













RESEARCH

Open Access



Exposing new taxonomic variation with inflammation — a murine model-specific genome database for gut microbiome researchers

Ikaia Lelewi^{1,2} , Josué Rodríguez-Ramos^{2,3} , Michael Shaffer² , Anice Sabag-Daigle⁴ , Katherine Kokkinias^{2,5} , Rory M. Flynn² , Rebecca A. Daly² , Linnea F. M. Kop^{6,7} , Lindsey M. Solden² , Brian M. M. Ahmer⁴ , Mikayla A. Borton²  and Kelly C. Wrighton^{1,2,3,5*} 

Abstract

Background The murine CBA/J mouse model widely supports immunology and enteric pathogen research. This model has illuminated *Salmonella* interactions with the gut microbiome since pathogen proliferation does not require disruptive pretreatment of the native microbiota, nor does it become systemic, thereby representing an analog to gastroenteritis disease progression in humans. Despite the value to broad research communities, microbiota in CBA/J mice are not represented in current murine microbiome genome catalogs.

Results Here we present the first microbial and viral genomic catalog of the CBA/J murine gut microbiome. Using fecal microbial communities from untreated and *Salmonella*-infected, highly inflamed mice, we performed genomic reconstruction to determine the impacts on gut microbiome membership and functional potential. From high depth whole community sequencing (~42.4 Gbps/sample), we reconstructed 2281 bacterial and 4516 viral draft genomes. *Salmonella* challenge significantly altered gut membership in CBA/J mice, revealing 30 genera and 98 species that were conditionally rare and unsampled in non-inflamed mice. Additionally, inflamed communities were depleted in microbial genes that modulate host anti-inflammatory pathways and enriched in genes for respiratory energy generation. Our findings suggest decreases in butyrate concentrations during *Salmonella* infection corresponded to reductions in the relative abundance in members of the *Alistipes*. Strain-level comparison of CBA/J microbial genomes to prominent murine gut microbiome databases identified newly sampled lineages in this resource, while comparisons to human gut microbiomes extended the host relevance of dominant CBA/J inflammation-resistant strains.

Conclusions This CBA/J microbiome database provides the first genomic sampling of relevant, uncultivated microorganisms within the gut from this widely used laboratory model. Using this resource, we curated a functional, strain-resolved view on how *Salmonella* remodels intact murine gut communities, advancing pathobiome understanding beyond inferences from prior amplicon-based approaches. *Salmonella*-induced inflammation suppressed *Alistipes* and other dominant members, while rarer commensals like *Lactobacillus* and *Enterococcus* endure. The rare and novel species sampled across this inflammation gradient advance the utility of this microbiome resource to benefit the broad

*Correspondence:

Kelly C. Wrighton

wrighton@colostate.edu

Full list of author information is available at the end of the article



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023. **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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

research needs of the CBA/J scientific community, and those using murine models for understanding the impact of inflammation on the gut microbiome more generally.

Keywords CBA/J mouse, *Salmonella*, Metagenome, Pathogen, Pathobiome

Introduction

Non-typhoidal *Salmonella* (NTS) is one of the leading causes of gastroenteritis and associated mortality worldwide, resulting in nearly 1 million cases and more than 50,000 deaths in 2017 [1–3]. *Salmonella enterica* serovar Typhimurium (referred to hereon as *Salmonella*) is a NTS model enteric pathogen that exploits inflammation to increase its pathogenicity and fitness relative to other bacteria [4–6]. Prior research in murine models showed gut microbiota are remodeled during *Salmonella*-induced inflammation because of innate immunity, diminished resources, and altered chemical environment [6, 7]. Similar inflammation-associated changes in commensal gut ecology are observed in patients with Crohn's disease, irritable bowel disease, and metabolic syndrome [4, 8]. The *Salmonella* disease model may represent a human-relevant system for investigating host–pathogen–microbiota interactions in the inflamed GI tract germane to microbiome changes during other human chronic inflammatory conditions.

Earlier work showed that *Salmonella* induced inflammation created reactive oxygen and nitrogen species in the luminal environment, increasing the availability of oxygen and potentiating formation of tetrathionate and nitrate [9]. Increased concentrations of these terminal electron acceptors allowed *Salmonella* to respire and out-compete obligate fermentative commensals like members of the *Clostridia* [6, 10]. Additionally, in the remodeled gut ecosystem respiring *Salmonella* benefited from unique access to non-competitive carbon sources like propionate and ethanolamine [11, 12]. Prior competition studies often focused on decreases in the *Clostridia* [7, 13], ignoring implications for the other members of the community, and those that can withstand *Salmonella* infection. Here we provide a holistic, strain resolved view of the changes in microbiome membership and function during *Salmonella* infection.

Previous investigations of the impacts of *Salmonella* on the gut microbiome relied on the use of pre-treated reduced diversity communities [14–16]. Mice in past studies (e.g., BALB/c or C57BL/6) required antibiotic conditioning prior to pathogen introduction, preventing investigation of *Salmonella* physiology in response to an intact microbial community [14, 15, 17]. As an alternative, CBA/J mice are gaining appreciation as a model to interrogate *Salmonella* pathogenicity as

they support longer non-systemic *Salmonella* infection without prior antibiotic perturbation, similar to enteric disease manifestation in humans [17–20]. While in vitro and in vivo studies with reduced microbiota consortia provide an important theoretical framework, additional research is needed in the presence of native microbial communities to understand how specific inflammation-induced changes to microbiota membership and function impact *Salmonella* physiology and pathogenicity.

Currently, the capability to study intact resident gut microbiota during *Salmonella*-induced inflammation is hindered because CBA/J mice are missing from available murine microbiome genomic catalogs [21–23]. Furthermore, existing murine gut genomic databases exclude inflamed mice, and mice colonized by enteric pathogens (e.g., like *Salmonella*, *Klebsiella*, and *Citrobacter*), thus limiting the extension of these existing microbiome resources to pathobiome models. Despite being recognized as important contributors to human health [24], these existing curated murine gut microbial genome catalogs also lack virome sampling [25, 26]. Accurate interrogation of complete microbial community functions during *Salmonella* infection requires comprehensive model-specific knowledge of gene content and community membership both in healthy and inflamed guts.

To evaluate the functional potential of microbial communities during *Salmonella*-induced inflammation, and to explore if the CBA/J inflammation model harbors unique and previously understudied microorganisms, we constructed a metagenome assembled genome catalog from healthy and *Salmonella*-infected CBA/J mice. We employed high-depth metagenomic sequencing and used several assembly strategies to increase the de novo reconstruction of viral and bacterial genomes. These efforts resulted in a comprehensive culture-independent genome resource that (i) revealed novel taxonomy unique to CBA/J and inflamed mice, (ii) included taxa with relevance to human systems and with anti-inflammatory effector potential, and (iii) showed how enteric inflammation remodels the functional profile of the gut by selecting for bacteria that encode mechanisms to withstand oxidative stress. Ultimately our findings reframe existing responses of the microbiota during *Salmonella* infection and provide new insights into specific bacteria

that can withstand inflammation to maintain critical gut functions, perhaps revealing promising future probiotic targets.

Results

Pathogen perturbation extends the genomic sampling of the CBA/J microbiome

To examine the microbial community response to *Salmonella* colonization, 14 CBA/J mice were infected with 10^9 CFU *Salmonella enterica* serovar Typhimurium strain 14,028 with results compared to 16 uninoculated control mice sampled at the same time points ($n=30$ mice, Fig. 1A). Feces were collected prior to infection (day -1) and in late stages of infection (days 10 and 11), with 16S rRNA microbial community analyses performed at early and late time points ($n=60$) and lipocalin-2, an indicator of enteric inflammation, measured on late time points from select mice from each treatment group ($n=12$). The 60 fecal samples yielded 2,047,287 paired end 16S rRNA reads, which identified 23,022 unique amplicon sequencing variants (ASVs) from both inflamed and control treatments (Additional file 1: Data S1). To confirm infection, we established that inoculated mice had *Salmonella* relative abundance greater than 25% on day 11 and had significantly higher lipocalin-2 concentrations than control mice on day 10. From these mice, we selected feces at day 11 from 3 *Salmonella* infected mice and 3 uninfected mice for deep metagenomic sequencing.

The 16S rRNA gene findings confirmed *Salmonella* infection resulted in statistically discernable microbial communities by day 11 following infection (Fig. 1B). A *Salmonella enterica* relative abundance increase ($\geq 25\%$) was concomitant with increased inflammation evidenced by a 2.5 log-fold rise in lipocalin-2 compared to levels in uninfected mice (Fig. 1B, Additional file 1: Data S1). The microbial community of inflamed mice statistically differed from control uninfected mice at the same time point, and pre-pathogen-treated mice from both treatments (Fig. 1B,C, Fig. S1A). Pre-infection (day -1) mice that later became *Salmonella* inoculated, and uninoculated mice had fecal microbial communities that were not discernable from each other, indicating observed community differences by

day 11 were due to *Salmonella* infection (Fig. S1B). As others have reported [13, 27], *Salmonella*-induced inflammation significantly changed gut microbial diversity, it reduced ASV richness by more than half (76.2%) and decreased Shannon's diversity by 2.6-fold. These findings demonstrate that pathogen perturbation changes microbial membership and structure, offering a strategy for differential genomic sampling of the CBA/J gut microbiome.

To extend the relevance of our findings, we compared uninfected amplicon sequenced communities to communities from CBA mice in two other studies and showed a strong overlap in taxonomy between the three (Fig. S2) [27, 28]. Dominate taxa in the CBAJ-DB were shared across studies despite myriad differences including mouse breeding facility, chow type, experimental facility, and experimental methods (Fig. S2). These findings highlight the relevance of this first CBA mouse genomic resource to microbiome research in this model more broadly by expounding notable community consistency between pathogen-free CBA mice despite other confounding factors.

These 16S rRNA analyses revealed inflamed communities were enriched in members of the *Gammaproteobacteria* and *Bacilli*, while gut communities of uninoculated mice included higher relative abundances of *Bacteroidia*, *Mollicutes*, and *Clostridia* (Fig. 1C). From the mice also sampled for metagenomic analysis ($n=6$), *Alistipes* sp. was the most dominant commensal and the most reduced during inflammation (from 37.2%), but notably still detectable (7.68%). *Salmonella enterica* Typhimurium dominated the *Gammaproteobacteria* in inflamed communities, contributing up to a mean relative abundance of 94% in infected samples. Certain low abundant members of the CBA/J microbiome significantly increased in relative abundance following pathogen treatment, including some members of *Lactobacillus*, *Enterococcus*, and *Lachnospiraceae* (Fig. 1D). Control mouse communities are consistent with findings from prior work showing uninfected CBA/J mouse gut community membership dominated by *Bacteroidetes* and *Firmicutes*, especially *Clostridia* of various *Lachnospiraceae* and *Ruminococcaceae* families

(See figure on next page.)

Fig. 1 Amplicon sequencing of the CBA/J gut reveals shifts in microbiome composition with inflammation. **A** Experimental design shows the number of mice in healthy (green) and infected (brown, infected with 10^9 CFU *Salmonella*) treatments with fecal sampling times and corresponding analysis indicated by black circles. **B** Boxplots show lipocalin-2 (Lc2) levels of mice in each treatment during peak infection (top), with *Salmonella* relative abundance indicated by circle size. Non-metric multidimensional scaling of Amplicon Sequence Variant (ASV) Bray-Curtis distances showing significantly distinct communities between treatments at time of sampling with points scaled to *Salmonella* relative abundance (bottom). Asterisks indicate mice used to create the CBAJ-DB. **C** ASV Class distribution is depicted by stacked bar charts of healthy and infected communities, with each bar representing a single mouse at the day 11 timepoint. **D** Rank abundance curve of mean ASV relative abundance by treatment in mice sampled to create the CBAJ-DB. Bars represent a single ASV and are colored by treatment, with ASVs ranked separately for each treatment to show the changes in rank and abundance with inflammation. Bars are labeled with taxonomic identity in text if greater than 3% mean relative abundance in either treatment, with text color denoting treatment. Black text indicates both treatments have the same ASV taxonomy within that rank

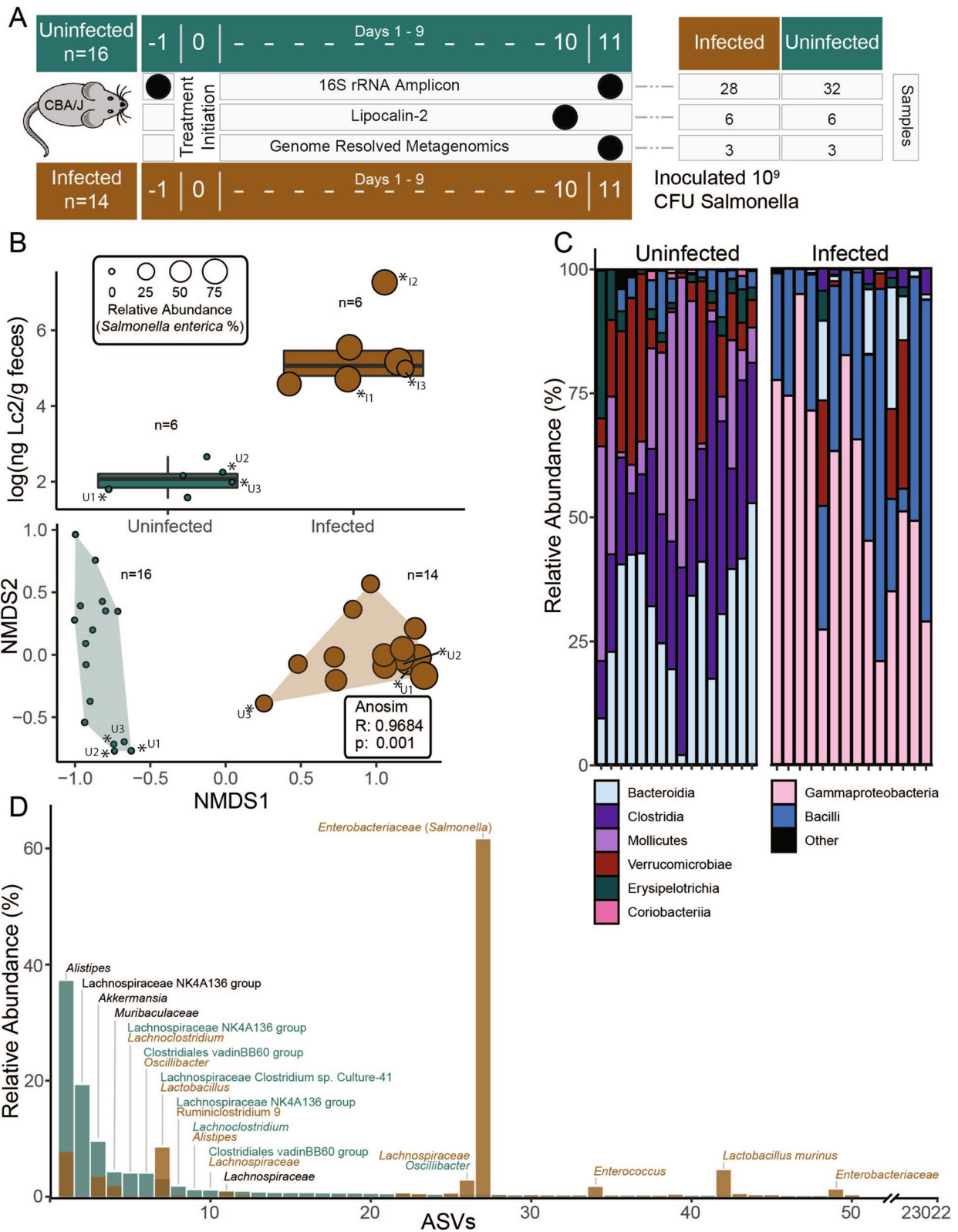


Fig. 1 (See legend on previous page.)

[27, 28]. These 16S rRNA gene analyses revealed abundant members in both healthy and inflamed CBA/J gut microbiomes that represented microorganismal genome “targets” for our database.

Microbial genomic reconstruction from CBA/J mice recovers relevant members sampled in amplicon surveys

To thoroughly catalog the CBA/J gut microbiota, high sequencing depth was required to sequence through *Salmonella* dominance (25.8–94.2% by amplicon analyses) and recover some of the first genomes from rare, but persistent co-occurring members of the pathogen inflamed gut. We obtained 254.2 Gbps of metagenomic sequencing data (Additional file 2: Data S2) from 6 representative mice (inflamed $n=3$, uninfected $n=3$,

Fig. 1A), sevenfold more sequencing/sample than is commonly done in murine catalogs (Fig. S3A) [21, 22, 29]. Additionally, we used iterative, targeted assembly approaches (single, co-assembly, subtractive assembly) as well as two different assemblers to attempt to enhance genome quality and recovery, especially from less dominant members (Fig. 2A, Fig. S3B). Subtractive and co-assembly methods derived 259 additional metagenomically assembled genomes (MAGs) beyond those from single sample assemblies, with the distribution of MAGs from each assembler reported (Fig. 2A). In total, we recovered 2281 MAGs. Quality assessment revealed 504 MAGs to be either medium or high-quality (MQHQ) with sufficiently low contamination to be included in further analyses. Of the genome quality

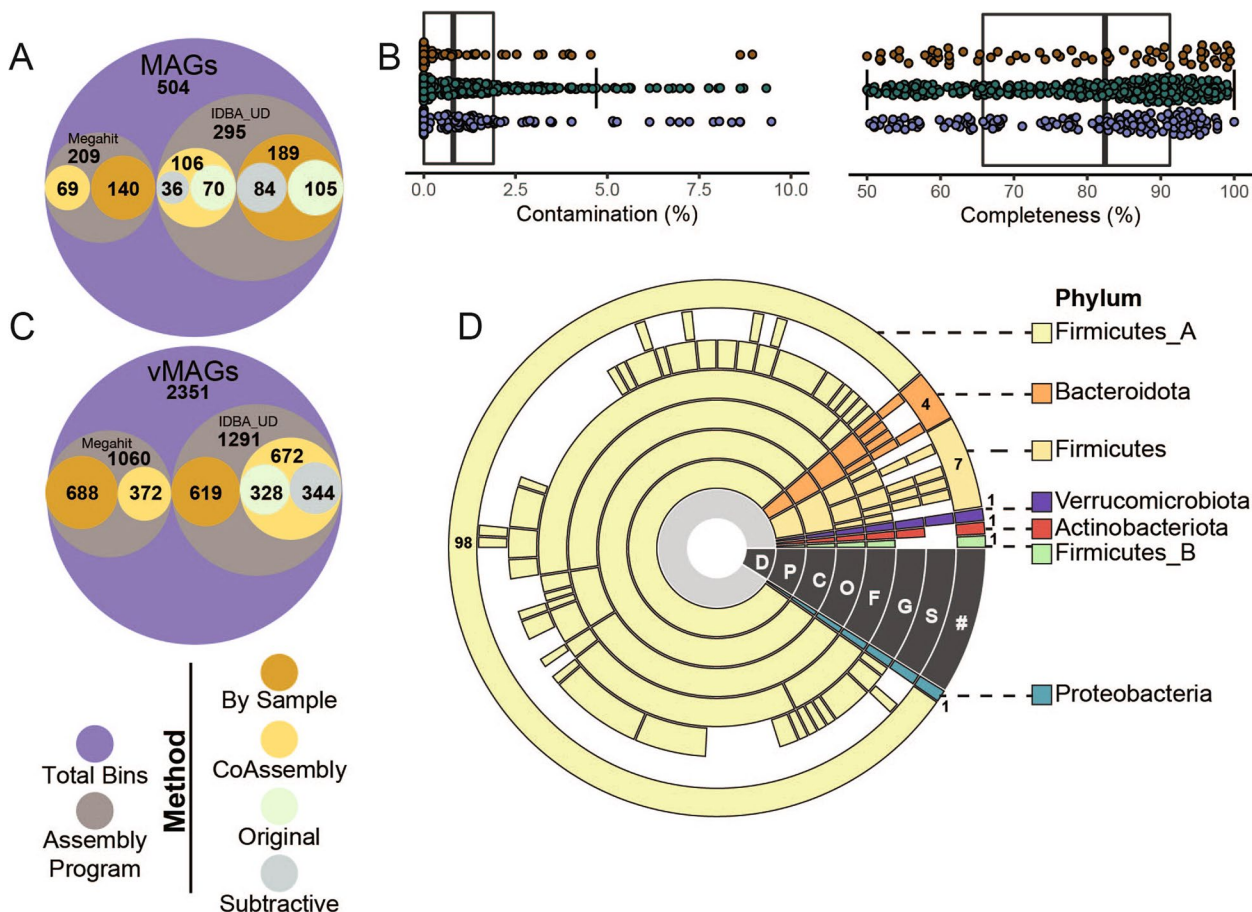


Fig. 2 CBA/J mouse database (CBAJ-DB) genomic methods and composition. **A** Circle plot shows the number of medium and high quality (MQHQ) metagenome assembled genomes (MAGs) reconstructed from CBA/J mouse gut metagenomes and corresponding assembly method and assembly software. **B** Completeness and contamination of all MQHQ MAGs colored by sample treatment origin (brown = infected; green = uninfected; purple = co-assembly) are shown by box plots, with bold horizontal lines indicating median across all MQHQ MAGs. **C** Circle plot shows the number of viral metagenome assembled genomes (vMAGs) reconstructed from CBA/J mouse gut metagenomes and corresponding assembly method and assembly software. **D** Dereplicated MQHQ (dMQHQ) MAG phylum distribution is shown by sequential colored rings listed from least specific (Domain, D) to most specific (Species, S) moving outwards from the plot center. Gaps at each taxonomic level represent MAGs that are previously undescribed. The outer ring is labeled with the number of MAGs from the dMQHQ CBAJ-DB database within each Phylum

tools used, CheckM [30] provided the most conservative MAG set ($n=504$ MQHQ MAGs), compared to CheckM2 [31] ($n=531$ MQHQ MAGs) and GUNC [32] ($n=790$, MQHQ MAGs) as assessed by contamination and completeness. There existed no significant differences between the quality metrics of MAGs from either treatment (uninfected, infected) or between assembly methods (Fig. S4A). These quality genomes contained 156,921 uniquely called predicted genes [33, 34] (Fig. 2, Additional file 3: Data S3).

Dereplication of our metagenome assembled genomes (99% identity) resulted in 113 medium and high-quality MAGs (dMQHQ) from both treatments. These MAGs were assigned to 7 Phyla – *Actinobacteriota* (1), *Bacteroidota* (4) *Firmicutes* (7), *Firmicutes_A* (98), *Firmicutes_B* (1), *Proteobacteria* (1), and *Verrucomicrobiota* (1) (Fig. 2D, Additional file 2: Data S2). Nearly a third (30 of the 113) of the dereplicated MAGs were assigned to 30 genera and 98 to species that were only recognized by alphanumeric numbering in GTDB-Tk, hinting that novelty sampled here may be undescribed not only in murine but larger MAG collections. Reflecting the richness of these samples, the majority of MAGs originated from uninfected mice (59%) and their co-assemblies (35%) while 13% came from inflamed mice. Specifically, *Enterococcus_D gallinarum*, *Erysipelatoclostridium cocleatum*, *Kineothrix sp000403275*, and *Lactobacillus_B animalis* MAGs were uniquely recovered from inflamed mice, consistent with their 16S rRNA membership (Fig. 1D).

This finding indicates how perturbation can aid in the sampling of genomes from conditionally rare members; however, we also acknowledge the risk of facility specificity and other factors influencing these microbiota.

Expanding this resource beyond solely bacterial genomes, we also reconstructed viral genomes from our CBA/J assemblies, recovering 4516 viral metagenome assembled genomes (vMAGs). Of these, 2351 vMAGs were ≥ 10 kb which were then dereplicated into 609 viral genomes (Figs. 2C and 4D, Additional file 4: Data S4).

We first sought to verify if this microbial dereplicated MAG set represented the key members identified in our amplicon sequenced CBA/J communities from both inflamed ($n=14$) and uninfected ($n=16$) individuals (Fig. 1B). The relative abundance of dMQHQ MAGs

closely mirrored the full community 16S rRNA amplicon at the class level from both uninfected day 11 ($\rho=0.68$) and inflamed day 11 ($\rho=0.86$) mice, indicating the dMQHQ database is representative of CBA/J untreated and inflamed communities (Figs. 1C and 3A,). More specifically, a linear discriminant analysis of MAG relative abundance indicated similar dynamics between our genome and amplicon data sets. For example, *Salmonella* and *Enterococcus_D* were the most significant genomes in determining infected communities, while genomes from *Alistipes*, *Duncaniella*, and *Lachnospiraceae* COE1 were most significant in determining uninfected communities (Fig. 3B). Additional to these genera, relative abundance of other key taxa is consistent with amplicon sequencing, including *Akkermansia*, and *Muribaculaceae* prominence in uninfected mice and persistence in infected mice. *Lactobacillus* genome and ASV relative abundance also similarly increased during infection (Fig. 3C).

To link these reconstructed genomes more precisely to the amplicon data, we identified 96 MAGs that contained a partial to full 16S rRNA gene sequence. A relatively low proportion of MAGs containing 16S rRNA sequences may be attributed to the difficulty de novo assembly algorithms have with the conserved regions and tetranucleotide variation associated with this gene [35]. A pairwise comparison of MAG-derived 16S rRNA sequences and the V4 region sequences from our ASVs identified 33 unique genomes containing sequences matching ASVs in our 16S rRNA dataset. Many MAGs with 16S rRNA matches were among the most enriched taxa including *Lactobacillus johnsonii*, *Alistipes* sp002428825, and *Clostridia* in order 4C28d-15 (Fig. S4B, Additional file 1: Data S1). Together these findings indicate significant membership congruence in our MAG database and our amplicon data, demonstrating that inferences made with the CBAJ-DB have relevance to the more broadly sampled amplicon sequenced gut communities from inflamed and uninfected mice.

This CBA/J microbial genomic resource includes mouse and human relevant lineages

One goal of developing a genome-resolved CBA-specific microbiome resource is to advance future multi-omics studies in this mouse model. Metaproteomes and metatranscriptomes are mapped using high stringency to

(See figure on next page.)

Fig. 3 The CBA/J database (CBAJ-DB) genomes are representative key members in amplicon microbiome data. **A** Class distribution of dereplicated MQHQ (dMQHQ) metagenome assembled genome (MAG) relative abundance across individual metagenomes (U# = uninfected, I# = infected) labeled by treatment. **B** Linear discriminant analysis Effect Size (LEfSe) scores of most important genera in either treatment in metagenomes, including genera LEfSe linear discriminant analysis (LDA) scores for amplicon sequencing variants (ASVs) with matching taxonomy. **C** Rank abundance of dMQHQ MAGs ($n=113$) showing mean relative abundance (RA) in uninfected (green) and infected (brown) treatments (TRT). Circles below bars highlight LDA significant species (black) and genera (gray) (top row) and treatment origin of each MAG (bottom row). MAGs are labeled with the most resolved GTDB-Tk taxonomy

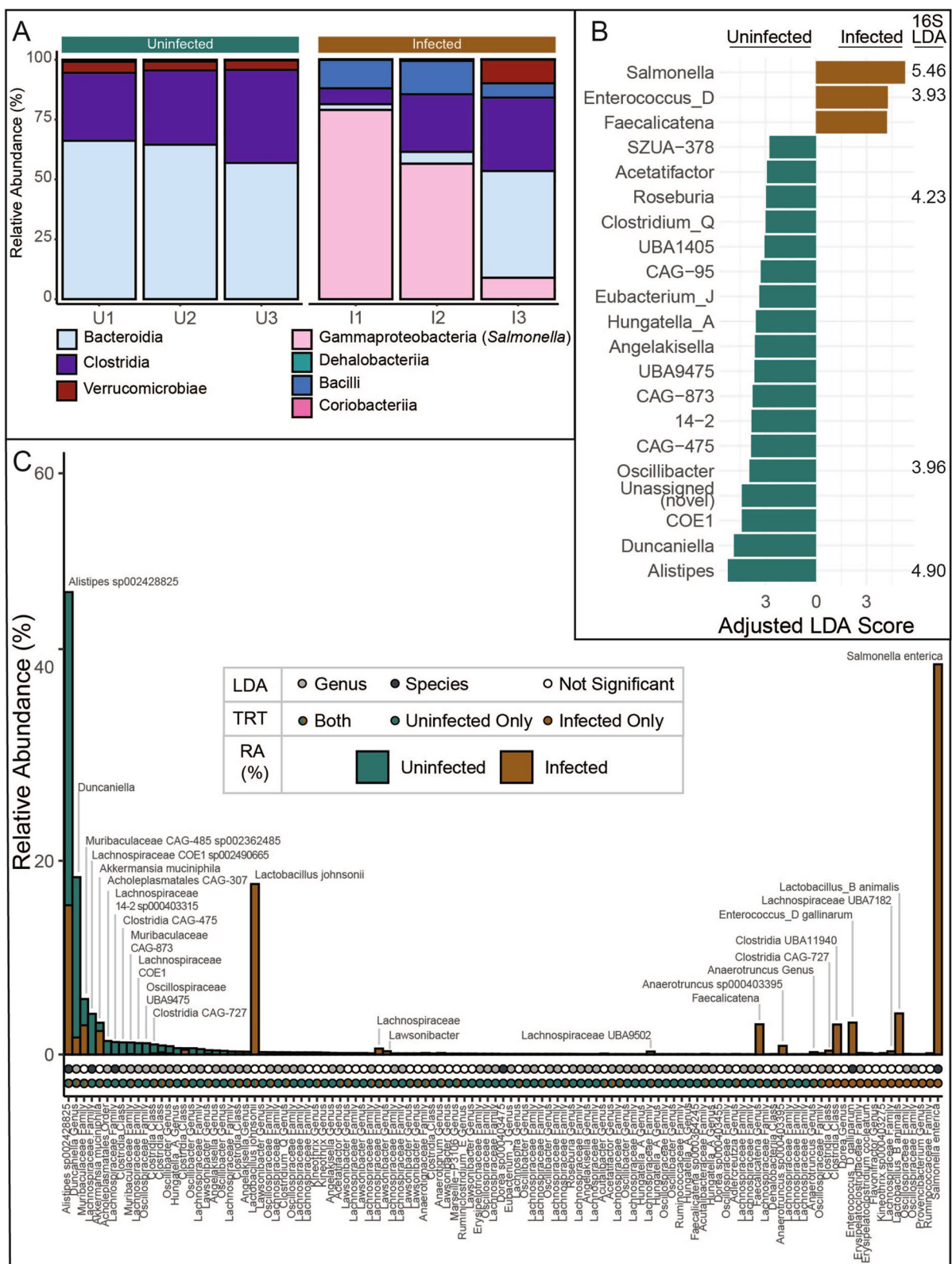


Fig. 3 (See legend on previous page.)

illuminate changes in gene expression, often at a genome-resolved level, typically the strain level (>99% average nucleotide identity) [36–38]. To assess the unique members captured in this resource, we compared strain level identity of our sampled MAGs to similar quality MAGs from two prevalent mouse gut genome catalogs: (i) Integrated Mouse Gut Metagenomic Catalog (iMGMC) [21] and (ii) The Mouse Gastrointestinal Bacteria Catalogue (MGBC) [22]. Notably, many of these CBA-derived genomes represented unique strains from the classes *Bacilli* ($n=3$), *Bacteroidia* ($n=2$), *Clostridia* ($n=24$), *Coriobacteriia* ($n=1$), and *Dehalobacteriia* ($n=1$) not represented in iMGMC, and MAGs from *Bacilli* ($n=1$), *Dehalobacteriia* ($n=1$), and *Clostridia* ($n=30$) not represented in MGBC (Fig. 4A). Additionally, of the strains that were sampled in our dataset and prior curated catalogs, 33 (30 *Clostridia*, 3 *Bacilli*) received a higher quality score indicating the value of these recovered MAGs to advance knowledge of cultivated and uncultivated genomes in murine models more broadly. At a coarser taxonomic level (e.g. 95% genome sequence identity), we still detected novel taxa in the CBAJ-DB that were absent in these existing genomic collections. For example, clustered at 95% identity, the CBAJ-DB harbored a novel *Faecalicatena* sp. and *Provencibacterium* sp. not found in the iMGMC.

We also examined CBAJ-DB MAGs against genomes derived from human hosts. To analyze shared genera and species, our dMQHQ database was dereplicated with isolate genomes from the Human Microbiome Project (HMP) [39] ($n=813$) and MQHQ MAGs ($n=2560$) from a human cohort (PRJNA725020) [40] (Additional file 5: Data S5) [39]. *Akkermansia muciniphila* (CBAJDB_482) and *Enterococcus_D gallinarum* (CBAJDB_497), two defining members of the commensal and inflamed gut respectively, clustered with species previously recovered from human hosts. Recovery of *Enterococcus_D gallinarum* from the uninfected CBA/J gut demonstrates the

applicability of perturbation techniques to uncover conditionally rare members. As has been reported by others, there was more similarity at higher taxonomic levels (e.g., genus) between our murine and human gut microbial members [41, 42], with 27 MAGs from *Bacilli*, *Bacteroidia*, *Clostridia*, *Corriobacteriia*, *Gammaproteobacteria*, and *Verrucomicrobiae* sharing similarity (Fig. 4A).

We were particularly interested if the microbial members recovered from our pathogen-inflamed CBA/J had relevance to inflammation in humans. To test this, sequencing reads from the Lloyd-Price et al. cohort [43] containing 972 inflamed and 365 healthy gut metagenomes were stringently mapped to the CBAJ-DB MQHQ MAGs (Additional file 6: Data S6) [43]. Consistent with their distribution across our treatments, sequencing reads from healthy and inflamed humans mapped to 11 of our *Akkermansia muciniphila* MAGs, while 3 *Enterococcus_D gallinarum* MAGs derived only in our inflamed treatments recruited sequences from inflamed human subjects (Fig. 4B, C). While it can be challenging to extend specific microorganismal findings from murine to human conditions [22, 41, 42], inferences from critical lineages (e.g., *A. muciniphila* or *E. gallinarum*) in our database may have more direct human relevance as *A. muciniphila* is recognized to promote gut barrier integrity and *E. gallinarum* has multiple documented cases as a pathobiont [44, 45]. Beyond inflammation, we also dereplicated the CBAJ-DB with the Unified Human Gastrointestinal Genome (UHGG) catalog [46]. Species from the *Muribaculaceae* (including genus CAG-485) and *Oscillospiraceae* family (including *Lawsonibacter* spp., *Oscillibacter* spp., genus UBA9475, and an undescribed genus) were well represented in the CBAJ-DB and human fecal samples (Data S5). Together these findings show that the CBAJ-DB recovered new species, but also many that have genomic coherence with members in the human gut.

(See figure on next page.)

Fig. 4 CBA/J database (CBAJ-DB) genomes link to other murine and human studies. **A** Medium and high quality (MQHQ) metagenome assembled genomes (MAGs) that clustered with genomes from either murine databases (light purple columns) or the HMP/Human cohort (yellow columns). CBAJ-DB MAGs that were the highest quality genome in each cluster are marked with an asterisk and are displayed in the stacked bar chart grouped by database. MAGs are grouped by class (first bar annotation) and treatment origin (second bar annotation), with the lowest assigned taxonomy indicated in gray scale (right bar). For each database in black outline, blue cells indicate CBAJ-DB MAGs that clustered, while green cells show no clustering. Accompanying bar chart shows the number of CBAJ-DB MAGs with higher quality scores corresponding genomes in other databases. The MAG indicated with red font in the heatmap was determined to contain possible chimeric by GUNC but not CheckM. **B** CBAJ-DB MAGs that recruited human reads are shown by blue (*Akkermansia muciniphila*) and orange nodes (*Enterococcus_D gallinarum*), with size indicating number of CBAJ-DB MAGs. MAG nodes are linked to databases shown as pie charts, where green (healthy human) and tan (inflamed human) indicate sequencing origin of mapped reads and node size indicates sample number. **C** SRA accession ID's of samples from the Lloyd-Price cohort that mapped to CBAJ-DB MAGs *Akkermansia muciniphila* (blue) or *Enterococcus_D gallinarum* (orange). **D** vContact2 network that shows 609 clustered viral metagenome assembled genomes (vMAG) populations present in CBAJ-DB. The network colors represent the vMAG study origin. Pie chart shows proportion of CBAJ-DB vMAGs that clustered to other studies that are cosmopolitan genera (brown), singletons (gray), novel genera (clusters of > 1 vMAGs from this study only, green), or known taxonomy (yellow)

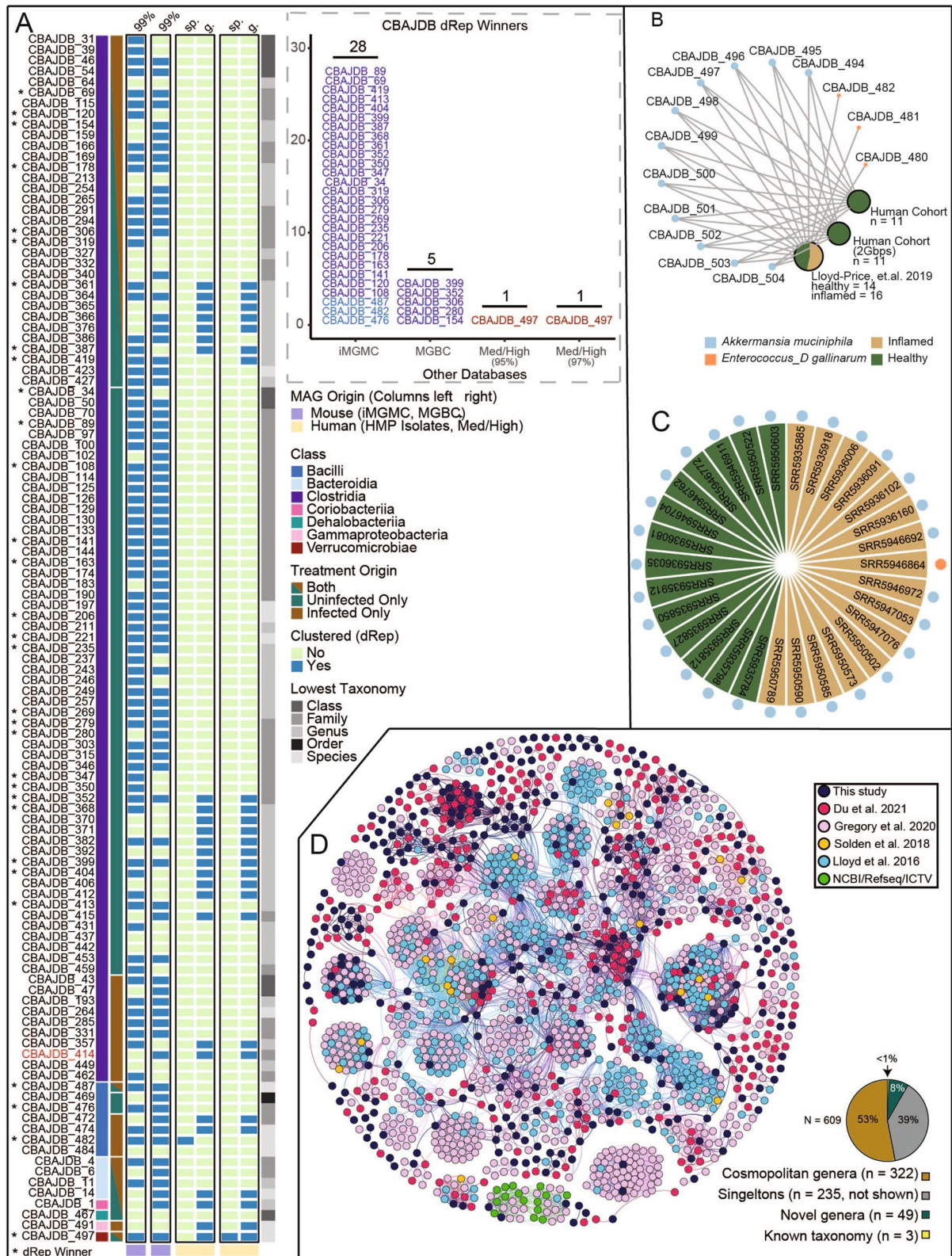


Fig. 4 (See legend on previous page.)

Salmonella infection and inflammation restructures the metabolic potential of the murine gut microbiome

Given this is one of the first genome-resolved analyses of a pathogen-impacted microbiota, and the first for *Salmonella*, it offered a new opportunity to assess functional potential remodeling during infection. Prior reports indicated that pathogen-induced inflammation created oxidative conditions that generated terminal electron acceptors like oxygen, tetrathionate, nitrate, and sulfate [9, 19]. As such, we wanted to evaluate the respiratory capacity of inflamed communities and compare it to uninfamed communities. Interestingly, individual respiration functions were not significantly different between treatments; however, when analysis was expanded to consider the entire respiration category, we found MAGs with respiration capability to be significantly enriched in infected samples (ANOVA $p=1.23e^{-4}$), likely due to the increased n afforded at the category level of analysis (Fig. 5A). In infected communities, *Salmonella* has the highest mean genome relative abundance and encoded gene sets for respiring oxygen (both high and low affinity oxidases), fumarate, tetrathionate, and trimethylamine N-oxide (TMAO) (Fig. 5A). Outside *Salmonella*, no other organisms had the capability for respiring with low affinity oxidases, but we infer *Enterococcus* and *Lactobacillus* have the capability to reduce low levels of oxygen for detoxification (due to the absence of complex I in electron transport chain) while *Akkermansia muciphila* and *Muribaculaceae* likely respire low levels of oxygen using high affinity oxygenase. Similarly, we observed genes for detoxifying reactive oxidative damage (SOD, catalase, thioredoxin reductase) were more enriched in the inflamed community than the uninfected community. Together these findings demonstrated organisms co-existing with *Salmonella* in the inflamed gut encode the metabolic abilities to withstand or leverage the oxidative redox conditions caused by inflammation (Fig. 5C). Markedly, there were members in the uninfamed gut with respiratory metabolic potential that were not maintained in the inflamed gut (*Duncaniella* sp, *Hungatella_A* sp), demonstrating there are other selective forces besides the ability to respire that dictate persistence in response to pathogen colonization (Fig. 5C).

Prior reports by our team and others demonstrated that butyrate, a key gut short-chain fatty acid (SCFA), decreased by 15-fold in the *Salmonella* inflamed gut, most likely due to inflammation induced redox changes with detrimental impacts on members of the class *Clostridia* [13, 27]. Here we sought to better understand the relationship between taxonomy and SCFA production potential. In uninfamed communities, the most prevalent butyrate-producing bacteria were members of the *Alistipes* and *Lachnospiraceae*, members of classes *Bacteroidia* and *Clostridia* respectively. Interestingly, while the most dominant *Clostridia* did decrease in relative abundance with inflammation, replacement *Clostridia* members (*Lachnospiraceae*, *Dorea*, *Faecalicatena*) were enriched which encoded overlapping butyrate production potential. For example, a MAG belonging to the genus *Dorea* within the *Clostridia* was enriched 16-fold and likely most contributed to butyrate production stability, while the dominant *Alistipes* MAG (a member of the *Bacteroidia*) reduced in abundance by a third was not replaced by taxonomically similar members. Together, these data suggest the notion that decreased butyrate concentrations observed in the CBA/J mouse model during *Salmonella* infection [27] may be attributed to *Bacteroidia* reduction and less so to *Clostridia*, a hypothesis needing further validation using gene expression to track butyrate production and consumption activities in the inflamed gut.

Salmonella-induced inflammation alters carbon usage patterns with more favorable redox conditions enabling the use of less energetically favorable substrates like 1,2-propanediol and ethanolamine [47, 48]. While *Salmonella* encodes this metabolic capacity, we were interested if any of the other persisting microorganisms could compete for use of these substrates. *Enterococcus_D* and multiple *Oscillospirales* genomes contain genes from the *eut* gene cluster for ethanolamine utilization and *pdu* genes for 1,2-propanediol utilization. These genera increase in relative abundance with inflammation, particularly *Enterococcus_D*, which is one of the next most abundant members after *Salmonella* (expanding to 2.6% of the inflamed community). Additionally, we showed that the polymer utilization

(See figure on next page.)

Fig. 5 The CBA/J database (CBAJ-DB) highlights differential metabolisms encoded in uninfected and inflamed mice. **A** Normalized relative abundance is shown by a heatmap. Values are mean GeTMM relative abundance of all MAGs with each function center scaled across rows. Functions that are significantly different between treatments as determined by analysis of variance (ANOVA) ($p \leq 0.05$) are indicated by horizontal bars between heatmaps with red highlighting significance at the function (first bar) or functional group level (second bar). Gray bars on each heatmap indicate the number dereplicated medium- and high-quality (dMQHQ) MAGs that comprise at least 0.5% of the community with a given function. **B** Percent change in relative abundance (RA) between treatments the 17 most divergent MAGs. Points are colored by the treatment (uninfected, green; infected, brown) with higher RA. Individual MAGs are uniquely colored by surrounding boxes, acting as color legend for subsequent figure sections. **C** Individual MAG contribution to specific functions is shown, with bar magnitude denoting mean RA in either treatment. **D** *Clostridia* MAGs with mean RA greater than 0.5% in both treatments are shown. **E** *Clostridia* contribution to significant functions (top) and contribution of other taxa (bottom). Plot fields are colored by treatment and bar magnitude indicates mean RA within a treatment

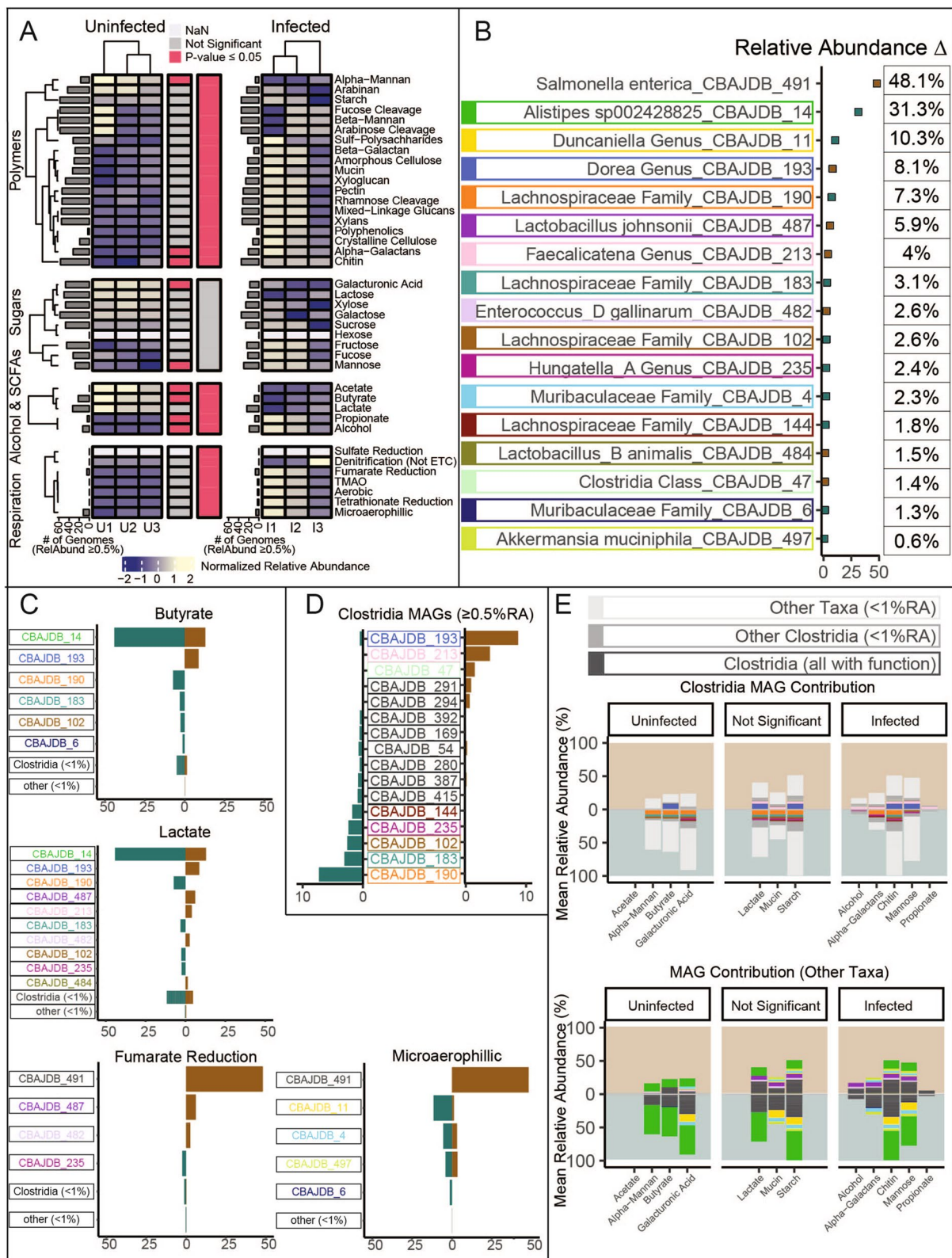


Fig. 5 (See legend on previous page.)

profile was also impacted with inflammation, as infected communities can utilize more alpha-galactan and chitin (Fig. 5). In a similar fashion, the community utilization potential of sugars fructose, fucose, and mannose increased with inflammation. Collectively, these data can inform probiotic approaches for controlling *Salmonella* abundance through competitive exclusion targeting select substrate use patterns using inflammation resistant strains.

Next, we quantified genes commonly reported in humans to impact inflammation and examined if they were depleted in this inflamed mouse model. Consistent with literature reporting healthy individuals have a greater potential for tryptophan degradation [49–51], we observed the potential for tryptophanase-mediated conversion of tryptophan to indole by members of *Bacteroidia*, *Clostridia*, and *Verrucomicrobiae* in both inflamed and uninfected mice. However, the proportion of *Bacteroidia* with this gene was much lower in inflamed guts (Fig. 6B). Tryptophan Indole/AhR pathway representation in infected mice is concurrent with lower proportions of *Verrucomicrobiae* and *Bacteroidia* spp. (Fig. 6). Also, like human microbiomes, we observed microbial genes responsible for cleaving taurine or glycine from primary bile acids and metabolizing secondary bile acid products (*bsh*, *baiN*, *baiA*, and *hdhA*) were significantly lower in relative abundance in mice infected with *Salmonella* (Fig. 6). These data provide promising insights that the functional gene profiles for modulating inflammation are present in the CBA/J model and may suggest its relevance for study of similar inflammation-associated mechanisms in humans.

Viral AMGs contribute to the bacterial community functional potential in CBA/J mice via *Firmicutes*

In the creation of the first murine gut viral database, we sought to compare viral genomic content cataloged here to other mammalian gut systems. Of the 609 dereplicated vMAGs that were recovered from both treatments, less than 1% had taxonomic assignments (Additional file 4: Data S4). These three vMAGs were assigned to the Caudovirales in the families Siphoviridae ($n=2$) and

Myoviridae ($n=1$). To perform biogeographic analyses, we collated phage genomes previously reported from mammalian guts [24, 38, 43, 52, 53] and clustered these with our mouse recovered vMAGs. We found that 322 of the CBA/J-derived vMAGs (53%) had similar representatives in other phage gut metagenome studies, meaning over half of our vMAGs clustered with viruses from at least one additional study (Fig. 4D). This suggests a potentially cosmopolitan phage seedbank that may be conserved across a wide variety of animals, geographies and, in the case of humans, ethnicities and health statuses. Ultimately, viral content in the CBAJ-DB can have relevance to other mouse models and human guts.

To explore if viral communities could potentially influence the structure and function of the CBAJ-DB uninfamed and inflamed microbial communities, we verified that microbial and viral genome-based ordinations were coordinated (Fig. S5). With informatics we conservatively determined that of the 609 vMAGs, 11.5% were putatively linked to 43 MAGs that encompassed 27 unique taxonomies (Fig. S5). All putative hosts corresponded to members of the *Firmicutes*, and included members of the *Lachnospiraceae*, *Ruminococcaceae*, *Oscillospiraceae*, *Anaerotignaceae*, and *Acutalibacteraceae* families. Among the vMAGs that putatively infected hosts, we identified 36 auxiliary metabolic genes (AMGs) with functionalities including regulation of the TCA cycle (citrate synthase), glycolysis (orthophosphate dikinase), phosphate metabolism (PhoH), and oxidative stress response (rubrerythrin). These phage genomes also encoded AMGs for the induction of germination (Peptidase A25), spore formation (M50B), the cleavage of amorphous cellulose (GH2), and low pH resistance (ornithine carbamoyltransferase). Among the putative viral hosts were members within the *Clostridia* class, exhibiting some of the largest MAG relative abundance differences between inflammation states. For example, *Dorea* and *Faecalicatena* enriched in inflamed mice, and *Lachnospiraceae* COE1 enriched in uninfected mice. Together these findings indicate phages may be underappreciated top-down (predation) and bottom-up (resource) controllers of microbiota functionality in the murine gut.

(See figure on next page.)

Fig. 6 Tryptophan and bile acid metabolism in inflamed and uninfected gut microbiomes. **A** Mean relative abundance summed for each function (rows). Functions that are significantly different between treatments as determined by analysis of variance (ANOVA) ($p \leq 0.05$) are indicated by horizontal bar between heatmaps with red highlighting significance at the function level. Gray bars indicate the number medium- and high-quality (MQHQ) metagenome assembled genomes (MAGs) that comprise at least 0.5% of the community with a given function. **B** Relative abundance (point color) and number of MAGs (size) in each class with each gene for tryptophan degradation separated by MAG presence in each treatment, where both indicate MAGs that recruited strictly mapped reads from both treatments. **C** Tryptophan degradation to indole and indole derivatives pathway with pie charts colored by proportion of MAGs in each class (coloring from B) for each treatment. **D** Relative abundance (bars) of MAGs in each treatment encoding bile salt hydrolase (*bsh*), points show sequence similarity to *bsh*, or *hdhA* (K22605), *baiN* (K00076), or *baiA* (K07007) involved in secondary bile acid metabolism. *Dorea* are highlighted as MAGs with more than one gene for metabolizing secondary bile acid products

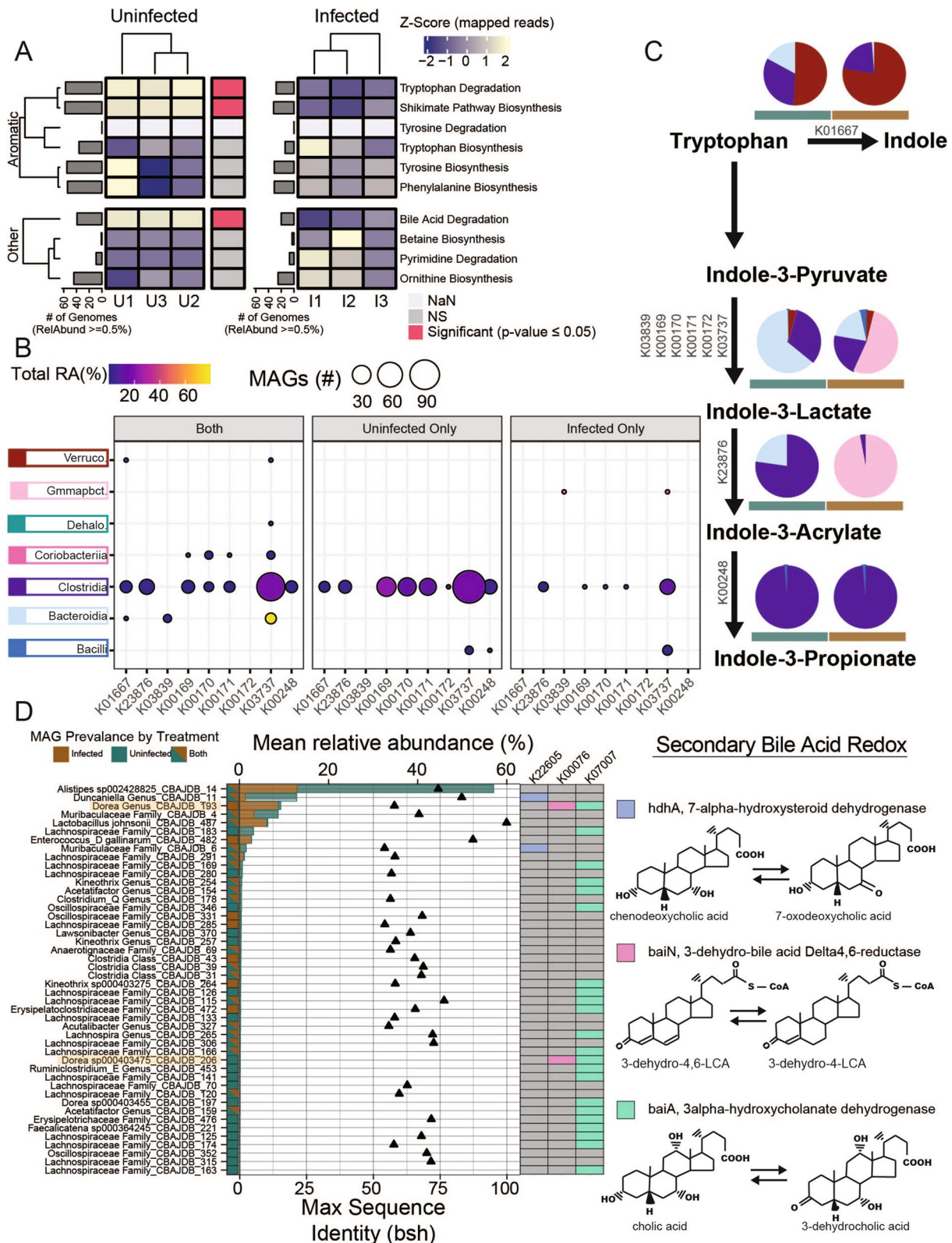


Fig. 6 (See legend on previous page.)

Discussion

Perturbation expanded the microbial and viral genomic cataloging of the CBA/J gut microbiome

Genome resolved catalogs like CBAJ-DB are valuable resources for interrogating metabolic potential of the microbiome, yet these databases are biased by the environment, organism, or disease state they were generated from, and host associated microbiomes can vary drastically between different species, model organisms, and even within an individual [54–56]. Previous murine gut bacterial databases lacked membership from inflamed individuals and CBA/J mice, and none have curated viral content [21, 22, 57], underscoring a clear value of this resource to the community. Our findings highlight the power of using biological perturbation, in this case *Salmonella*-induced inflammation, to genomically sample taxa that are conditionally rare and obscured by their low abundance only to be critical contributors to ecosystem functionality under altered states.

While assemblies and binning are prone to missing key lineages or under-sampling diversity, we used paired 16S rRNA analyses to affirm the representation of critical community members in the inflamed and uninfected gut. Our paired amplicon sequencing indicates the CBAJ-DB contains membership similar to previously reported *Salmonella*-inflamed CBA/J communities and proportional representation of similar bacteria to those found in other mouse breeds [13, 27, 58]. It is our intent to create a resource for other researchers conducting microbiome analyses using CBA/J mice, such that this genome content can be accessed by taxonomic naming or linkages to the 16S rRNA gene. Likewise, this genome library can be used by others for read recruitment of future metagenome and metatranscriptome sequencing, or to substantiate metabolomic insights from the CBA/J microbiome.

The vMAG database also provides interesting context for researchers in other mammalian gut habitats, but especially mouse gut which has been historically under-sampled in this regard. While collections of human gut viruses are available, the mouse virome is understudied [24, 59]. This collection of over 4000 vMAGs contains a significant number of cosmopolitan genera also found in other mammalian guts including humans. The existence of a core mammalian gut virome is an exciting proposition that alludes to an intricacy in gut community function and begs further exploration. The gut microbiome is a complex system involving the interplay of host, microbes, and abiotic elements [60–62]. Beyond functional characterization of gut communities during *Salmonella* infection, the CBAJ-DB offers a bacterial and

viral resource for holistic microbiome study in a common mouse model.

A genome resolved inventory of functional potential changes in the pathogen inflamed gut

Previous studies of gut microbiomes highlight the role microbial metabolites play in gastrointestinal inflammation as signaling effectors in host immune regulation [63, 64]. The CBAJ-DB showcases the juxtapose of pro-inflammatory and anti-inflammatory microbial membership and gene content in CBA/J mice and during *Salmonella* infection. *Salmonella*-induced inflammation shifted the functional potential of infected communities favoring respiring organisms, marked by a reduction of butyrate and acetate producers and an increase in bacteria with anaerobic respiration capability. Convention indicates butyrate agonism of PPAR- γ and SCFA engagement with G-protein coupled receptors respectively help to maintain luminal anaerobiosis and promote colonic regulatory T cell development [6, 65, 66]. We showed specific bacteria reduction coincides with an inflamed state and diminished SCFA production potential in the gut.

Our findings indicate *Alistipes* reduction following inflammation as a possible cause of butyrate production potential loss, contrasting with current dogma linking butyrate production in the gut chiefly to *Clostridia* abundance [11, 67–69]. Furthermore, *Salmonella* infection enriched certain *Clostridia* including *Dorea*, *Faecalicatena*, and a novel bacteria described only at the class level, highlighting the functional redundancy that may be provided within this class. Interestingly, the mouse with the lowest *Salmonella* relative abundance had the greatest diversity of *Clostridia*. It is interesting to speculate that this lineage and the diversity within it may be important for microbiota recolonization and return to homeostasis following gastric infection, a notion supported by previous research [70].

Salmonella-induced inflammation also caused a reduction in bacteria with the capability to mediate anti-inflammatory microbial metabolites. Bacteria with genes for secondary bile acid production and tryptophan catabolism were decreased in inflamed metagenomes, a response previously shown to increase host susceptibility to infection [71]. Specifically, reduced bile acid can limit ligands like pregnane X (PXR) and farnesoid X (FXR) which are important regulators of the host anti-inflammatory response, thus reduced production of these genes could have further feedback on inflammation [64, 65]. Similarly, indole and indole derivatives like indole acrylic acid, indole-3-probionate,

and indole-3-lactate are AhR and PXR agonists and thus anti-inflammatory [49, 72]. Here we demonstrate how the reduction of bacteria capable of producing these important host pathway modulators can further promote inflammation and *Salmonella* expansion, evidenced by high lipocalin-2 levels concomitant with high *Salmonella* relative abundance and lower functional potential for bile acid deconjugation, secondary bile acid production, and tryptophan catabolism in inflamed mice. Future research directly measuring transcription and metabolite concentrations can be used in concert with the CBAJ-DB to determine the anti-inflammatory impact of individual bacteria on the microbiome.

Commensal bacteria that can withstand inflammation may represent future biological therapeutic opportunities

A recent rise of antibiotic-resistant *Salmonella* strains globally underpins the need for alternative treatments and prevention measures against foodborne pathogens. One avenue may be the use of probiotic bacteria to reduce the intensity or duration of infection [49, 73, 74]. *Alistipes* sp002428825 and *Akkermansia muciniphila* clustered closely with genomes from human microbial communities and we explore here their potential as anti-inflammatory probiotics.

Akkermansia muciniphila is a well-known commensal gut bacterium in mammals that lives in the lumen mucosal layer and contributes to epithelial gut barrier integrity [75, 76]. Nevertheless, one study showed *Akkermansia muciniphila* exacerbated inflammation and increased *Salmonella typhimurium* relative abundance [77]. These findings are inconsistent with our data however, where *Akkermansia muciniphila* is relatively abundant and consistently present in uninfected and *Salmonella*-infected mice. Genome analysis reveals the capacity of *Akkermansia muciniphila* to produce indole, potentially an important anti-inflammatory mechanism and *Salmonella* deterrent. Other studies have shown the effectiveness of indole from *E. coli* increasing tight junctions of the gut epithelial and decreasing *Salmonella* pathogenicity, though more work is needed to confirm similar action from *Akkermansia* in buffering gut inflammation [78].

Alistipes sp002428825 was also consistently detected in both treatments. Analysis of this genome suggests it can respire oxygen and directly compete with *Salmonella* for arabinan, arabinose, and pectin, while maintaining critical gut homeostasis functionalities like butyrate production. *Alistipes* spp. are often associated with healthy human microbiomes [79, 80] and have even been shown to facilitate microbiota recolonization following perturbation [81]. We have shown *Alistipes* has saccharolytic genes to metabolize rhamnose and

fucose, and it is possible that Gram-positive bacteria eliminated by inflammation and fucosylated proteins from host epithelial erosion could support *Alistipes* during *Salmonella* infection as sources of these sugars [82, 83]. Given this genus, like *Akkermansia muciniphila*, remained abundant despite host inflammation, and both bacteria contain genes for indole derivative production, it also may be directly antagonistic to *Salmonella* through anti-inflammatory effector potential.

It also bears mentioning the presence of lactic acid bacteria *E. gallinarum* and *L. johnsonii* in the CBAJ-DB in relatively high abundance during *Salmonella* infection. In fact, these members were previously illuminated in prior studies from mice as well as clinical patients, demonstrating co-occurrence may exist beyond this single study or mouse model [84]. Beyond resisting or maybe even responding to conditions created by *Salmonella* infection, we provide genomic evidence supporting *L. johnsonii* nutrient competition for arabinose, mixed-linkage glucans, and amorphous cellulose, while *E. gallinarum* may compete for chitin, pectin, and arabinan. Additionally, closely related species to these have been shown to produce anti-*Salmonella* agents like bacteriocin and organic acids putatively decreasing *Salmonella* abundance over time [85–87]. Given these species have many members already approved as probiotics and our data indicate CBA/J mice harbor at least one species of *Enterococcus* common to human guts and *Lactobacillus* spp. resistant to host inflammation, a probiotic lactic acid bacteria strain resistant to *Salmonella* may reside in the CBAJ-DB [88]. This genome resolved research identifies future targets with promising potential for exploration as probiotics robust to *Salmonella* infection. Yet, we recognize that commensal bacteria like these can be pathobionts provided the right setting [54, 77, 89, 90], such that future research in this mouse model would first include challenge experiments with isolates or even targeted consortia that provide multiple avenues of overlapping pathogen colonization resistance.

Conclusion

The CBAJ-DB uniquely captures gut community variation in CBA/J mice. MAG reconstruction from metagenomic sequencing enabled us to profile the functional potential of the murine gut microbiome during acute *Salmonella* inflammation, contrasting community membership and gene content with uninfected mice. Persisting taxa in the inflamed gut encoded the capability to withstand or utilize changing redox conditions, while bacteria producing SCFA and producing host anti-inflammatory effectors decreased in mice with high *Salmonella* burden. Further, our phage analyses leave open the possibility that

phage infection could alter *Firmicutes* energy regulation and spore formation.

Together, these findings validate physiological investigations performed with reduced complexity synthetic or modified gut microbiota. We also provide new perspectives that advance the understanding of *Salmonella* effect on an intact microbiome and provide model specificity to CBA/J gut consortia. Our efforts show novel bacteria unique to the CBA/J mouse model and enriched by *Salmonella* infection. An exploration of potential probiotic targets in the CBAJ-DB revealed multiple lactic acid bacteria capable of withstanding the host immune response to *Salmonella* and that may be indifferent to *Salmonella* competition. Additionally, genomes were recovered of *A. muciniphila* and *E. gallinarum* with species similarity to bacteria in human guts. The CBAJ-DB is the first culture-independent murine genome catalog to include sampling from CBA/J mice and inflamed individuals, providing a resource with application to multi-omic microbiome investigation, gut inflammation research, and studies involving the CBA/J mouse model broadly.

Methods

Strains and media

S. enterica serovar Typhimurium strain 14,028 (*S. typhimurium* 14,028) was cultured overnight in Luria-Bertani (LB) broth at 37 °C with constant agitation. Overnight culture was washed and resuspended in water. *S. typhimurium* 14,028 ASV was determined manually from an identical sequence match to the 16S region of NCBI Reference Sequence NZ_CP034230.1 mapped with Geneious Prime® 2020.1.2.

Animals and experimental design

Female CBA/J mice were purchased from The Jackson Laboratory (Bar Harbor, ME) and housed 5 per cage in conventional enclosures maintained in a temperature controlled 12-h light/dark cycle. To mitigate microbiome differences caused by variables other than *Salmonella*, we housed all mice in the same room, and mice were chosen at random when populating cages and assigning treatment. Irradiated mouse chow (Teklad, 7912) was made available ad libitum to mice in the control group ($n=16$) and to infected mice ($n=14$) housed separately. Individuals in this study were chosen based on fecal sample availability at day 11, and only infected mice with $\geq 25\%$ *S. typhimurium* 14,028 at the sampling time were used. Mice in the infected group received 10^9 CFU *S. typhimurium* 14,028 oral gavage on day 0 with no subsequent treatment, and control group mice were left without treatment. Animal experiment protocol was approved by The Ohio State University Institutional

Animal Care and Use Committee (IACUC; OSU 2009A0035).

Sample collection

Fecal pellets were collected from each mouse 1 day prior to treatment and 10 and 11 days after treatment initiation on autoclaved aluminum foil. Pellets were immediately placed in labeled microcentrifuge tubes and flash frozen with EtOH/dry ice prior to storage at -80 °C until further processing.

Lipocalin-2 quantification

Vortex homogenization of fecal sample in PBS containing 0.1% Tween 20 (100 mg/ml) for 20 min was performed prior to centrifugation of the resultant suspension at 12,000 rpm for 10 min at 4 °C. The resulting supernatant was used to measure levels of inflammation marker Lipocalin-2 using the DuoSet murine Lcn-2 ELISA kit (R&D Systems, Minneapolis, MN).

DNA extraction and sequencing

Total nucleic acids were extracted using the *Quick-DNA* Fecal/Soil Microbe Microprep Kit (Zymo Research) and stored at -20 °C until amplicon or metagenomic sequencing. DNA was submitted for amplicon sequencing at Argonne National Lab at the Next Generation Sequencing facility using Illumina MiSeq with 2×251 bp paired end reads following established HMP protocols [91]. Briefly, universal primers 515F and 806R were used for PCR amplification of the V4 hypervariable region of 16S rRNA gene using 30 cycles. The 515F primer contained a unique sequence tag to barcode each sample. Both primers contained sequencer adapter regions. DNA for metagenomes was submitted to the Genomics Shared Resource facility at Ohio State University and was prepared for sequencing with a Nextera XT library system followed by solid-phase reversible immobilization size selection. Libraries were quantified and then sequenced using an Illumina HiSeq platform.

16S rRNA preprocessing

Amplicon sequencing fastq data were processed in a QIIME2 2019.10.0 environment [92]. Reads were demultiplexed and then denoised with DADA2 [93]. For all sequencing runs ($n=4$), forward reads were truncated at 246 bps and reverse reads were truncated at 167 bps. Feature tables from each sequencing run were combined, and ASVs were assigned taxonomy with the silva-132-99-515-806-nb-classifier [94]. Before further analysis, the ASV table was filtered with R version 4.0.2 to (i) remove samples with no ASVs, (ii) to remove samples

with a combined ASV count < 1000 across all samples, (iii) to remove ASVs with 0 abundance in every sample, and (iv) to remove ASVs designated as mitochondria and chloroplast. The resulting filtered feature table contained 23,022 ASVs. Raw reads are deposited on NCBI (PRJNA348350) and the final ASV table is published in the supplementary materials (Additional file 1: Data S1). Reads for external study comparison were obtained from the NCBI Sequence Read Archive under accession number SRP057511 (O'loughlin [28]) and under bioproject PRJNA348350 (Borton [27]). All 16S rRNA amplicon data for comparison analyses were treated as described above. ASV relative abundances were then summed within genera or the lowest annotated taxonomy level when no classification was provided.

Genome reconstruction from metagenomes

All bioinformatic tools were run with default parameters unless otherwise specified. Quality scores of raw metagenomic reads were evaluated using FastQC (v0.11.9, [95]). Reads were trimmed, adapters removed, and mouse reads removed using BBDuk (ktrim=r, k=23, mink=11, hdist=1, qtrim=rl, trimq=20, minlen=75, maq=10) from BBTools (v38.89, <https://jgi.doe.gov/data-and-tools/bbtools>). Trimmed reads from each individual sample were assembled with Megahit (v1.1.1, [96]) and with IDBA_UD (v1.1.3, [97]). Each assembler was also used to perform co-assembly of all the samples ($n=6$) at once.

Subsequently, each single-sample and co-sample assembly was binned separately. To obtain bins, reads were mapped to assembly contig or scaffold set filtered to ≥ 2500 bps using BBDuk (<https://jgi.doe.gov/data-and-tools/bbtools>), sorted using SAMtools (v1.9, [98]), and then binned with Metabat2 (v2.12.1, [99]). The resulting MAGs were checked for quality and contamination with CheckM (v1.1.2, [30]), CheckM2 (v0.1.3, [31]), and GUNC (v1.0.5, [32]). Using the resulting medium-quality (completeness $\geq 50\%$, contamination < 10%) and high-quality (completeness $\geq 90\%$, contamination < 5%) MAGs from the initial single-sample and co-sample assemblies, trimmed reads were mapped using BBDuk in perfect mode. Reads that did not map were assembled individually by-sample and co-assembled by treatment with IDBA_UD (v1.1.3, [97]) to create subtractive assemblies. The resulting subtractive contigs or scaffolds were then subject to previously described filtering (≥ 2500 bps), processing, binning, and quality check.

To construct the final CBAJ-DB, all MQHQ MAGs from all assemblies were assigned taxonomy with GTDB-Tk (v1.3.0, r95, [100], classify_wf) and

dereplicated with dRep (v2.5.4, 99% ANI, [101]). Conventional assembly and binning of reads rarified with BBTools for MAG recovery sensitivity to read depth was performed with the Snakemake [102] workflow included here (https://github.com/ilelewi/metaG_pipeline/blob/main/Snakefile_5.88_onesamp).

16S rRNA linked to MAGs

Mining of MAGs for 16S rRNA genes was performed with Wrighton Lab software (https://github.com/WrightonLabCSU/join_asvbins) MMseqs2 (v13.45111 [103],) and Barrnap (v0.9 [104],) and SILVA reference database (SILVA_138_SSURef_NR99_tax_silva [94],) followed by pairwise comparison to the V4 region sequences from our amplicon sequencing.

Database mapping and comparison to other MAG resources

To calculate relative abundance of individual dMQHQ MAGs, BBmap (<https://jgi.doe.gov/data-and-tools/bbtools>) was used to randomly map trimmed reads to the dMQHQ with minid=0.95. Then CoverM (v0.6.0, <https://github.com/wwood/CoverM>) was used to estimate read counts per scaffold and per bin. Values are the mean number of aligned reads calculated after removing positions with the most and the least coverage as determined by default values (methods:proper-pairs-only -m trimmed_mean -min-read-percent-identity-pair 0.95, -proper-pairs-only -m mean -min-read-percent-identity-pair 0.95 -min-covered-fraction 0.75, -proper-pairs-only -m reads_per_base -min-read-percent-identity-pair 0.95 -min-covered-fraction 0). Relative abundances of mapped reads were then GeTMM normalized (Additional file 2: Data S2) [105] using the edgeR package (v3.36.0 [106],). To place MAGs in either treatment (control, infected) or both treatments, CoverM was used and mapping was only considered if the subject covered fraction exceeded 75% with 95% sequence identity and a minimum of 3 reads per base depth.

To estimate the relative abundance of each vMAG, the metagenomic reads were mapped using Bowtie2 [107]. Reads were mapped using BBMap with minid=0.95. Afterwards, CoverM was run using the -mean option to consider only those vMAGs that have >75% of their fraction covered. Relative abundances for each vMAG were calculated as their coverage proportion from the sum of the whole coverage of all bins for each set of metagenomic reads prior to GeTMM normalization and are reported in Additional file 4: Data S4.

Isolate MAGs were obtained from the Human Microbiome Project [39], and medium- and high-quality MAGs were obtained from <https://doi.org/10.5281/>

[zenodo.4977712](https://zenodo.org/record/4977712). Metagenomic reads from the Human cohort and Lloyd-Price cohort [43] were obtained from PRJNA725020 and the SRA Database Commons respectively. Human cohort rarefaction to 2Gbps was performed with BBTools reformat. Human metagenomic reads were mapped to MQHQ bins with fastANI (v1.32, [108]). Genome matches with ANI $\geq 94\%$ and alignment fraction (AF) $\geq 33\%$ were deemed of the same species and genome matches with ANI $\geq 73\%$ and AF $\geq 33\%$ of the same Genus [109].

Comparison to mouse database MAGs was performed on MGBC [22] non-redundant high-quality genomes (Additional file 5: Data S5, MGBC-hqnr_26640) and iMGMC [21] dereplicated medium- and high-quality MAGs (Additional file 5: Data S5, iMGMC-mMAGs-dereplicated_genomes). iMGMC MAGs were re-classified with GTDB-Tk version 1.3.0 r95 to align with CBAJ-DB and MGCB taxonomy, and dereplication with each MAG set and CBAJ-DB was performed with dRep (v2.5.4, 99% ANI).

MAG function analysis

CBAJ-DB MAGs were annotated using DRAM (v1.1.1) [34] and dbCAN (v3.0.2, dbCAN-HMMdb-V10, [110]). CAZy genes called with DRAM were updated with the latest dbCAN database and parsed with HMMer (v3.3, [111]) to include only significant hits (e -value $< 10^{-18}$) with $> 35\%$ coverage. Wrighton lab software rule_adjectives (https://github.com/rmFlynn/rule_adjectives) and functions_pa (https://github.com/ilelewi/functions_pa) were used to parse KEGG ids, Enzyme Commission numbers, and dbCAN ids from gene annotations referencing function rule sheets (Additional file 7: Data S7) to determine function presence or absence in each MAG. Function relative abundance significance was determined with ANOVA performed using the R stat package (v4.1.3).

Viral host-linkage and AMGs

Metagenomic assemblies and subassemblies ($n=16$) were screened for DNA viral sequences using VirSorter2 (v2.2.2, [112]) using the published protocol in Protocols.io [113]. Briefly, VirSorter2 was run using parameters “-include-groups dsDNAphage, ssDNA,” “-min-length 10,000,” and “-min-score 0.5.” The resulting VirSorter2 output was then run through CheckV [114] to ensure quality viral sequences using the “end_to_end” function. The trimmed viral sequences output by CheckV were then once again run through VirSorter2 using options above with additional flags “-seqname-suffix-off,” “-viral-gene-enrich-off,” and “-prep-for-dramv.” The final output was then manually curated using the CheckV output as described in Protocols.io. Briefly, (1) viral-like scaffolds that had more than 1 viral gene were kept

and deemed viral, (2) viral-like scaffolds with no viral genes, host genes equal to zero, or score ≥ 0.95 , or that had 1 host gene with a length of ≥ 10 kb were separated and further inspected. Scaffolds not meeting the above criteria or manually inspected to be non-viral were discarded. After generation of a curated, quality-controlled viral vMAGs, they were clustered at 95% identity across 85% of the shortest contig representing viral populations using ClusterGenomes [113].

To determine taxonomic affiliation, vMAGs were clustered to viruses belonging to viral reference taxonomy databases NCBI Bacterial and Archaeal Viral RefSeq V85 with the International Committee on Taxonomy of Viruses (ICTV) and NCBI Taxonomy using the network-based protein classification software vConTACT2 (v0.9.8, [115]) using default methods [116]. To determine if the viruses present in this study represented relevant communities across mammalian gut ecosystems, we included viruses mined from 5 publicly available datasets in our vConTACT2 analyses from (1) human guts [24, 43, 52], moose rumen [38], and bird / other mammalian guts [53]. The viral sequences that were identified from these systems and the genes used for vConTACT2 are deposited on Zenodo with <https://doi.org/10.5281/zenodo.7144664> with more information of downloaded datasets found in Additional file 4 Data S4.

Viral contigs were annotated with DRAM-v [34]. Auxiliary scores were assigned by DRAM, based on the following ranking system: A gene is given an auxiliary score of 1 if there is at least one hallmark gene on both the left and right flanks, indicating the gene is likely viral. An auxiliary score of 2 is assigned when the gene has a viral hallmark gene on one flank and a viral-like gene on the other flank. An auxiliary score of 3 is assigned to genes that have a viral-like gene on both flanks. All vMAG annotations are reported in Additional file 4: Data S4. To identify likely vMAG hosts, we used two strategies which included (1) linking viral spacers found in CRISPR systems assembled using CRASS [117] and (2) oligonucleotide frequencies between virus and hosts using VirHostMatcher and a threshold of $d2^*$ measurements of < 0.25 [118]. The lowest $d2^*$ value for each viral contig < 0.2 was used, and only vMAGs for which the top 3 hits had taxonomic consensus at the genera level were considered “good” hits [118]. All virus-host links are reported in Additional file 4: Data S4.

Spearman correlation of metagenomic and amplicon communities

Relative abundance of mapped reads to dMQHQ MAGs were averaged within treatment, and the total relative abundance of each MAG in *Bacteroidia*, *Clostridia*, *Verrucomicrobiae*, *Gammaproteobacteria*, *Bacilli*, and

Coriobacteriia Classes was summed respectively. Relative abundance of MAGs from all other classes were summed together into an additional category. Similarly, 16S amplicon sequence ASV relative abundances from high responder mice and control mice were summed within each treatment and by taxonomy according to the classes previously mentioned, combining all other ASV abundances into an additional category. Spearman correlation was then performed with the R stats package comparing Class abundances between metagenomic communities and ASV communities within the same treatment.

Statistical analysis

Alpha (Shannon's diversity) and Beta diversity (Bray–Curtis dissimilarity) were calculated with ASV or MAG relative abundances from the filtered data using the Vegan package 2.5.7 in R [119]. The NMDS Beta diversity visualization was produced using ggplot in R [120]. To determine significant grouping of samples by treatment, we performed an analysis of similarities and mrpp [121], with a stress test determining goodness of fit [121]. Lipocalin-2 (ng/g) and *S. typhimurium* 14,028 relative abundance significance between treatments was determined using a Wilcoxon rank sum test, the same test used to compare class relative abundance between different treatments and timepoints. Linear discriminate analysis on MAGs and ASVs was done with LEfSe [122]. MAG and vMAG ordination coordination was determined by Procrustes analysis [119]. MAG quality quartile significance between groups was calculated with a chi squared test in R.

Abbreviations

CBAJ-DB	The CBAJ database
MAG	Metagenome assembled genome
vMAG	Viral metagenome assembled genome
ASV	Amplicon sequencing variant
NTS	Non-Typhoidal <i>Salmonella</i>
MQHQ	Medium and high-quality
dMQHQ	Dereplicated medium and high-quality
AMG	Auxiliary metabolic gene
iMGMC	Integrated Mouse Gut Metagenomic Catalog
MGBC	Mouse Gastrointestinal Bacterial Catalogue
PXR	Pregnane X receptor
FXR	Farnesoid X receptor
AhR	Aromatic hydrocarbon receptor
SCFA	Short-chain fatty acid

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-023-01529-7>.

Additional file 1. Additional data pertaining to 16S rRNA amplicon sequencing including IDs of mice sequenced and their treatment (sheet: 16S_rRNA_Metadata). The file also includes sheets with Lipocalin-2 measurements (sheet: Lipocalin_2), ASV abundances (sheet: ASV_table), ASV taxonomy (sheet: Taxonomy (Silva 132)), and 16s sequences found in MAGs (sheet: 16S_in_MAGs).

Additional file 2. MAG IDs of medium and high-quality MAGs along with their completeness and contamination scores, bin size, and contig information (sheet: MQHQ_stats). tRNA location in each bin (sheet: MQHQ_tRNA) and sampling depth (sheet: Sampling_Depth) along with read mapping results (sheets: Mapping (95%), Mapping (strict)).

Additional file 3. DRAM gene annotations of the MQHQ MAG in the CBAJ-DB.

Additional file 4. All data pertaining to the recovered vMAGs in the CBAJ-DB including the dereplicated vMAG set (sheet: UViG_info_final_609), read mapping abundances (sheet: read_mapping_abunces_609_vMAGs), host linkages (sheet: virus_host_linkages) and AMG information (sheet: 36_host_linked_AMGs).

Additional file 5. dRep and FastANI clustering results of CBAJ-DB with iMGMC, MGBC, and human MAGs.

Additional file 6. Read mapping data from Jason Lloyd Price cohort and human cohort mapped to the CBAJ-DB.

Additional file 7. Individual MAG function presence and relevant gene counts along with the rule sets used to assign genome functionality.

Additional file 8: Fig. S1. Relative abundance of classes in pre-infection communities are not statistically different than uninfected communities, indicating a shared starting microbiome prior to infection. **Fig. S2.** CBAJ-DB uninfected (no *Salmonella*) amplicon sequenced communities show considerable taxonomic overlap with communities from other CBA studies. **Fig. S3.** A) Gigabase pairs (Gbps) per sample of prevalent murine genome databases. **Fig. S4.** Contamination and completion statistics and the most resolved taxonomy groups for MAGs containing amplicon sequencing variants (ASVs). **Fig. S5.** Procrustes analysis of dereplicated medium and high quality (dMQHQ) metagenome assembled genomes (MAGs) and viral metagenome assembled genomes (vMAGs).

Acknowledgements

The authors would like to thank Tyson Claffey and Richard Wolfe for Colorado State University server management; Sandy Shew for management of computing resources retained from The Ohio State University Unity cluster; and Dr. Pearly Yan at the Genomics Shared Resource Core at The Ohio State University Comprehensive Cancer Center for management of metagenomic sequencing.

Authors' contributions

IL, MS, JRR, and MAB analyzed and interpreted sequencing data. ASD, IL, MS handled mice and sample collection and ASD performed lipocalin-2 assays. RAD was responsible for DNA extractions and quality control. KK, LK, and LMS provided conceptual framework and editorial support. RMF and IL wrote function parsing scripts and MAB and KCW provided function rules. ASD and BA handled *Salmonella* aspects of the project. The author(s) read and approved the final manuscript.

Funding

This work was supported by NIH NIAID R01A1143288 awarded to B.M.A. and K.C.W. I.L. was funded by a fellowship awarded through the NIH Predoctoral Training in Quantitative Cell & Molecular Biology grant T32GM132057. This work was also partially supported by DOE Office of Science, Office of Biological and Environmental Research (BER), grant no. DE-SC0021350.

Availability of data and materials

The sequence data supporting the results of this article are available in the National Center of Biotechnology Information (NCBI) under Bioproject number PRJNA348350. All sequencing outputs including the ASV table and fasta files are included on Zenodo (<https://doi.org/10.5281/zenodo.7144664>).

Declarations

Ethics approval and consent to participate

Mouse experiments in this study were performed in accordance with protocols approved by The Ohio State University Institutional Animal Care and Use Committee (IACUC; OSU 2009A0035-R4).

Consent for publication

All authors consent to the manuscript submission to *Microbiome*.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Cell and Molecular Biology, The Colorado State University, Fort Collins, CO, USA. ²Department of Soil and Crop Sciences, The Colorado State University, Fort Collins, CO, USA. ³Graduate Degree Program in Ecology, The Colorado State University, Fort Collins, CO, USA. ⁴Department of Microbial Infection and Immunity, The Ohio State University, Columbus, OH, USA. ⁵Department of Microbiology, Immunology, and Pathology, The Colorado State University, Fort Collins, CO, USA. ⁶Department of Microbiology, RIBES, Radboud University, Nijmegen, The Netherlands. ⁷Department of Microbiology and Biophysics, The Ohio State University, Columbus, OH, USA.

Received: 4 November 2022 Accepted: 21 March 2023

Published online: 20 May 2023

References

- Stanaway JD, Parisi A, Sarkar K, Blacker BF, Reiner RC, Hay SI, et al. The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Infect Dis Elsevier*. 2019;19:1312–24.
- Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, et al. The Global Burden of Nontyphoidal Salmonella Gastroenteritis. *Clin Infect Dis*. 2010;50:882–9.
- Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson M-A, Roy SL, et al. Foodborne illness acquired in the United States—major pathogens. *Emerg Infect Dis*. 2011;17:7–15.
- Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, et al. Salmonella enterica Serovar Typhimurium exploits inflammation to compete with the intestinal microbiota. *PLOS Biol*. 2007;5:e244 (Public Library of Science).
- Barman M, Unold D, Shifley K, Amir E, Hung K, Bos N, et al. Enteric salmonellosis disrupts the microbial ecology of the murine gastrointestinal tract. *Infect Immun*. 2008;76:907–15.
- Rogers AWL, Tsois RM, Bäumlner AJ. Salmonella versus the microbiome. *Microbiol Mol Biol Rev*. 2020;85:e00027-19 (American Society for Microbiology).
- Stecher B. Establishing causality in Salmonella-microbiota-host interaction: the use of gnotobiotic mouse models and synthetic microbial communities. *Int J Med Microbiol*. 2021;311:151484.
- de Vos WM, Tilg H, Hul MV, Cani PD. Gut microbiome and health: mechanistic insights. *Gut* BMJ Publishing Group. 2022;71:1020–32.
- Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, et al. Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature*. 2010;467:426–9 (Nature Publishing Group).
- Shelton CD, Yoo W, Shealy NG, Torres TP, Zieba JK, Calcutt MW, et al. Salmonella Typhimurium uses anaerobic respiration to overcome propionate-mediated colonization resistance. *bioRxiv*. Cold Spring Harbor Laboratory; 2021;2021.05.25.445690.
- Rivera-Chávez F, Bäumlner AJ. The pyromaniac inside you: Salmonella metabolism in the host gut. *Annu Rev Microbiol*. 2015;69:31–48.
- Walker GT, Raffatellu M. Salmonella respiration turns the tables on propionate. *Trends Microbiol*. 2022;30:206–8 (Elsevier).
- Rivera-Chávez F, Zhang LF, Faber F, Lopez CA, Byndloss MX, Olsan EE, et al. Depletion of butyrate-producing Clostridia from the gut microbiota drives an aerobic luminal expansion of Salmonella. *Cell Host Microbe*. 2016;19:443–54.
- Sekirov I, Tam NM, Jogova M, Robertson ML, Li Y, Lupp C, et al. Antibiotic-induced perturbations of the intestinal microbiota alter host susceptibility to enteric infection. *Infect Immun*. 2008;76:4726–36 (American Society for Microbiology).
- Woo H, Okamoto S, Guiney D, Gunn JS, Fierer J. A model of Salmonella colitis with features of diarrhea in SLC11A1 wild-type mice. *PLOS ONE*. 2008;3:e1603 (Public Library of Science).
- Ferreira RBR, Gill N, Willing BP, Antunes LCM, Russell SL, Croxen MA, et al. The intestinal microbiota plays a role in Salmonella-induced colitis independent of pathogen colonization. *PLoS ONE*. 2011;6:e20338.
- Ahmer BM, Gunn JS. Interaction of Salmonella spp. with the intestinal microbiota. *Front Microbiol*. Frontiers. 2011 [cited 2020 Apr 16];2. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2011.00101/full>
- Karlinsey JE, Maguire ME, Becker LA, Crouch M-LV, Fang FC. The Phage Shock Protein PspA facilitates divalent metal transport and is required for virulence of Salmonella enterica sv Typhimurium. *Mol Microbiol*. 2010;78:669–85.
- Shelton CD, Yoo W, Shealy NG, Torres TP, Zieba JK, Calcutt MW, et al. Salmonella enterica serovar Typhimurium uses anaerobic respiration to overcome propionate-mediated colonization resistance. *Cell Rep*. Elsevier; 2022 [cited 2022 Jul 14];38. Available from: [https://www.cell.com/cell-reports/abstract/S2211-1247\(21\)01680-6](https://www.cell.com/cell-reports/abstract/S2211-1247(21)01680-6)
- Spiga L, Winter MG, de Carvalho TF, Zhu W, Hughes ER, Gillis CC, et al. An oxidative central metabolism enables Salmonella to utilize microbiota-derived succinate. *Cell Host Microbe*. 2017;22:291–301.e6.
- Lesker TR, Durairaj AC, Gálvez EJC, Lagkourdos I, Baines JF, Clavel T, et al. An integrated metagenome catalog reveals new insights into the murine gut microbiome. *Cell Rep*. 2020;30:2909–2922.e6.
- Beresford-Jones BS, Forster SC, Stares MD, Notley G, Viciani E, Browne HP, et al. The Mouse Gastrointestinal Bacteria Catalogue enables translation between the mouse and human gut microbiotas via functional mapping. *Cell Host Microbe*. 2022;30:124–138.e8.
- Wong E-OY, Brownlie EJE, Ng KM, Kathirgamanathan S, Yu FB, Merrill BD, et al. The CIAMIB: a large and metabolically diverse collection of inflammation-associated bacteria from the murine gut. *mBio*. 2022;13:e02949-21 (American Society for Microbiology).
- Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The gut virome database reveals age-dependent patterns of virome diversity in the human gut. *Cell Host Microbe*. 2020;28:724–740.e8.
- Adiliaghdam F, Jeffrey KL. Illuminating the human virome in health and disease. *Genome Med*. 2020;12:66.
- Cao Z, Sugimura N, Burgermeister E, Ebert MP, Zuo T, Lan P. The gut virome: a new microbiome component in health and disease. *eBioMedicine*. Elsevier; 2022 [cited 2022 Jul 11];81. Available from: [https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964\(22\)00294-8/fulltext](https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(22)00294-8/fulltext)
- Borton MA, Sabag-Daigle A, Wu J, Solden LM, O'Banion BS, Daly RA, et al. Chemical and pathogen-induced inflammation disrupt the murine intestinal microbiome. *Microbiome*. 2017 [cited 2019 Aug 26];5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408407/>
- O'Loughlin JL, Samuelson DR, Braundmeier-Fleming AG, White BA, Halldorson GJ, Stone JB, et al. The intestinal microbiota influences Campylobacter jejuni colonization and extraintestinal dissemination in mice. *Appl Environ Microbiol*. 2015;81:4642–50 (American Society for Microbiology).
- Xiao L, Feng Q, Liang S, Sonne SB, Xia Z, Qiu X, et al. A catalog of the mouse gut metagenome. *Nat Biotechnol*. 2015;33:1103–8 (Nature Publishing Group).
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ Inc*; 2015 May. Report No.: e1346. Available from: <https://peerj.com/preprints/554>
- Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *bioRxiv*; 2022 [cited 2023 Feb 5]. p. 2022.07.11.499243. Available from: <https://www.biorxiv.org/content/10.1101/2022.07.11.499243v1>
- Orakov A, Fullam A, Coelho LP, Khedkar S, Szklarczyk D, Mende DR, et al. GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome Biol*. 2021;22:178.
- Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*. 2017;35:725–31.
- Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*. 2020;48:8883–900.
- Mise K, Iwasaki W. Unexpected absence of ribosomal protein genes from metagenome-assembled genomes. *ISME Commun*. 2022;2:1–9 (Nature Publishing Group).
- Microbial Genome-Resolved Metaproteomic Analyses Frame Intertwined Carbon and Nitrogen Cycles in River Hyporheic Sediments. 2021 [cited 2023 Feb 27]. Available from: <https://www.researchsquare.com>

37. McGivern BB, Tfaily MM, Borton MA, Kosina SM, Daly RA, Nicora CD, et al. Decoding bacterial polyphenol metabolism in an anoxic wetland soil. *Nat Commun.* 2021;12:2466 (Nature Publishing Group).
38. Solden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, et al. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat Microbiol.* 2018;3:1274–84 (Nature Publishing Group).
39. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The human microbiome project. *Nature.* 2007;449:804–10 (Nature Publishing Group).
40. Borton MA, Shaffer M, Hoyt DW, Jiang R, Ellenbogen J, Purvine S, et al. Targeted curation of the gut microbial gene content modulating human cardiovascular disease. *bioRxiv*; 2022 [cited 2022 Aug 12]. p. 2022.06.20.496735. Available from: <https://www.biorxiv.org/content/10.1101/2022.06.20.496735v1>
41. Nguyen TLA, Vieira-Silva S, Liston A, Raes J. How informative is the mouse for human gut microbiota research? *Dis Model Mech.* 2015;8:1–16.
42. Park JC, Im S-H. Of men in mice: the development and application of a humanized gnotobiotic mouse model for microbiome therapeutics. *Exp Mol Med.* 2020;52:1383–96 (Nature Publishing Group).
43. Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature Nature.* 2019;569:655–62 (Publishing Group).
44. Ghaffari S, Abbasi A, Somi MH, Moaddab SY, Nikiiaz L, Kafil HS, et al. Akkermansia muciniphila: from its critical role in human health to strategies for promoting its abundance in human gut microbiome. *Crit Rev Food Sci Nutr.* 2022;0:1–21 (Taylor & Francis).
45. Choi S-H, Lee S-O, Kim TH, Chung J-W, Choo EJ, Kwak YG, et al. Clinical features and outcomes of bacteremia caused by Enterococcus caseliflavus and Enterococcus gallinarum: analysis of 56 cases. *Clin Infect Dis.* 2004;38:53–61.
46. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, et al. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol.* 2021;39:105–14 (Nature Publishing Group).
47. Faber F, Thiennimitr P, Spiga L, Byndloss MX, Litvak Y, Lawhon S, et al. Respiration of microbiota-derived 1,2-propanediol drives Salmonella expansion during colitis. *PLoS Pathog.* 2017;13:e1006129.
48. Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, Huseby DL, et al. Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proc Natl Acad Sci.* 2011;108:17480–5 (Proceedings of the National Academy of Sciences).
49. Hyland NP, Cavanaugh CR, Hornby PJ. Emerging effects of tryptophan pathway metabolites and intestinal microbiota on metabolism and intestinal function. *Amino Acids.* 2022;54:57–70.
50. Agus A, Planchais J, Sokol H. Gut microbiota regulation of tryptophan metabolism in health and disease. *Cell Host Microbe.* 2018;23:716–24.
51. Cusotto S, Delgado I, Anesi A, Dexpert S, Aubert A, Beau C, et al. Tryptophan metabolic pathways are altered in obesity and are associated with systemic inflammation. *Front Immunol.* 2020 [cited 2022 Jul 5];11. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2020.00557>
52. Du J, Zayed AA, Kigerl KA, Zane K, Sullivan MB, Popovich PG. Spinal cord injury changes the structure and functional potential of gut bacterial and viral communities. *mSystems.* 2021;6:e01356-20 (American Society for Microbiology).
53. Wang H, Ling Y, Shan T, Yang S, Xu H, Deng X, et al. Gut virome of mammals and birds reveals high genetic diversity of the family Microviridae. *Virus Evol.* 2019;5:vez013.
54. Yang Y, Nguyen M, Khetrapal V, et al. Within-host evolution of a gut pathobiont facilitates liver translocation. *Nature.* 2022;607:563–70. <https://doi.org/10.1038/s41586-022-04949-x>.
55. Hugenholtz F, de Vos WM. Mouse models for human intestinal microbiota research: a critical evaluation. *Cell Mol Life Sci.* 2018;75:149–60.
56. Ursell LK, Metcalf JL, Parfrey LW, Knight R. Defining the human microbiome. *Nutr Rev.* 2012;70:S38-44.
57. Kieser S, Zdobnov EM, Trajkovski M. Comprehensive mouse microbiota genome catalog reveals major difference to its human counterpart. *PLoS Comput Biol.* 2022;18:e1009947 (Public Library of Science).
58. Velazquez EM, Nguyen H, Heasley KT, Saechao CH, Gil LM, Rogers AWL, et al. Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nat Microbiol.* 2019;4:1057–64 (Nature Publishing Group).
59. Nayfach S, Pérez-Espino D, Call L, Low SJ, Sberro H, Ivanova NN, et al. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol.* 2021;6:960–70 (Nature Publishing Group).
60. Mousa WK, Chehadeh F, Husband S. Recent advances in understanding the structure and function of the human microbiome. *Front Microbiol.* 2022;13:825338.
61. Yadav M, Verma MK, Chauhan NS. A review of metabolic potential of human gut microbiome in human nutrition. *Arch Microbiol.* 2018;200:203–17.
62. Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome.* 2020;8:103.
63. Zheng D, Liwinski T, Elinav E. Interaction between microbiota and immunity in health and disease. *Cell Res.* 2020;30:492–506 (Nature Publishing Group).
64. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell.* 2014;157:121–41 (Elsevier).
65. Zeng MY, Inohara N, Nuñez G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* 2017;10:18–26 (Nature Publishing Group).
66. Ang Z, Ding JL. GPR41 and GPR43 in obesity and inflammation – protective or causative? *Front Immunol.* 2016 [cited 2022 Jul 11];7. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2016.00028>
67. Louis P, Flint HJ. Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol.* 2017;19:29–41.
68. Parada Venegas D, De la Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front Immunol.* 2019 [cited 2022 Jul 11];10. Available from: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00277>
69. Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. *FEMS Microbiol Lett.* 2009;294:1–8.
70. Kim Y-G, Sakamoto K, Seo S-U, Pickard JM, Gilliland MG, Pudlo NA, et al. Neonatal acquisition of Clostridia species protects against colonization by bacterial pathogens. *Science.* 2017;356:315–9 (American Association for the Advancement of Science).
71. Sinha SR, Haileselassie Y, Nguyen LP, Tropini C, Wang M, Becker LS, et al. Dysbiosis-induced secondary bile acid deficiency promotes intestinal inflammation. *Cell Host Microbe.* 2020;27:659-670.e5.
72. Guziar DV, Quinn RA. Review: microbial transformations of human bile acids. *Microbiome.* 2021;9:140.
73. Qin X, Yang M, Cai H, Liu Y, Gorris L, Aslam MZ, et al. Antibiotic resistance of Salmonella typhimurium monophasic variant 1,4,[5],12:i:-in China: a systematic review and meta-analysis. *Antibiotics.* 2022;11:532 (Multidisciplinary Digital Publishing Institute).
74. Sabag-Daigle A, Blunk HM, Gonzalez JF, Steidley BL, Boyaka PN, Ahmer BMM. Use of attenuated but metabolically competent Salmonella as a probiotic to prevent or treat Salmonella infection. *Infect Immun.* 2016;84:2131–40 (American Society for Microbiology Journals).
75. Ouyang J, Lin J, Isnard S, Fombuena B, Peng X, Marette A, et al. The Bacterium Akkermansia muciniphila: a sentinel for gut permeability and its relevance to HIV-related inflammation. *Front Immunol.* 2020;11:645.
76. Zhang T, Li Q, Cheng L, Buch H, Zhang F. Akkermansia muciniphila is a promising probiotic. *Microb Biotechnol.* 2019;12:1109–25.
77. Ganesh BP, Klopffleisch R, Loh G, Blaut M. Commensal Akkermansia muciniphila exacerbates gut inflammation in Salmonella typhimurium-infected gnotobiotic mice. *PLoS ONE.* 2013;8:e74963.
78. Bansal T, Alaniz RC, Wood TK, Jayaraman A. The bacterial signal indole increases epithelial-cell tight-junction resistance and attenuates indicators of inflammation. *Proc Natl Acad Sci U S A.* 2010;107:228–33.
79. Rautio M, Eerola E, Väisänen-Tunkelrott M-L, Molitoris D, Lawson P, Collins MD, et al. Reclassification of Bacteroides putredinis (Weinberg et al., 1937) in a New Genus Alistipes gen. nov., as Alistipes putredinis comb. nov., and Description of Alistipes finegoldii sp. nov., from Human Sources. *Syst Appl Microbiol.* 2003;26:182–8.
80. Parker BJ, Wearsch PA, Veloo ACM, Rodriguez-Palacios A. The Genus Alistipes: gut bacteria with emerging implications to inflammation, cancer, and mental health. *Front Immunol.* 2020;11:906.

81. Chng KR, Ghosh TS, Tan YH, Nandi T, Lee IR, Ng AHQ, et al. Metagenome-wide association analysis identifies microbial determinants of post-antibiotic ecological recovery in the gut. *Nat Ecol Evol*. 2020;4:1256–67.
82. Pickard JM, Chervovskiy AV. Intestinal fucose as a mediator of host-microbe symbiosis. *J Immunol Baltim Md*. 1950;2015(194):5588–93.
83. Mistou M-Y, Sutcliffe IC, van Sorge NM. Bacterial glycobiology: rhamnose-containing cell wall polysaccharides in Gram-positive bacteria. *FEMS Microbiol Rev*. 2016;40:464–79.
84. Singh P, Teal TK, Marsh TL, Tiedje JM, Mosci R, Jernigan K, et al. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome*. 2015;3:45.
85. De Keersmaecker SCJ, Verhoeven TLA, Desair J, Marchal K, Vanderleyden J, Nagy I. Strong antimicrobial activity of *Lactobacillus rhamnosus* GG against *Salmonella typhimurium* is due to accumulation of lactic acid. *FEMS Microbiol Lett*. 2006;259:89–96.
86. Jiang H, Li P, Gu Q. Heterologous expression and purification of plantaricin NC8, a two-peptide bacteriocin against *Salmonella* spp. from *Lactobacillus plantarum* ZJ316. *Protein Expr Purif*. 2016;127:28–34.
87. Perumal V, Venkatesan A. Antimicrobial, cytotoxic effect and purification of bacteriocin from vancomycin susceptible *Enterococcus faecalis* and its safety evaluation for probiotization. *LWT*. 2017;78:303–10.
88. Fijan S. Microorganisms with claimed probiotic properties: an overview of recent literature. *Int J Environ Res Public Health*. 2014;11:4745–67.
89. Hanchi H, Mottawea W, Sebei K, Hammami R. The genus *Enterococcus*: between probiotic potential and safety concerns—an update. *Front Microbiol*. 2018 [cited 2022 Aug 15];9. Available from: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01791>
90. Antoun M, Hattab Y, Akhrass F-A, Hamilton LD. Uncommon pathogen, *Lactobacillus*, causing infective endocarditis: case report and review. *Case Rep Infect Dis*. 2020;2020:e8833948 (Hindawi).
91. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci*. 2011;108:4516–22 (Proceedings of the National Academy of Sciences).
92. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol*. 2019;37:852–7.
93. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
94. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590–6.
95. Babraham Bioinformatics - FastQC A quality control tool for high throughput sequence data. [cited 2021 May 10]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
96. 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*. 2015;31:1674–6.
97. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.
98. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
99. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019 [cited 2019 Nov 18];7. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6662567/>
100. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Hancock J, editor. Bioinformatics*. 2019;btz848.
101. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11:2864–8.
102. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28:2520–2.
103. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017;35:1026–8 (Nature Publishing Group).
104. Seemann T. Barrnap. 2022 [cited 2022 Aug 16]. Available from: <https://github.com/tseemann/barrnap>
105. Smid M, Coebergh van den Braak RJJ, van de Werken HJG, van Riet J, van Galen A, de Weerd V, et al. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics*. 2018;19:236.
106. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
107. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
108. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*. 2018;9:5114 (Nature Publishing Group).
109. Barco RA, Garrity GM, Scott JJ, Amend JP, Nealson KH, Emerson D. A genus definition for bacteria and archaea based on a standard genome relatedness index. *mBio*. 2020;11:e02475-19.
110. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2012;40:W445–51.
111. Eddy SR. Accelerated profile HMM searches. *PLOS Comput Biol*. 2011;7:e1002195 (Public Library of Science).
112. Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*. 2021;9:37.
113. Bolduc B, Roux S. Clustering viral genomes in iVirus. *protocols.io*. 2017 [cited 2022 Aug 15]. Available from: <https://www.protocols.io/view/clustering-viral-genomes-in-ivirus-gwebxbe>
114. Nayfach S, Camargo AP, Schulz F, et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol*. 2021;39:578–85. <https://doi.org/10.1038/s41587-020-00774-7>.
115. Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat Biotechnol*. 2019;37:632–9 (Nature Publishing Group).
116. Merchant N, Lyons E, Goff S, Vaughn M, Ware D, Micklos D, et al. The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. *PLOS Biol*. 2016;14:e1002342.
117. Skennerton CT, Imelfort M, Tyson GW. Crass: identification and reconstruction of CRISPR from unassembled metagenomic data. *Nucleic Acids Res*. 2013;41: e105.
118. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. Alignment-free S_d2^Δ*\$ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res*. 2017;45:39–53.
119. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci*. 2003;14:927–30.
120. Create Elegant Data Visualisations Using the Grammar of Graphics. [cited 2022 Aug 12]. Available from: <https://ggplot2.tidyverse.org/>
121. Anderson MJ. Analysis of Ecological Communities: Bruce McCune and James B. Grace, MjM Software Design, Gleneden Beach, USA, 2002, ISBN 0 9721290 0 6, US\$ 35 (Pbk). *J Exp Mar Biol Ecol*. 2003;289:303–5.
122. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60.

Publisher's Note

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