.5×10^8 CFU/g¹² and the bacterial fraction, based on 16S rRNA gene copy numbers in one organic apple, averages 4.85×10^7 16S rRNA gene copy numbers per gram¹⁷. Tubers carry from 1×10^3 to 5.5×10^8 CFU/g in their pulps¹⁸. The consumption of raw fruits and vegetables in salads is considered healthy and provides abundant and diverse microorganisms that may be functionally active along the passage through the GI tract of the consumer.

Fermented foods are another dietary component rich in microbes. They were consumed more frequently in the past as pickling and fermenting were means to preserve and extend food shelf life prior to the use of chemical preservatives and refrigeration. They carry endogenous communities of Lactic Acid Bacteria (LAB) that become enriched during the fermentation process. For example, brined olives carry significant amounts of *Lactobacillus* suggesting a possible probiotic role in olives^{19,20}. Another class of fermented foods is represented by cheeses and other dairy products. Passoli et al.²¹ analyzed 303 metagenomes from different fermented foods, cheese, yogurt, and dietary supplements and recovered Metagenome-Assembled Genomes (MAGs). Authors compared these MAGs to reference genomes and MAGs from human stool and established a link between some microbial members present in foods that can be shaping the gut microbiome.

Shotgun metagenomic approaches are required to characterize the microbial community at higher resolution including strain and functional gene levels. Cernava et al.²² analyzed the arugula phyllosphere and reported the resistome as part of the gene content with ecological significance. These types of analyses are useful to describe the microbial diversity and the linkage with metabolic pathways associated with plant fiber degradation, vitamin biosynthesis, and short chain fatty acid metabolism in plant-associated microbiomes.

In this pilot study we developed methods to extract microbial cells including the endophytic communities from plant tissues and sequenced them extensively to generate a shotgun metagenomic dataset from tomatoes, baby spinach, green olives, and dry figs. We observed a high degree of intraspecific diversity and possibly novel genera in the samples when MAGs were recovered and analyzed. In addition, we isolated the dominant bacterial species of each sample and sequenced their genomes. Our results suggest the Edible Plant Microbiome is a functional component of the diet that might contribute to the gut microbiome with potential implications for human nutrition and health.

Results

Microbial community structure of the edible plant microbiota. We focused our study on fruits and vegetables that are part of a common diet, easily accessible to consumers, and commonly consumed raw; therefore carrying viable microbes. The samples for this study were acquired from supermarkets and prepared in a similar way for human consumption by rinsing vigorously with tap water. The samples were then homogenized by blending, and the plant-associated microbiome was purified from plant cells by filtration, centrifugation, and the use of detergents to preferentially lyse plant cells and chloroplasts.

DNA was extracted from the purified microbiomes and analyzed by shotgun metagenomic sequencing. Plant cell removal efficiency was verified by fluorescence microscopy using DNA stains, indicating ranges between 70 and 95% efficacy, but in all samples, plant cells and DNA remained present in the shotgun metagenomic datasets (ranging from 0.1 to 69.3% of the total number of sequencing reads) (Supplementary Table 1).

After host DNA removal, shotgun metagenomic samples ranged from 3.6 to 23 million sequencing reads with an average read length of 100 bp. The coverage estimation of the microbial community based on read redundancy indicated that more than 80% of the communities were covered in the samples, suggesting our sequencing effort captured the majority of abundant microbes present in the sample (>0.1% abundance of the total) (Supplementary Fig. 1). Comparison of sequence diversity based on metagenomic reads (similar to alpha diversity obtained from Nonpareil curves) with other metagenomes from different environments showed that samples from this study present similar diversity to the ones seen in the human stool and slightly higher than Nunu, a west African fermented yogurt-like milk product, but lower than marine, fresh water, and soil samples (Fig. 1A).

Taxonomic annotation based on metagenomic reads indicated that the microbial communities were dominated by Proteobacteria in all samples, except for green olives in which the most abundant phylum was Firmicutes (69.7% of the total). Actinobacteria were also 10× more abundant in green olives compared to the other samples (4.2% vs. 0.4% on average). At the genus level, vegetables and fruits were dominated by different microbial members (Fig. 1B). For instance, *Pseudomonas* was the most abundant taxon in early girl tomatoes (54%) followed by *Rahnella* (22%) and *Pantoea* (9%). Baby spinach were dominated by *Pantoea* (30%) followed by *Erwinia* (20%) and *Enterobacter* (19%). Black mission figs were dominated by *Janthinobacterium* (38%), followed by *Pantoea* (12%) and *Herbaspirillum* (12%). Green olives were dominated by *Lactobacillus* (62%) followed by *Pseudomonas* (21%), *Acidipropionibacterium* (3%), and *Pediococcus* (3%). The microbial composition in these

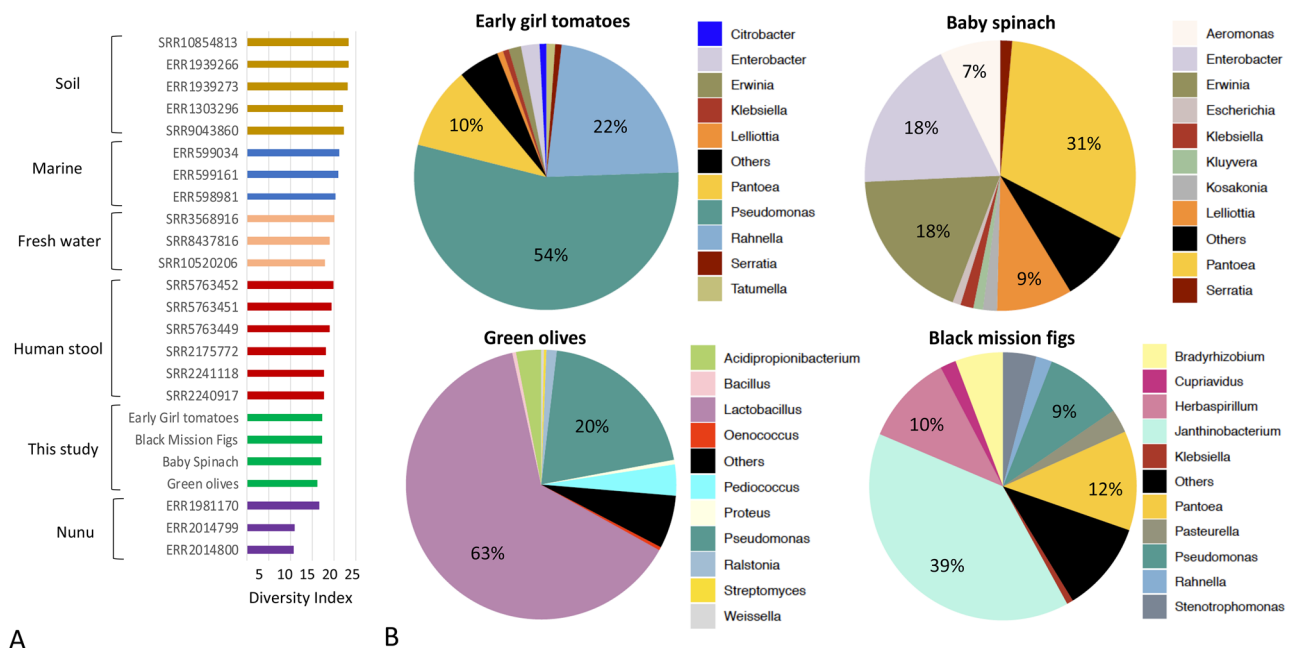


Figure 1. Diversity and taxonomic composition of shotgun metagenomes from fruits and vegetables analyzed in this study. **(A)** Comparison of alpha diversity index values of metagenomic samples from this study with other metagenomes from diverse ecosystems. Diversity index values were calculated based on the rarefied coverage of metagenomic reads in each sample using Nonpareil v3²³. Bars are colored by sample type and each sample is labeled by its accession number submitted at the NCBI Sequence Read Archive (SRA) or the EBI European Nucleotide Archive (ENA). Metadata associated with these metagenomic samples can be found in Supplementary Table 2. **(B)** Pie charts showing the relative abundance of the top ten most abundant microbial members at the genus level based on k-mer analysis using kraken2²⁴.

samples is consistent with previous reports on Gammaproteobacteria as a relevant group in plant-associated microbiomes and the genus *Pseudomonas*^{10–12,22}, a dominant member across samples. However, further comparisons between samples would require additional sampling as microbiomes can be variable in response to multiple factors such as farming practices.

The taxonomic profile based on 16S and 18S rRNA gene sequences recovered from the metagenomes showed that black mission figs harbored the highest proportion of sequencing reads belonging to fungi, mainly classified in the *Aspergillus* (30%), Unclassified Eurotiales (29%), and Unclassified Trichocomaceae (26%) genera (Supplementary Table 3). No archaeal sequences were detected in any of the metagenomes.

We also generated a strain culture collection from these samples using different media types and identified representative species from multiple genera that were present in high abundance in the metagenomes. For instance, bacterial isolates from *Pseudomonas*, *Rahnella*, and *Pantoea* were isolated from early girl tomatoes; *Pseudomonas*, *Pantoea*, and *Stenotrophomonas* were isolated from baby spinach; *Lactobacillus*, *Pseudomonas*, *Bacillus*, *Leuconostoc*, and *Pantoea* from green olives; and *Janthinobacterium*, *Herbaspirillum*, *Pantoea*, and *Pseudomonas* from black mission figs. Most of the isolates belong to known species represented in public databases except for one isolate from figs (*J. lividum*, 93.35% ANI), which likely represents a new species. Fragment recruitment plots of the genome of some of these bacterial isolates against the metagenomic reads for each sample showed the presence of the isolate in the sample (nucleotide identity > 99%) in high abundance. For instance, the average sequencing depth of *P. fluorescens* from early girl tomatoes was 12.7 × while the one for *J. lividum* from black mission figs was 53.9 ×. In addition, to the target genome, there is a gradient of identities that reveal possible ecotypes (Fig. 2).

To identify individual uncultured populations that represent abundant members in the microbial community, genome binning analysis was performed by assembly of sequencing reads into contigs and then binning of these contigs into metagenome-assembled genomes (MAGs) (Supplementary Table 4). Quality of MAGs was evaluated following the standard criteria reported by Bowers et al.²⁵. MAGs with more than 90% completeness and less than 5% contamination were classified as high quality, while MAGs with > 50% completeness and less than 10% contamination were classified as medium quality. In total, 12 MAGs (8 high and 4 medium quality) were recovered from the metagenomes with identities matching dominant genera seen by Kraken2²⁴.

The majority of recovered MAGs were obtained from green olives (n = 7), likely because of the higher average coverage of the microbial community (Nonpareil coverage: > 95%) and low complexity of the community (*Lactobacillus* and *Pseudomonas* encompass approx. 82% of the total abundance). Five near-complete MAGs were affiliated with lactic acid bacteria confirming the dominance of this group in green olives with high coverage for each. More specifically, MAG3 showed 99.7% whole-genome average nucleotide identity (ANI)²⁶ to *Lactobacillus buchneri*. MAG9 showed 98.8% ANI to *Lactobacillus acetotolerans*, but MAG1, MAG10, and MAG6 could not be assigned to a known species following the 95% ANI threshold for species demarcation or to a genus (> 80%

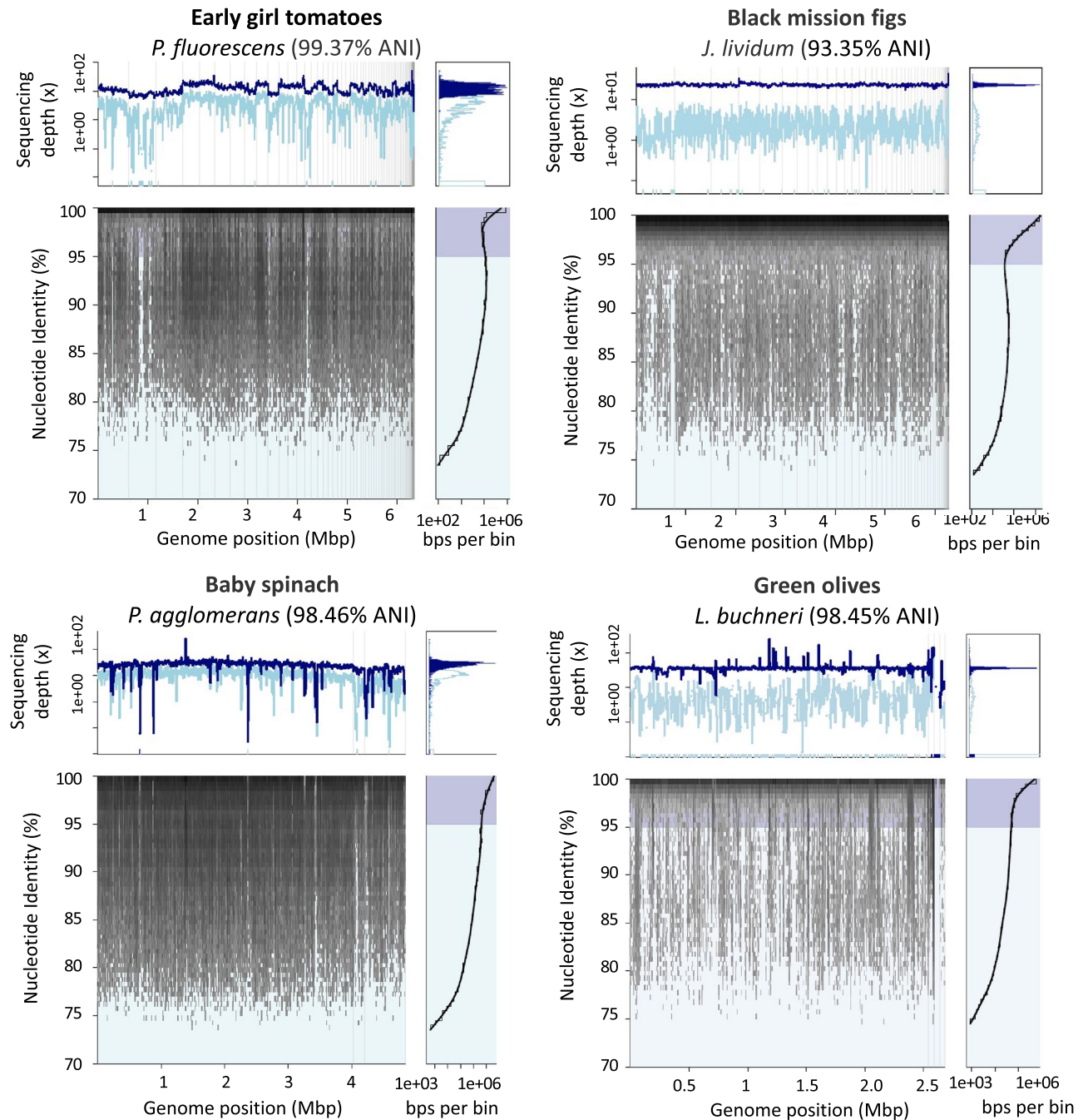


Figure 2. Fragment recruitment plots of bacterial isolates derived from fruits and vegetables and fermented foods against metagenomic sequencing reads from each sample. The y-axis in the top left panel corresponds to the average sequencing depth values in logarithm scale of reads that mapped to the genome sequence (x-axis). The y-axis in the bottom left panel corresponds to the percentage of nucleotide identity (%) of each mapped read and the x-axis to the position of the read on the genome. The right bottom panel shows the identity histogram of mapped reads (logarithmic scale). The dark blue peak in the histogram (right top panel) corresponds to the average coverage of mapped reads with $\geq 95\%$ nucleotide identity, while the light blue indicates the coverage of sequencing reads with $< 95\%$ nucleotide identity. A sequence-discrete population is represented by reads with high nucleotide identity ($> 95\%$ nucleotide identity, indicates the cut-off for species demarcation) to the reference genome sequence and with even coverage across the reference sequence.

ANI), most likely representing undescribed microbial members²⁶. These MAGs can only be assigned to the family Lactobacillaceae. One of the two medium quality MAGs was assigned to the genus *Pseudomonas* and the other one to Gammaproteobacteria.

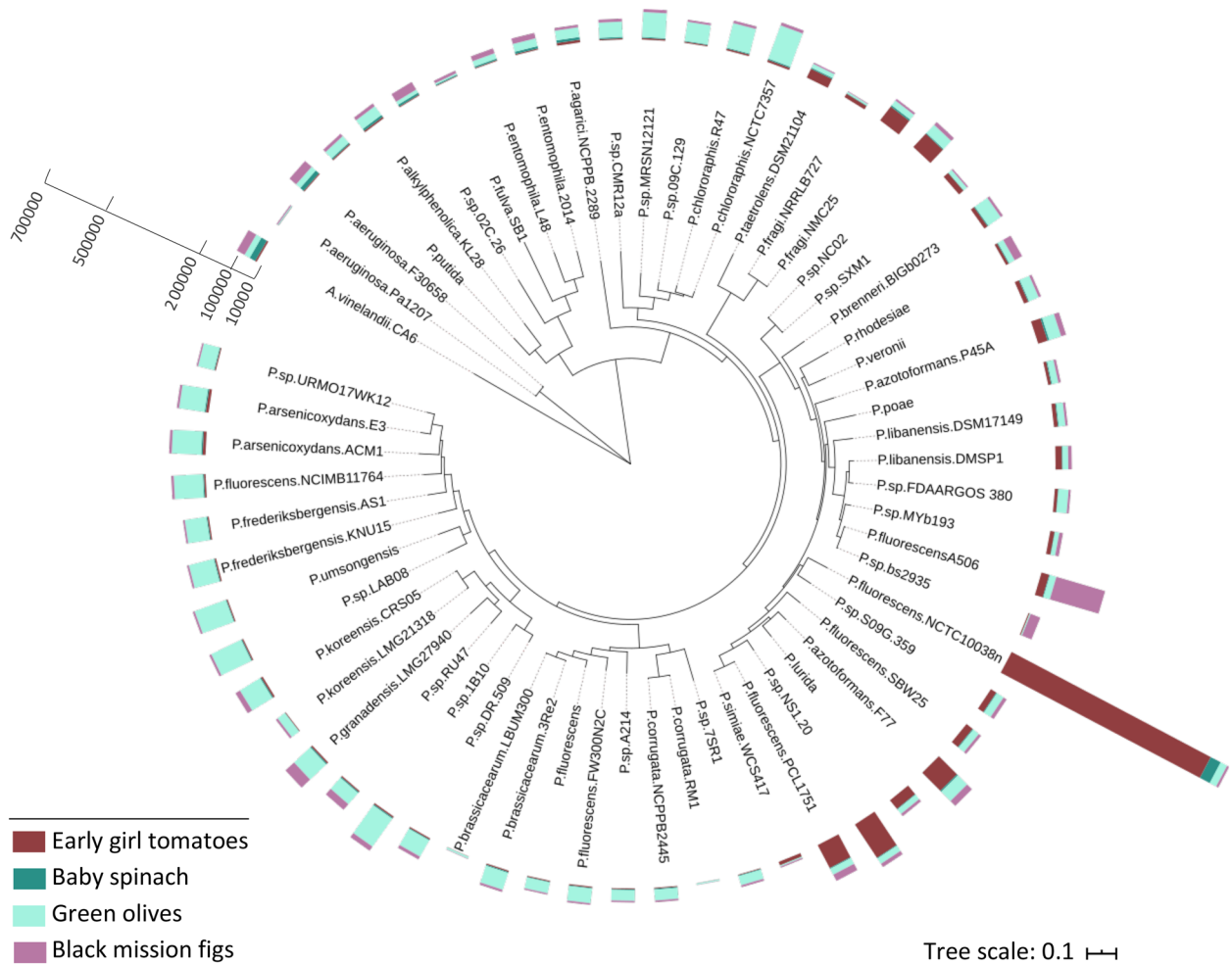


Figure 3. *Pseudomonas* population diversity in fresh fruits and vegetables. Maximum-likelihood tree constructed using 100 universal single copy genes from 26 *Pseudomonas* reference genomes. *Azotobacter vinelandii* was used as an outgroup. Bar plots colored by sample type show the number of sequencing reads that mapped each genome. Phylogenetic tree was visualized and edited using iTOL v6 (<https://itol.embl.de/>)²⁷.

Two medium quality MAGs from the metagenomic sample taken from baby spinach were annotated as *Erwinia persicina* (93.4% ANI) and *Aeromonas* sp. (97.7% ANI). In black mission figs, three high quality MAGs were recovered affiliated with *Janthinobacterium* (93% ANI), *Pantoea* (91% ANI), and *Herbaspirillum* (87.5% ANI) genera, most likely representing undescribed bacterial species given their nucleotide divergence relative to known species. These MAGs are consistent with the taxonomic description of the samples based on Kraken2.

Isolation of representative bacterial strains from each sample and genome sequencing. In order to evaluate if the bacterial diversity in the samples was amenable to culture, strains were isolated using different formulations of solid media and selective conditions. In addition, colony forming units (CFU) were counted on the microbial preparations showing counts between 1×10^4 and 1×10^6 CFU/g consistent with other studies. The most abundant species on each sample was isolated and its genome was sequenced. Fragment recruitment plots of the dominant genome of each sample against the metagenome from where it originated revealed high relative abundance in the sample, high sequence identity to the recruited genome, and in most cases a high degree of intraspecific diversity (Fig. 2). Some of the isolates from green olives matched the MAGs reconstructed from the metagenome as in the case of MAG3 and the isolate *L. buchneri*.

Diversity of *Pseudomonas* populations across samples. The distribution of metagenomic reads mapped onto a phylogenetic tree of *Pseudomonas* constructed with 100 marker genes from representatives of different species revealed the presence of multiple *Pseudomonas* species across samples as well as different distributions of these species per sample (Fig. 3).

Due to the elevated genetic diversity previously observed in *P. fluorescens* in soil and other habitats/environments, its importance as a plant growth-promoting bacterium, and presence and abundance across samples, we were interested in evaluating the level of intraspecific population diversity of this taxon in food ingredients that are part of our diet (Supplementary Fig. 2). Here, the genome from the reference strain *P. fluorescens* NCTC

10038 was used to recruit metagenomic reads and visualize the nucleotide diversity of *Pseudomonas* populations in the samples. Fragment recruitment plots showed that early girl tomatoes contained high-identity nucleotide sequencing reads evenly across the entire reference genome, most likely belonging to members of the same species (>95% nucleotide identity and 12X of average sequencing depth). In addition, diverse co-occurring populations seem to be present as well (mapped reads with 80% to 95% nucleotide identity), indicating other species in the same genus. Baby spinach also showed reads mapping to the reference genome, but in less abundance (0.61X coverage) and lower species diversity. *Pseudomonas* was not one of the top 10 genera in this sample but the *P. fluorescens* NCTC 10038 genome could be recovered. Metagenomes from green olives and figs do not contain *P. fluorescens* NCTC 10038 but display more distantly related populations present in the sample (nucleotide identity ranging from 80 to 95%).

Genomic diversity and functional profile of LAB-like MAGs recovered from green olives. Since five MAGs belonging to LAB from green olives were recovered, we were interested in evaluating their genome diversification and function which can reflect ecological niches in the community. For this, metagenomic reads were recruited against the MAG sequences in the sample (Supplementary Fig. 3). The resulting fragment recruitment plots showed that LAB-like MAG6 and MAG3 harbored considerable intraspecific diversity (mapped metagenomic reads with nucleotide identity values between 95 and 100%), and that moderately-to-distantly related populations (reads with <90% identity) are co-occurring in green olives. In contrast, recruitment of reads by the LAB-like MAG1 and MAG10 indicated one dominant population (48.7X and 47.2X mean sequencing depth, respectively) with less population sequence diversity (e.g., few reads mapping >95% and 99% nucleotide identity to MAG sequence). The conditions of pickled green olives favor the dominance of multiple LAB species, some with high genetic diversity and others distantly related to known species and likely representing a novel genus.

LAB, in particular, *Lactobacillus*, are some of the best characterized and widely studied probiotic bacteria, and have been isolated from sources such as stool, milk, and soil. Their probiotic attributes include antimicrobial activities, immunomodulatory cell wall components, adhesive pili, and enzymes for SCFA production^{20,28,29}. We screened for genes with probiotic-linked mechanisms in LAB-like MAGs with functions that are related with host colonization and anti-inflammatory properties (Fig. 4).

The LAB-like MAGs showed the presence of MFS sugar secondary transporters, Phosphotransferase Transporter System (PTS) permease families, and anchor proteins including LPXTG cell wall anchor domain-containing proteins that have previously been involved in conferring beneficial effects to the host. PTS permeases in *Lactobacillus* are associated with fermentation of various sugar substrates³². In addition, PTS permeases have been associated with the long-persistence phenotype in the gastrointestinal tract observed in the probiotic *L. johnsonii* NCC533³³. LPXTG cell wall anchor domain-containing proteins have been widely studied in *Lactobacillus* because of their role in epithelial cell adhesion in the gastrointestinal tract and in host immune response modulation^{29,34}. Additional beneficial features associated with probiotics are tolerance to environmental stress, including acidity. The gene coding for the heat shock protein 60 family co-chaperone GroEL that confers environmental stress resistance was identified in the MAGs, which allows *Lactobacillus johnsonii* to bind to mucin and intestinal epithelial cells to enhance intestinal persistence³⁵. Altogether, these data suggest that green olives and their microbes can offer some of the same functionalities provided by commercial probiotics.

Functional annotation of the edible plant microbiota. Beyond taxonomic annotation, shotgun metagenomic sequencing facilitates extensive analysis of all functional genes present in a microbial community. With this, we can begin to better understand not only who is present, but what functions they can perform. Here, functional annotation of genes from the metagenomes using UniProt DB³⁶ revealed that the majority of the genes predicted in the sample from early girl tomatoes, baby spinach, and green olives were annotated (>90% of the total genes), while only around 65% of the genes predicted in the sample from figs were annotated (Table 1). The lower annotation of the genes in figs can be explained by a large fungal population (10%) and the paucity in fungal databases and genomes. Prediction of Gene Ontology (GO) biological processes showed that a significant proportion of metagenomic reads were related to housekeeping functions such as nucleic acid metabolism, protein metabolism, and carbohydrate metabolism (Supplementary Table 5).

Comparison of the annotated genes across the four samples revealed that not all biological processes are equally distributed. In baby spinach, we identified a high frequency of reads assigned to processes including the pectin catabolism, glycerol ether metabolism, trehalose biosynthesis, L-threonine catabolism to glycine, and L-serine biosynthetic process compared to the other samples. In figs, we found that the xyloglucan metabolism and the poly-hydroxybutyrate biosynthesis were more frequent. In tomatoes, glyoxylate catabolism was more frequent in comparison with the other samples.

We also identified the presence of metabolic pathways associated with human gut microbiome homeostasis including biosynthesis of glutamate, terpenoids, riboflavin, thiamine, folic acid, glutamine, and biotin. Carotenoid biosynthetic process was present in all samples, except in green olives, while response to redox state, lactose metabolic process, as well as lipoteichoic acid, teichoic acid, and isopentenyl diphosphate biosynthetic process was identified only in green olives. Overall, these data revealed the Edible Plant Microbiome harbors a diverse genomic repertoire that varies among samples.

Diversity of carbohydrate-active enzymes. To further understand the functional capacity of these plant microbiomes, we investigated the abundance of carbohydrate-active enzymes (CAZymes), a group of enzymes that are involved in the breakdown of plant cell walls with potential for further conversion to beneficial metabolites in the human intestinal tract. CAZymes are categorized based on protein sequence and

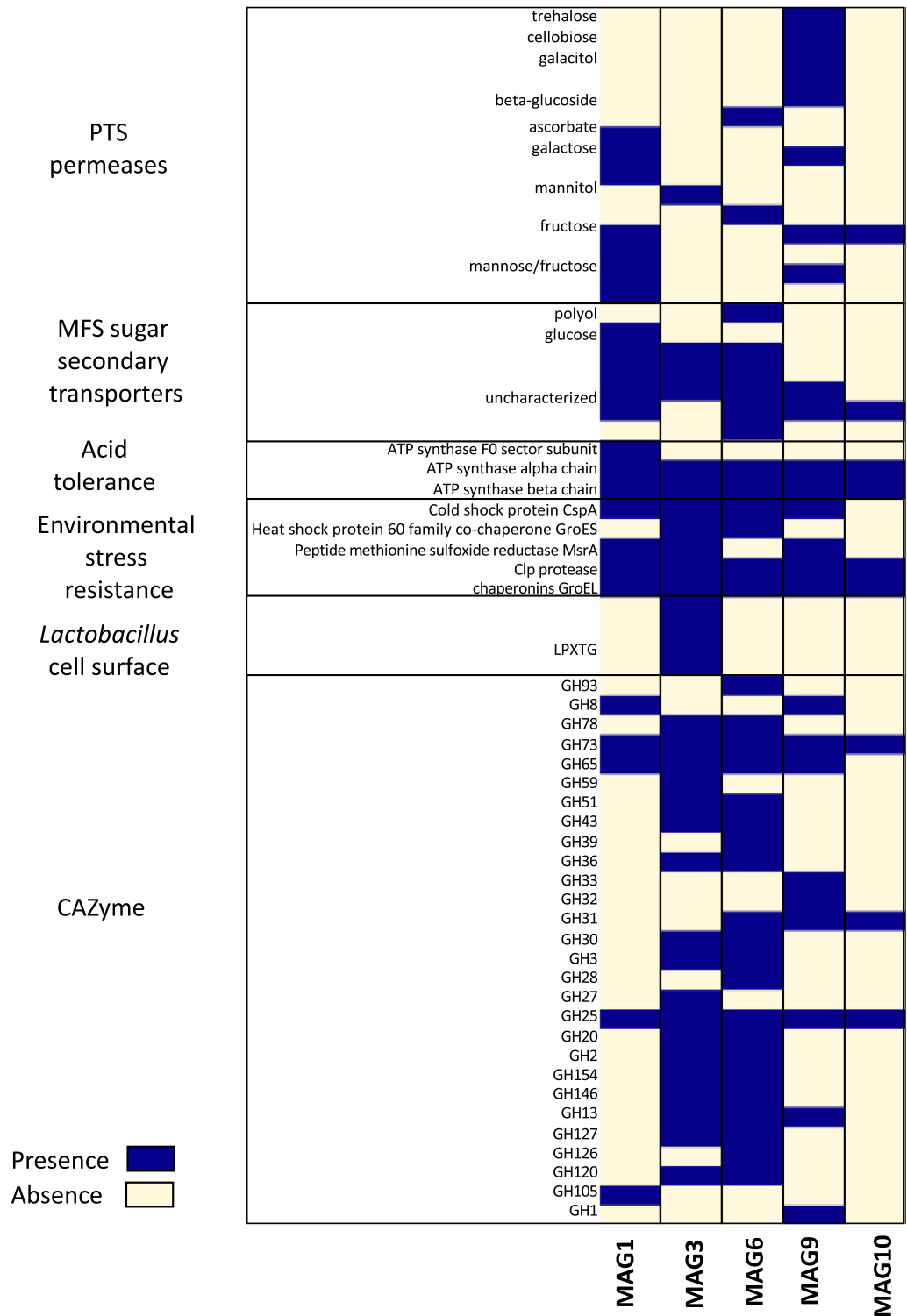


Figure 4. Functional profile of LAB-like MAGs recovered from the green olives metagenome. Functional properties identified in the LAB-like MAGs that have been previously reported in commercial probiotic strains^{30,31}. These proteins include Phosphotransferase Transporter System (PTS) permeases, secondary sugar transporters from the major facilitator superfamily (MFS), LPXTG cell wall anchor domain-containing protein, among others.

structural similarities including glycosyltransferases (GTs), glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), auxiliary activities (AA), and carbohydrate-binding modules (CBMs). Puta-

	Early girl tomatoes	Baby spinach	Green olives	Black mission figs
No. of annotated genes	841	1011	881	748
AA (%)	3.1	2.0	1.2	6.1
CBM (%)	5.0	5.4	2.3	3.6
CE (%)	5.4	3.7	4.7	9.1
GH (%)	43.2	46.9	56.9	38.6
GT (%)	40.9	39.8	34.2	40.6
PL (%)	2.5	2.3	0.8	1.9

Table 1. Functional annotation of predicted genes from metagenomics samples using carbohydrate-active enzymes (CAZyme) and UniProt databases. Columns indicate the number of predicted proteins, percentage of each CAZyme module. AA auxiliary activity, CBM carbohydrate binding module, CE carbohydrate esterase, GH glycoside hydrolase, GT glycosyltransferase, PL polysaccharide lyase, and the percentage of annotated proteins using UniProt database³⁷.

tive CAZymes in the plant-associated microbiota characterized in this study were identified using predicted open reading frames (ORFs) from the metagenomes and the CAZymes database³⁸. GHs and GTs were the most abundant CAZymes predicted in the metagenomes (approx. 30 to 57% of the total). Fewer percentages of AAs (approx. 1–6%), CBMs (2–5%), CEs (3–9%), and PLs (1–4%) enzymes were identified in the samples (Table 1 and Supplementary Table 6).

Specifically, 79 of the 133 GHs families in the CAZy database were detected among samples. GHs are enzymes that hydrolyze the glycosidic bond leading to the formation of carbohydrate hemiacetal³⁸. Our results showed common GHs present in all metagenomes with GH13, GH23, and GH1 being the most frequent families. GH13 is one of the most common families identified in the human gut microbiome and is characterized for breaking down starch³⁹. Starch is the major alpha-glucose polymer that plants synthesize to store energy and an important carbohydrate source consumed worldwide⁴⁰. We also observed GHs that were distinctly distributed in specific samples and that have been associated with beneficial effects in the human gut. For example, the enzyme tomatinase (EC 3.2.1.-) is a member of GH10 along with xylanases. Tomatinase facilitates the degradation of the antifungal tomatine into nonfungitoxic substrates: tomatidine and beta-lycotetraose. The presence of fungi in black mission figs may be linked to the presence of GH10 enzymes rendering a fungal toxin inert.

GH31 was prevalent in green olives and baby spinach. The enzyme sulfoquinovosidase (EC 3.2.1.) in GH31 hydrolyses alpha-sulfoquinovoside. Sulfoquinovoside is produced by plants, degraded by aerobic bacteria into carbon and sulfur⁴¹, and present in leafy green vegetables such as spinach⁴². Sulfoquinovosidase has also been implicated in benefitting human gut commensals⁴¹. GH1 comprises B-glucosidases involved in cellulose degradation, one of the most abundant fibers in leafy vegetables. GH1 enzymes are used by gut microbes to hydrolyze glycans into glucose and glycolytic precursor glucose-6-phosphate. GH65 containing trehalose phosphorylases was the most abundant in the green olive sample. Trehalose is a disaccharide which prevents desiccation in plants and is present in olive leaves⁴³. Trehalase is also present in the small intestine of vertebrates, and bacteria such as *Enterococcus* have been shown to metabolize trehalose. The family GH28 contains enzymes related to pectin degradation. Pectin is considered a prebiotic, and bacteria belonging to the genera *Erwinia* and *Pseudomonas*, present in tomatoes, are common pectinolytic enzyme producers⁴⁴. GH28 enzymes produced by bacteria have been found to be active in the oral cavity commencing the breakdown of pectin when ingested⁴⁵. Taken together, these data further support the notion that the microbes present in plants could provide metabolic benefit to humans when consumed as part of the diet, as they contain a large number of carbohydrate active enzymes that play crucial roles in human health and the substrates which are present in plant cell walls.

Metabolic pathways associated with SCFAs production. The most widely studied microbial metabolites in relation to human health are short chain fatty acids (SCFAs). SCFAs are derived from the microbial fermentation of complex carbohydrates, and act as both important regulatory signaling molecules and an additional fuel source for the host^{46,47}. The most prevalent SCFA pathways across the four samples analyzed were lactate oxidation and L-lysine catabolic process to acetate (Supplementary Fig. 4A). Lactate metabolic process was also identified in the metagenomes of both baby spinach and green olives. The least common pathways represented across the metagenomes were butyrate metabolic processes, as the metagenomes contained the fewest reads mapping to butyrate kinase. Interestingly, the sample with the broadest SCFA metabolic potential was the pickled green olive, suggesting the pickling process may select for microbes with extensive SCFA productive capacity. In agreement with Supplementary Fig. 4A, the enzymes with the greatest representation across samples are those related to acetate (acetate kinase) and lactate (D-lactate dehydrogenase) production. These data indicate that the microbiomes of fruit and vegetables contain the genes and pathways required to produce acetate and propionate, a family of microbial metabolites associated with human health, but are missing butyrate biosynthesis as it is mostly linked to strict anaerobes not seen in the samples.

Biosynthetic gene pathways associated with vitamin production. Microbes are known to synthesize vitamins, which are essential for several metabolic reactions in their respective hosts. The identification of enzymes that participate in vitamin biosynthesis and uptake transporters in the metagenomes of this study

was based on genes annotated to KEGG Orthologies (KOs) predicted as previously identified in human fecal metagenomes⁴⁸ and in the probiotic strain *Bacillus clausii* ENTPro³⁰ (Supplementary Fig. 4B).

Our results showed the presence of genes encoding enzymes in the plant-associated metagenomes that are part of biosynthetic pathways for cobalamin, biotin, and pantothenate. Few clusters of genes were observed for folate, menaquinone, and niacin biosynthesis. Moreover, a high proportion of genes (33/66) associated with cobalamin metabolism was detected in the metagenomes and a slightly higher abundance of cobalamin metabolic genes was found in early girl tomatoes and green olives (Supplementary Fig. 4B). Cobalamin (vitamin B12) is an essential cofactor for humans synthesized by some bacteria and archaea and it has been involved in mediating microbe-microbe interactions and host-microbe communication⁴⁹. Cobalamin deficiency can result in anemia, nervous system disease, and neuropathy⁵⁰.

Discussion

The presence and the importance of microbes in food have been previously studied, suggesting that the microbes we consume in our diet can modulate the gut microbial structure. Based on these findings, together with the Edible Plant Microbiome concept previously described by Berg et al.⁵, we were interested in characterizing the diversity and functional capabilities of the Edible Plant Microbiome in common dietary fruits and vegetables. In this pilot study, we showed that microbes residing in fresh fruits and vegetables harbor a great deal of taxonomic and functional diversity that could potentially play a previously unappreciated role in shaping the human gut microbiota with downstream implications for human health. Our study was made possible by developing plant cell removal/filtration methods and performing shotgun metagenomic sequencing to provide enough sequencing coverage of the microbial communities (Supplementary Fig. 1).

Consistent with previous studies of plant associated microbiomes, Proteobacteria, Firmicutes, and Bacteroidetes were the most abundant phyla across samples (Fig. 1B). At the genus level, we observed that each sample harbored a different community structure, which may be driven by adaptations to distinct environmental factors including farming, and processing (e.g., figs were dried, and olives were pickled). We observed that several *Pseudomonas* species were present in all samples, confirming its ecological adaptation to colonize multiple crops and diverse metabolic capacity to use different organic compounds as carbon sources⁵¹. Specifically, *P. fluorescens* has been recognized as a bacterial complex with strains assigned to more than 50 species due to its elevated genetic diversity and gene content differences⁵². Although the genomic diversity of this taxon has been widely studied in soil and plants because of its relevance as a biocontrol agent and plant growth properties, little is known about its role and the extent of diversity in daily food items.

Fragment recruitment plots of the *P. fluorescens* reference genome against metagenomic reads from early girl tomatoes and baby spinach revealed the presence of an abundant population as well as a gradient of nucleotide diversity, most likely representing multiple discrete populations. In the case of early girl tomatoes, we observed one abundant genome population co-occurring with populations from other species, reflecting genomic differences associated with distinct ecological roles (Supplementary Fig. 2). This result echoes previous reports in other environments such as soil and lakes where microbial communities harbor discrete, but still related, populations^{53–55}.

This genomic diversity can represent different ecotypes adapted to changes in environmental conditions through the plant lifecycle, soil environment, farming practices, and post-harvest handling. Therefore, each time we eat fresh fruits and vegetables we consume a variety of *Pseudomonas* species. Tracing studies will reveal their survival through the gut and potential role in the human microbiome. In addition, the poor representation of plant-associated microbiomes and the updating of new species in public databases constrain our current understanding of the ecological role of these microbes in the human gut microbiota. MAGs belonging to *Pseudomonas* have been previously recovered from gut metagenomes of human populations around the world^{56,57}, but it remains unknown whether they are commensals or transient members in the GI tract.

Green olives are part of the Mediterranean diet and are widely consumed around the world. They have been recognized as an important food with respect to human health as they are rich in phenolic compounds, fatty acids, and antioxidant properties^{58,59}. Moreover, during the pickling process, green olives are enriched by LAB, mainly *Lactobacillus* populations, many of which (i.e., *L. plantarum*, *L. pentosaceus*) have been recognized as probiotics due to their health benefits in humans^{19,20}. By applying binning analysis, we were able to recover five near-complete MAGs that were closely related to *L. buchneri* and *L. acetotolerans*, 99.7% and 98.2% ANI respectively, as well as MAGs that might represent novel genera within the Lactobacillaceae family based on the ANI metrics for genus definition (>80% ANI) (Supplementary Table 4). Fragment recruitment plots of these MAGs against the metagenomic reads showed a high degree of nucleotide diversity of LAB populations that are co-habiting in pickled green olives. This taxonomic diversity reflects the diverse functional profile observed in gene content differences among MAGs such as genes involved in sugar uptake, stress response, adhesion mechanisms as well as distinct CAZyme families targeting multiple substrates. This genomic diversity might be associated with the ability to colonize different niches in the plant, survive during the pickling process, and tolerate the human stomach pH to reach different compartments of the GI tract. Still, it remains unexplored the impact that the microbial diversity observed in green olives can have during its transit in the GI tract as well as in the resident *Lactobacillus* populations, that represent on average 0.3% in the colon⁶⁰ and 6% of the total bacterial cells in the duodenum⁶¹.

Fungi constitute a significant component of the plant microbiome. Among the four metagenomes analyzed here, dried figs showed the highest number of 18S rRNA sequence reads mapping to *Aspergillus* detected at relatively high frequency. The presence of this fungus and other filamentous species has previously been described in dried figs and their enrichment in the fruit has been linked to environmental conditions such as humidity and temperature during the drying, post-harvest, and manufacturing processes that favor fungal growth⁶². Some

of these fungi can produce mycotoxins, including aflatoxin B1, which has carcinogenic effects⁶³, but the limited sequencing depth prevents us from assessing whether mycotoxins are present in the metagenome.

SCFAs are the most widely studied microbial metabolites because of their relation to human health. In comparison with the other metagenomes, green olives showed the highest abundance of enzymes participating in lactate, propionate, and acetate production. The production of SCFAs has been experimentally detected in the Moresca and Kalamata cultivars⁵⁹ and in the Spanish-style green table olives, Gordal and Manzanilla olive cultivars^{28,64}. Lactate is produced during fermentation by LAB, which was the most abundant taxa in the sample. Moreover, *Acidipropionibacterium* has been positively correlated with propionic acid, acetic acid, and methyl propanoate, while *Ruminococcus* with butyric and propionic acids²⁸. Although *Ruminococcus* was not detected in the sample as it is a gut commensal, there are likely other microbial members that could be participating in the production of butyric acid but are present in low abundance or absent in fresh fruits and vegetables.

Vegetables, fruits, and grains are an important source of carbohydrates and dietary fibers in our diet. In particular, dietary fibers are comprised of cell wall polysaccharides that are indigestible to the human host, but are used as fermentable substrates by microbes residing in the large intestine to produce SCFAs^{39,65}. There is a large variety of plant cell wall polymers, each with unique chemical bonds that are degraded primarily by specific glycosyl hydrolase (GH) enzyme families. Analysis of the CAZyme families in the metagenomes (Supplementary Table 6) indicated that GH13 and GH23 were the most prevalent families as they are involved in breaking down starch, one of the most abundant and ubiquitous biopolymers derived from plants. In addition, certain families were more frequent in specific samples.

Overall, our study highlights a high microbial diversity in raw fruits and vegetables. Specifically, the role of genomic and functional diversity identified in *Pseudomonas* and *Lactobacillus* needs additional attention, since 100's million cells are consumed in fresh food items like apples¹⁷. Moreover, there is a lack of research on studying which part of the edible plant microbiota become possible autochthonous (resident) and/or complementary allochthonous (transient) members of the human gut, and the extent to which these microbes influence, if any, ecological processes such as stability, resilience, and microbial *in situ* evolution.

The large and widespread functional diversity observed in the edible plant microbiome suggests that they may need to be considered in studies of the human gut microbiome, especially when differentiating microbiome-mediated clinical outcomes and in particular during dietary interventions that contain them. This preliminary study warrants further investigation with larger sample size and samples from different sources to evaluate activity and colonization along the many sub-environments of the GI tract and related host effects in health and disease.

Methods

Sample collection and microbial isolation. A set of commonly consumed fruits and vegetables (Supplementary Table 1) was purchased in 2017 in a farmers' market or in a supermarket in San Francisco, California to represent the diversity at the point of sale. The plant materials used in this study were farmed and sold in the United States. We have purchased them at the point of sale and are in compliance with international laws, protocols, and regulations for the Convention of Biological Diversity. Every sample was processed in the same way as it is typically done for consumption, consisting of rinsing with tap water vigorously and blotting dry. For every sample, 50 g were weighed and blended in a kitchen-grade blender (Ninja) for 30 s with 200 ml of a homogenizing buffer containing phosphate-buffered saline (PBS) pH 7.2 and 1% Triton X and 2% 0.5 M EDTA pH 8.0 (10 mM final concentration EDTA). After blending, the slurry was filtered with a coarse and fine kitchen grade strainer and a 40 mM strainer (Cole-Parmer 410-0001-FLS-CP Cell Strainer) sequentially. The filtered suspension was spun down at 5000 rpm for 10 min at 4 °C to pellet down the bacterial and fungal cells and resuspended in PBS for DNA extraction.

To quantify the total bacterial load in the samples, 50 g of sample material was weighed out into a blender cup previously sterilized with 70% EtOH. 200 mL of sterile PBS was added to the blender cup and the sample was blended for 90 s before being strained through a Miracloth (Available from Millipore Sigma 475855). The filtrate was centrifuged for 10 min at 5000 rpm. The supernatant was aspirated. To obtain representative colonies, pelleted material was resuspended in 5 mL of PBS and diluted from 10⁰ to 10⁻⁷. The filtered suspension was plated on a variety of culture media including standard recipes of R2A, MRS, PDA and TSA agar plates under aerobic or anaerobic conditions, different pH, and temperature. Single colonies were inoculated in liquid medium for 2 days and used for DNA extraction.

DNA extraction and sequencing of metagenomes and whole genomes. To extract DNA from individual colonies, the Zymo quick DNA bacterial/fungal kit (Zymo) was used. DNA quality and concentration were measured using Nanodrop and Picogreen fluorescent quantification. Genomic libraries were built using the Nextera DNA library preparation kit (Illumina Inc) and sequencing was performed using an Illumina MiSeq instrument (Illumina Inc) with 2 × 250 bp paired-end reads. Alternatively, other bacterial genomes were sequenced using an Oxford Nanopore and their genomes were assembled using flye v2.8.3⁶⁶.

To extract DNA from the microbial pellets obtained from fresh fruits and vegetables, the MagaZorb DNA extraction kit (Promega) was used. DNA quality and concentration were measured using Nanodrop and Picogreen fluorescent quantification. DNA libraries were built using the Nextera XT library preparation kit (Illumina Inc) and DNA sequencing was performed using an Illumina HiSeqX instrument (Illumina Inc) using a 2 × 150 bp flow cell.

Whole-genome assembly and shotgun metagenomic analyses. Raw reads from genome sequences were filtered for quality control using Solexa QA (Cox et al. 2010) with a Phred score >20 and a minimum length of 50 bp. Trimmed reads were assembled de novo using IDBA-UD v1.1.3 with pre-corrections

and default parameters⁶⁷. Contigs smaller than 1 Kb were filtered from the assembly. Genome completeness and contamination were evaluated using CheckM v1.07⁶⁸.

Raw paired-end reads from the metagenomes were processed for quality control with Solexa QA⁶⁹ for trimming and removing of Illumina adaptors using a Phred score > 20 and minimum fragment length of 50 bp. Host DNA removal from the metagenomes was performed by mapping trimmed reads to each plant reference genome using Bowtie2⁷⁰ with default parameters. Nonpareil v3.0²³ with default parameters was used to estimate the sequencing average coverage and diversity (similar to Shannon index) for each sequenced and filtered library.

De novo metagenomic assembly, recovery of metagenome-assembled genomes (MAGs), and MAG quality control were performed as described previously⁵⁷. Briefly, assembly was performed using MEGAHIT v1.1.1 with default parameters⁷¹, binning analysis using MetaBAT2 v2.12.1; option '-m 1500'⁷², and estimation of the completeness and contamination of each MAG using CheckM v1.07 with lineage specific workflow⁶⁸. Fragment recruitment plots for bacterial isolates and MAG contigs were constructed using the *enveomics.R* package v1.4.1 from the *Enveomics* collection⁷³.

Taxonomic annotation of MAGs was performed by identifying the closest relative genome using reference genomes available at RefSeq from NCBI as of November 2019 and mash distances using Mash v2; option '-s 10000' for sketching⁷⁴. Then, the Average Nucleotide Identity (ANI) values were computed between each pair in order to define at which taxonomic level the MAGs belong to using the species (> 95% ANI) and genus (> 80% ANI) scores accordingly²⁶. Taxonomic classification of metagenomic reads was carried out using Kraken2 with default parameters²⁴.

Phylogenetic analysis of 26 *Pseudomonas* species present in the metagenomic samples with more than 0.1% relative abundance was conducted by using reference genomes available at NCBI and based on 100 universal single copy marker genes from each genome. Identification of single copy genes was performed using the script HMM.essential.rb⁷³ and HMMer 3.0⁷⁵, then aligned using muscle v3.8⁷⁶ and concatenated using the script Aln.cat.rb⁷³. A maximum-likelihood tree was constructed with RAXML 8.2.12⁷⁷ using the GTRGAMMA model and 100 bootstraps. The best scoring tree was visualized and edited using iTOL v6 (<https://itol.embl.de/>)²⁷.

Protein-coding genes in assembled contigs were identified by Prodigal⁷⁸ with default parameters. Functional annotation was performed using BLASTp⁷⁹ searches of amino acid sequences against the UniProt/SwissProt database³⁶ using a cut-off for a match of at least 40% amino acid identity and 70% of the query protein length. The abundance of protein functions in each dataset was calculated as reads per kilobase per million mapped reads (RPKM). Proteins were assigned to the functional categories using Gene Ontology terms⁸⁰. Functional Mapping and Analysis Pipeline⁸¹ was applied to estimate gene family abundance values associated with metabolic pathways and operons of interest using KEGG Orthologous groups (KOs) and the KEGG enzyme database⁸². Prediction of carbohydrate-active enzymes (CAZymes) in the metagenomes using run-dbcanc v2.0.6⁸³ and the dbCAN CAZyme database³⁸.

All metagenomic samples were deposited in the NCBI's Sequence Read Archive under the project PRJNA635436.

Received: 12 February 2021; Accepted: 24 November 2021

Published online: 15 December 2021

References

- Zmora, N., Suez, J. & Elinav, E. You are what you eat: Diet, health and the gut microbiota. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 35–56 (2019).
- David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**, 559–563 (2014).
- Johnson, A. J. *et al.* Daily sampling reveals personalized diet-microbiome associations in humans. *Cell Host Microbe* **25**, 789.e5–802.e5 (2019).
- Lang, J. M., Eisen, J. A. & Zivkovic, A. M. The microbes we eat: Abundance and taxonomy of microbes consumed in a day's worth of meals for three diet types. *PeerJ* **2**, e659 (2014).
- Berg, G., Erlacher, A. & Grube, M. The edible plant microbiome: Importance and health issues. In *Principles of Plant-Microbe Interactions: Microbes for Sustainable Agriculture* (ed. Lugtenberg, B.) 419–426 (Springer, 2015). https://doi.org/10.1007/978-3-319-08575-3_44.
- Wassermann, B., Rybakova, D., Müller, C. & Berg, G. Harnessing the microbiomes of Brassica vegetables for health issues. *Sci. Rep.* **7**, 1–12 (2017).
- Hardoim, P. R. *et al.* The hidden world within plants: Ecological and evolutionary considerations for defining functioning of microbial endophytes. *Microbiol. Mol. Biol. Rev.* **79**, 293–320 (2015).
- Scherling, C., Ulrich, K., Ewald, D. & Weckwerth, W. A metabolic signature of the beneficial interaction of the endophyte *paenibacillus* sp. isolate and in vitro-grown poplar plants revealed by metabolomics. *Mol. Plant Microbe Interact.* **22**, 1032–1037 (2009).
- Chanclud, E. & Lacombe, B. Plant hormones: Key players in gut microbiota and human diseases?. *Trends Plant Sci.* **22**, 754–758 (2017).
- Rastogi, G. *et al.* Leaf microbiota in an agroecosystem: Spatiotemporal variation in bacterial community composition on field-grown lettuce. *ISME J.* **6**, 1812–1822 (2012).
- Lopez-Velasco, G., Welbaum, G. E., Boyer, R. R., Mane, S. P. & Ponder, M. A. Changes in spinach phylloepiphytic bacteria communities following minimal processing and refrigerated storage described using pyrosequencing of 16S rRNA amplicons. *J. Appl. Microbiol.* **110**, 1203–1214 (2011).
- Jackson, C. R., Randolph, K. C., Osborn, S. L. & Tyler, H. L. Culture dependent and independent analysis of bacterial communities associated with commercial salad leaf vegetables. *BMC Microbiol.* **13**, 274 (2013).
- Leff, J. W. & Fierer, N. Bacterial communities associated with the surfaces of fresh fruits and vegetables. *PLoS ONE* **8**, e59310 (2013).
- López, S. M. Y. *et al.* Microbial endophytes that live within the seeds of two tomato hybrids cultivated in Argentina. *Agronomy* **8**, 136 (2018).
- Romero, F. M., Marina, M. & Pieckenstein, F. L. The communities of tomato (*Solanum lycopersicum* L.) leaf endophytic bacteria, analyzed by 16S-ribosomal RNA gene pyrosequencing. *FEMS Microbiol. Lett.* **351**, 187–194 (2014).

16. Köberl, M., Dita, M., Martinuz, A., Staver, C. & Berg, G. Agroforestry leads to shifts within the gammaproteobacterial microbiome of banana plants cultivated in Central America. *Front. Microbiol.* **6**, 91 (2015).
17. Wassermann, B., Müller, H. & Berg, G. An apple a day: Which bacteria do we eat with organic and conventional apples?. *Front. Microbiol.* **10**, 1629 (2019).
18. Kõiv, V. *et al.* Endophytic bacterial communities in peels and pulp of five root vegetables. *PLoS ONE* **14**, e0210542 (2019).
19. Randazzo, C. L., Restuccia, C., Romano, A. D. & Caggia, C. *Lactobacillus casei*, dominant species in naturally fermented Sicilian green olives. *Int. J. Food Microbiol.* **90**, 9–14 (2004).
20. Bautista-Gallego, J. *et al.* Screening of lactic acid bacteria isolated from fermented table olives with probiotic potential. *Food Res. Int.* **50**, 135–142 (2013).
21. Pasoli, E. *et al.* Large-scale genome-wide analysis links lactic acid bacteria from food with the gut microbiome. *Nat. Commun.* **11**, 2610 (2020).
22. Cernava, T. *et al.* Enterobacteriaceae dominate the core microbiome and contribute to the resistome of arugula (*Eruca sativa* Mill.). *Microbiome* **7**, 13 (2019).
23. Rodriguez, L. M. R., Gunturu, S., Tiedje, J. M., Cole, J. R. & Konstantinidis, K. T. Nonpareil 3: Fast estimation of metagenomic coverage and sequence diversity. *mSystems* **3**, e00039 (2018).
24. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
25. Bowers, R. M. *et al.* Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
26. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
27. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: An online tool for the display and annotation of phylogenetic and other trees. *Nucl. Acids Res.* **44**, W242–W245 (2016).
28. de Castro, A. *et al.* Microbiota and metabolite profiling of spoiled Spanish-style green table olives. *Metabolites* **8**, 73 (2018).
29. Remus, D. M. *et al.* Impact of *Lactobacillus plantarum* sortase on target protein sorting, gastrointestinal persistence, and host immune response modulation. *J. Bacteriol.* **195**, 502–509 (2013).
30. Khatri, I., Sharma, G. & Subramanian, S. Composite genome sequence of *Bacillus clausii*, a probiotic commercially available as Entergermina[®], and insights into its probiotic properties. *BMC Microbiol.* **19**, 1–15 (2019).
31. Koryszewska-Bagińska, A., Gawor, J., Nowak, A., Grynberg, M. & Aleksandrak-Piekarczyk, T. Comparative genomics and functional analysis of a highly adhesive dairy *Lactobacillus paracasei* subsp. *paracasei* IBB3423 strain. *Appl. Microbiol. Biotechnol.* **103**, 7617–7634 (2019).
32. Francl, A. L., Thongaram, T. & Miller, M. J. The PTS transporters of *Lactobacillus gasseri* ATCC 33323. *BMC Microbiol.* **10**, 77 (2010).
33. Denou, E. *et al.* Identification of genes associated with the long-gut-persistence phenotype of the probiotic *Lactobacillus johnsonii* strain NCC533 using a combination of genomics and transcriptome analysis. *J. Bacteriol.* **190**, 3161–3168 (2008).
34. Zhang, B. *et al.* Comparative genome-based identification of a cell wall-anchored protein from *Lactobacillus plantarum* increases adhesion of *Lactococcus lactis* to human epithelial cells. *Sci. Rep.* **5**, 1–12 (2015).
35. Bergonzelli, G. E. *et al.* GroEL of *Lactobacillus johnsonii* La1 (NCC 533) is cell surface associated: Potential role in interactions with the host and the gastric pathogen *Helicobacter pylori*. *Infect. Immun.* **74**, 425–434 (2006).
36. Wu, C. H. *et al.* The Universal Protein Resource (UniProt): An expanding universe of protein information. *Nucl. Acids Res.* **34**, D187–D191 (2006).
37. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B. & Wu, C. H. UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
38. Huang, L. *et al.* dbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucl. Acids Res.* **46**, D516–D521 (2018).
39. Kaoutari, A. E., Armougom, F., Gordon, J. I., Raoult, D. & Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **11**, 497–504 (2013).
40. Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M. & Henrissat, B. Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of α -amylase-related proteins. *Protein Eng. Des. Sel.* **19**, 555–562 (2006).
41. Speciale, G., Jin, Y., Davies, G. J., Williams, S. J. & Goddard-Borger, E. D. YihQ is a sulfoquinovosidase that cleaves sulfoquinovosyl diacylglyceride sulfolipids. *Nat. Chem. Biol.* **12**, 215–217 (2016).
42. Burrichter, A. *et al.* Anaerobic degradation of the plant sugar sulfoquinovose concomitant with H₂S production: *Escherichia coli* K-12 and *Desulfovibrio* sp. strain DF1 as co-culture model. *Front. Microbiol.* **9**, 2792 (2018).
43. Khanmohammadi, M., Moeini, M., Bagheri Garamarudi, A., Sotudehnia, A. & Zarrabi, M. Simultaneous determination of sucrose and trehalose in olive leaves by spectrophotometry utilizing partial least squares method. *Acta Physiol. Plant* **31**, 865–869 (2009).
44. Ragaert, P., Devlieghere, F. & Debevere, J. Role of microbiological and physiological spoilage mechanisms during storage of minimally processed vegetables. *Postharvest Biol. Technol.* **44**, 185–194 (2007).
45. Cantarel, B. L., Lombard, V. & Henrissat, B. Complex carbohydrate utilization by the healthy human microbiome. *PLoS ONE* **7**, e28742 (2012).
46. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **7**, 189–200 (2016).
47. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From dietary fiber to host physiology: Short-chain fatty acids as key bacterial metabolites. *Cell* **165**, 1332–1345 (2016).
48. Das, P., Babaei, P. & Nielsen, J. Metagenomic analysis of microbe-mediated vitamin metabolism in the human gut microbiome. *BMC Genom.* **20**, 208 (2019).
49. Fang, H., Kang, J. & Zhang, D. Microbial production of vitamin B12: A review and future perspectives. *Microb. Cell Fact.* **16**, 15 (2017).
50. Issac, T. G., Soundarya, S., Christopher, R. & Chandra, S. R. Vitamin B12 deficiency: An important reversible co-morbidity in neuropsychiatric manifestations. *Indian J. Psychol. Med.* **37**, 26–29 (2015).
51. Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B. & Jackson, R. W. *Pseudomonas* genomes: Diverse and adaptable. *FEMS Microbiol. Rev.* **35**, 652–680 (2011).
52. Garrido-Sanz, D. *et al.* Genomic and genetic diversity within the *Pseudomonas fluorescens* complex. *PLoS ONE* **11**, e0150183 (2016).
53. Garcia, S. L. *et al.* Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. *ISME J.* **12**, 742–755 (2018).
54. Oh, S. *et al.* Metagenomic Insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl. Environ. Microbiol.* **77**, 6000–6011 (2011).
55. Johnston, E. R. *et al.* Metagenomics reveals pervasive bacterial populations and reduced community diversity across the Alaska Tundra ecosystem. *Front. Microbiol.* **7**, 579 (2016).
56. Forster, S. C. *et al.* A human gut bacterial genome and culture collection for improved metagenomic analyses. *Nat. Biotechnol.* **37**, 186–192 (2019).
57. Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499–504 (2019).

58. Uylaser, V. & Yildiz, G. Fatty acid profile and mineral content of commercial table olives from Turkey. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* **41**, 518 (2013).
59. Sabatini, N., Mucciarella, M. R. & Marsilio, V. Volatile compounds in uninoculated and inoculated table olives with *Lactobacillus plantarum* (*Olea europaea* L., cv. Moresca and Kalamata). *LWT Food Sci. Technol.* **41**, 2017–2022 (2008).
60. Almonacid, D. E. *et al.* 16S rRNA gene sequencing and healthy reference ranges for 28 clinically relevant microbial taxa from the human gut microbiome. *PLoS ONE* **12**, e0176555 (2017).
61. Nistal, E. *et al.* Study of duodenal bacterial communities by 16S rRNA gene analysis in adults with active celiac disease vs non-celiac disease controls. *J. Appl. Microbiol.* **120**, 1691–1700 (2016).
62. Galván, A. I. *et al.* Effect of temperature during drying and storage of dried figs on growth, gene expression and Aflatoxin production. *Toxins (Basel)* **13**, 134 (2021).
63. Marchese S, Polo A, Ariano A, Velotto S, Costantini S, Severino L. Aflatoxin B1 and M1: Biological Properties and Their Involvement in Cancer Development. *Toxins (Basel)*. **10**(6), 214. (2018). <https://doi.org/10.3390/toxins10060214>
64. Cortés-Delgado, A. *et al.* Volatile profile of Spanish-style green table olives prepared from different cultivars grown at different locations. *Food Res. Int.* **83**, 131–142 (2016).
65. Lovegrove, A. *et al.* Role of polysaccharides in food, digestion, and health. *Crit. Rev. Food Sci. Nutr.* **57**, 237–253 (2017).
66. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
67. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
68. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
69. Cox, M. P., Peterson, D. A. & Biggs, P. J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinform.* **11**, 485 (2010).
70. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
71. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
72. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
73. Rodriguez, R. L. M. & Konstantinidis, K. T. *The Enveomics Collection: A Toolbox for Specialized Analyses of Microbial Genomes and Metagenomes*. <https://peerj.com/preprints/1900> (2016).
74. Ondov, B. D. *et al.* Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
75. Johnson, L. S., Eddy, S. R. & Portugal, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinform.* **11**, 431 (2010).
76. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797 (2004).
77. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
78. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
79. Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **12**, 402 (2011).
80. Ashburner, M. *et al.* Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
81. Kim, J., Kim, M. S., Koh, A. Y., Xie, Y. & Zhan, X. FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinform.* **17**, 1–8 (2016).
82. Kanehisa, M. Enzyme annotation and metabolic reconstruction using KEGG. *Methods Mol. Biol.* **1611**, 135–145 (2017).
83. Zhang, H. *et al.* dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucl. Acids Res.* **46**, W95–W101 (2018).

Acknowledgements

We thank Devin Pineda, Alfredo Hans Loch, Joshua Copeland, Alan Yu, and Sanam Sajjadi for help processing samples and to Courtney McCormick for her assistance in lab management at the Illumina Accelerator in San Francisco, CA.

Author contributions

G.T. and T.J.M. designed the experiment. J.D.W. performed the experiments. G.T. and T.J.M. provided guidance and assisted in the experiments. M.J.S.G. wrote the manuscript with input from G.T., E.S., C.T., and T.I. M.J.S.-G. and J.-N.K. performed the bioinformatics and data analyses. All authors contributed to manuscript revision and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03334-4>.

Correspondence and requests for materials should be addressed to G.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021