



Divergence of gut bacteria through the selection of genomic variants implicated in the metabolism of sugars, amino acids, and purines by small extracellular vesicles in milk

Fang Zhou, Peerzada Tajamul Mumtaz, Haluk Dogan, Roland Madadjim, Juan Cui & Janos Zempleni

To cite this article: Fang Zhou, Peerzada Tajamul Mumtaz, Haluk Dogan, Roland Madadjim, Juan Cui & Janos Zempleni (2025) Divergence of gut bacteria through the selection of genomic variants implicated in the metabolism of sugars, amino acids, and purines by small extracellular vesicles in milk, Gut Microbes, 17:1, 2449704, DOI: [10.1080/19490976.2025.2449704](https://doi.org/10.1080/19490976.2025.2449704)

To link to this article: <https://doi.org/10.1080/19490976.2025.2449704>



© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 06 Jan 2025.



[Submit your article to this journal](#)



Article views: 2200



[View related articles](#)



[View Crossmark data](#)

RESEARCH PAPER



Divergence of gut bacteria through the selection of genomic variants implicated in the metabolism of sugars, amino acids, and purines by small extracellular vesicles in milk

Fang Zhou^a, Peerzada Tajamul Mumtaz^a, Haluk Dogan^b, Roland Madadjim^b, Juan Cui^b, and Janos Zempleni^a

^aDepartment of Nutrition and Health Sciences, University of Nebraska-Lincoln, Lincoln, NE, USA; ^bSchool of Computing, University of Nebraska-Lincoln, Lincoln, NE, USA

ABSTRACT

Here, we report that small extracellular vesicles (sEVs) in milk mediate the communication between bacteria and animal kingdoms, increase the divergence of bacteria in the intestine, and alter metabolite production by bacteria. We show that bovine milk sEVs select approximately 55,000 genomic variants in 19 species of bacteria from the murine cecum *ex vivo*. The genomic variants are transcribed into mRNA. The selection of genomic variants by milk sEVs alters bacterial metabolism, leading to an up to 12-fold difference in the abundance of more than 1000 metabolites in bacteria cultured in milk sEV-free media compared to sEV-containing media. Evidence is particularly strong that selection of genomic variants by milk sEV changes the metabolism of sugars, amino acids, and purines which might contribute to the development of spatial learning and memory deficiencies and seizure phenotypes reported for milk sEV-depleted infants and mice. Human milk is a rich source of sEVs, whereas formula contains only trace amounts of milk sEVs. This report implicates nutritional sEVs in altered microbial metabolism beyond the mere selection of bacterial communities.

ARTICLE HISTORY

Received 6 May 2024
Revised 20 December 2024
Accepted 31 December 2024



KEYWORDS


Bacteria; exosomes;
extracellular vesicles;
genomic variants; milk

Introduction

Small extracellular vesicles (sEVs) facilitate cell-to-cell communication and cell-to-environment communication in mammals and gram-negative and -positive bacteria, respectively.^{1–3} In most cases, communication is facilitated by the transfer of regulatory cargo such as small RNAs from sEV donor cells to recipient cells.^{4,5} sEVs and small RNAs do not originate exclusively in endogenous synthesis but are abundant in the diet such as milk.^{6,7} Human milk contains 2.2×10^{11} sEVs/mL, which harbor more than 200 distinct microRNAs.⁷ An infant consuming 800 mL milk/d⁸ ingests 176 trillion milk sEVs (mEVs) per day. In contrast, infant formulas contain only trace amounts of mEVs and microRNA cargo.^{7,9} Nature has designed milk to be the sole source of nutrition in the early stages of mammalian life. Consistent with an important role of mEVs in the health and development of mammalian offspring, dietary mEV depletion elicits strong phenotypes in infants and neonate mice,

including a ninefold decrease in the performance of the Barnes maze test of spatial learning and memory, a fivefold increase in the severity of kainic acid-induced seizures, a 50% decrease in litter size, and a 75% decrease in survival to weaning in mice fed a mEV-depleted diet compared to controls fed a mEV-sufficient diet, as well as changes in the urinary metabolome in formula-fed compared to breastfed infants.^{10–12} Evidence from clinical and epidemiological studies is similar and suggests that dietary depletion of mEVs in formula-fed infants decreases neuronal myelination and brain white matter, as well as general, verbal, and non-verbal cognitive abilities.¹³ mEV supplementation improves intestinal health. For example, supplementation with mEVs decreased the severity of necrotizing enterocolitis by 50% in neonate mice.¹⁴ These are important considerations in infant nutrition because only 26% of parents in the United States follow the recommendation by the American Academy of Pediatrics to use human

CONTACT Janos Zempleni  jzemleni2@unl.edu  Department of Nutrition and Health Sciences, University of Nebraska-Lincoln, 1700 N. 35th St., Lincoln, NE 68583, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19490976.2025.2449704>

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

milk as the sole source of nutrition in the first 6 months of life.^{15,16} Infants exclusively fed with formula do not realize the benefits of mEVs conferred by breastfeeding.

The gut microbiome may play an important role in transmitting sEV signals in milk to the host. Approximately, 50% of orally administered mEVs escape absorption, survive digestion in the gastrointestinal tract, and reach the large intestine where most of the gut microbiome resides.^{9,17,18} Bacteria internalize mEVs (this study). Consumption of mEV-defined diets is associated with changes in bacterial communities in the murine gut,¹⁹ and bacterial communities and biosynthetic pathways are different in preterm infants fed mothers' own milk compared to formula-fed preterm infants.²⁰ It is plausible that mEV-dependent changes in bacterial communities in the gut alter both the composition and quantity of metabolites secreted by bacteria and absorbed by the host, a phenomenon referred to as signalome.²¹ Here, we report that the effects of mEVs on the gut microbiome extend beyond the mere selection of bacterial communities, and mEVs select 55,000 genomic variants in 137 genes, including 13 genes in purine metabolism, from 19 bacterial species in cultures of murine cecum extracts compared to reference genomes. The variant genes are transcribed into mRNA, leading to changes in bacterial metabolism. We further report that mEVs elicit an up to 12-fold change in the production of intermediates from purine metabolism, glycolysis and gluconeogenesis, pentose metabolism, and tricarboxylic acid cycle in murine gut bacteria *ex vivo*. We propose that studies of diet-microbiome interactions need to include the selection of genomic variants because the selection of variants drives bacterial divergence and speciation and may affect the health of the host.

Studies were conducted *ex vivo*, as opposed to *in vivo*, based on the following rationale. The data reported here were first generated using samples from mice in a previous study, in which we assessed the effects of mEVs on bacterial communities in the murine gut *in vivo*.¹⁹ Second, conducting the study *ex vivo* allowed us to tightly control mEV levels in media, which mimicked concentrations in the gut's aqueous-phase water. Third, we have used monocultures of commensal gut pathogens (*Clostridioides difficile* and

Enterococcus faecalis) to demonstrate that mEVs select genetic mutations *ex vivo* that make the pathogens less virulent *in vivo*, which speaks to the power of the *ex vivo* approach.²²

Results

Bioinformatics pipelines

We used two complementary bioinformatics pipelines to assess the effects of mEVs on the selection of genomic variants. The MIDAS (Metagenomic Intra-Species Diversity Analysis System) pipeline maps sequencing reads by using 30,000 bacterial reference genomes and reveals the landscape of variants selected by mEVs in species and strains.²³ The StrainPhlAn pipeline maps sequencing reads by using well-annotated clade-specific marker genes from over ~17,000 bacterial, viral, and eukaryotic reference genomes and allows for a higher level of resolution compared to MIDAS at the strain level.²⁴ Analyses were performed by using an average of approximately 130 million raw DNA sequencing reads from each cecum sample cultured in mEV-supplemented (mEVS) and mEV-free (mEVF) media for a total of 400 million reads per treatment group (Supplementary Table S1). After filtering, 367,237,399 and 395,031,435 sequencing reads remained for downstream analysis of mEVS and mEVF cultures, respectively.

Bovine milk EVs participate in the selection of genomic variants and population structures in mixed cultures of bacteria from the murine cecum

Analysis by MIDAS suggests that bovine milk EVs select genomic variants, including single nucleotide polymorphisms (SNPs), insertions and deletions in anaerobic cultures of cecum content from C57BL/6J mice. Cecum content from the same three mice was aliquoted and added to mEVS or mEVF media to eliminate variations in the input material as confounder. We analyzed DNA-sequencing data by using two levels of stringency, accepting genomic variants that were detected in either all three biological replicates (high stringency) or two replicates (lower stringency).

More than 200 and 190 million sequencing reads were mapped to 11 and 19 bacterial species in

Table 1. Bacterial species in mEVS and mEVF cultures.

Variable	mEVS ^a	mEVF
Annotated sequencing reads ^b	207,380,409	191,902,964
Species detected only in mEVS or mEVF ^c	<i>Anaerotruncus</i> sp_59696 <i>Clostridium</i> sp_57418 <i>Clostridium drakei</i> _61273 <i>Desulfotomaculum ruminis</i> _62150	<i>Aneurinibacillus aneurinilyticus</i> _62028 <i>Brevundimonas diminuta</i> _58385 <i>Clostridium</i> sp_60561 <i>Lactobacillus johnsonii</i> _57663 <i>Paenibacillus ehimensis</i> _60544 <i>Bacillus cereus</i> _56036 <i>Lysinibacillus boronitolerans</i> _54563 <i>Sporosarcina newyorkensis</i> _62264 <i>Clostridium termitidis</i> _57708 <i>Lysinibacillus fusiformis</i> _53142 <i>Staphylococcus xylosus</i> _56919 <i>Paenibacillus barengoltzii</i> _56244
Species detected in both mEVS and mEVF	<i>Bacteroides faecis</i> _58503 <i>Bacteroides thetaiotaomicron</i> _56941 <i>Burkholderiales bacterium</i> _56577 <i>Clostridium celerecrescens</i> _61145 <i>Clostridium</i> sp_60041 <i>Clostridium sporogenes</i> _58038 <i>Enterococcus faecalis</i> _56297	

^amEVF, milk extracellular vesicle-free media; mEVS, milk extracellular vesicle-supplemented media.

^bValues represent the combined total of annotated reads from all three sequencing experiments per treatment. Annotation was achieved by aligning sequencing reads with more than 30,000 bacterial reference genomes in MIDAS.

^cSpecies were detected by using MIDAS ($n = 3$ per group). Source data are provided as a Source Data file.

mEVS and mEVF cultures, respectively, by using a threshold of 10× sequencing coverage (Table 1). Seven species were detected in both mEVS and mEVF cultures, four species were unique to mEVS cultures, and 12 species were unique to mEVF cultures. For three of the seven species present in both mEVS and mEVF cultures, their relative abundance differed between cultures: *Clostridium celerecrescens*_61145, *Enterococcus faecalis*_56297, and *Burkholderiales bacterium*_56577 ($p < 0.05$; Figure 1). *C. celerecrescens* was a dominant bacterium because it serves as a reference strain in the MIDAS and StainPhlAn

pipelines used in bioinformatics analysis but also represents other Clostridia.

In mEVS cultures, 278 and 28,594 strain-level genomic variants were detected in high and low stringency datasets, respectively (Figure 2a). Ninety-five genes carried non-synonymous SNPs in mEVS cultures in the high stringency dataset. In mEVF cultures, 92 and 26,382 strain-level genomic variants were detected in high and low stringency datasets, respectively (Figure 2b). Forty-two genes carried non-synonymous SNPs in mEVF cultures in the high stringency dataset. Genomic variants were slightly underrepresented in coding sequences (CDS)

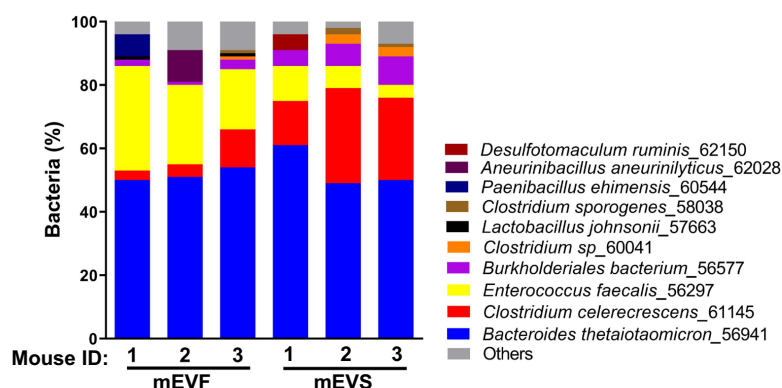


Figure 1. Percent distribution of bacterial species cultured in media defined by the content of milk extracellular vesicles. Gut content was collected from ceca of C57BL/6J mice and cultured in sterile M9 minimal salts media supplemented with milk EVs (mEVS) or vehicle (mEVF) under anaerobic conditions for 7 d. Bars represent the percent distribution of bacterial species from three mice per treatment, analyzed by MIDAS. Source data are provided as a Source Data file.

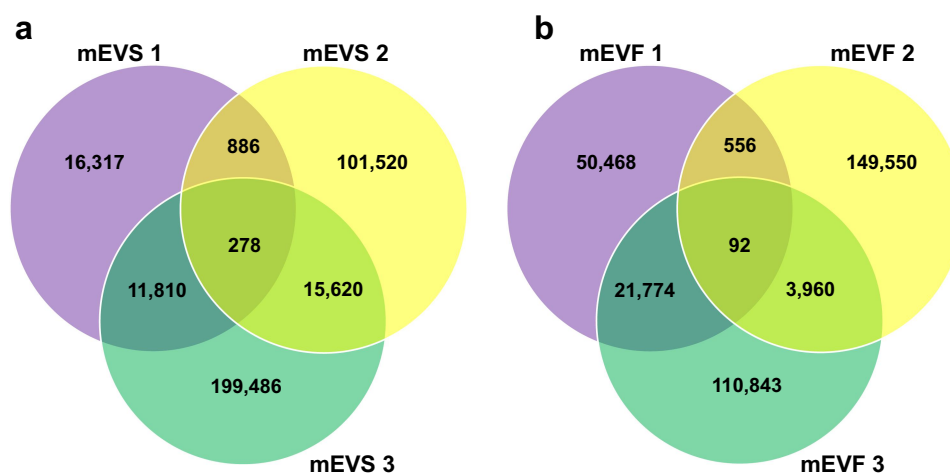


Figure 2. Venn diagrams of genomic variants in bacteria cultured in media defined by the content of milk extracellular vesicles. a) Genomic variants in mEVS cultures. b) Genomic variants in mEVF cultures. Gut content was collected from ceca of C57BL/6J mice and cultured in mEVS or mEVF under anaerobic conditions for 7 d. Genomic variants were analyzed by using MIDAS ($n = 3$ mice per group). Aliquots of cecum content from the same mice were used to inoculate mEVS and mEVF cultures, and the numerals after group designations identify the mice. Source data are provided as a Source Data file.

Table 2. Genomic variations in CDS or IGR in mEVS or mEVF cultures of murine cecum content.

	mEVS ^a		mEVF	
	CDS	IGR	CDS	IGR
<i>B. thetaiotaomicron</i> _56941	2490 ± 239	2475 ± 158	2310 ± 192	2320 ± 88
<i>B. bacterium</i> _56577	15,526 ± 2378*	2230 ± 369*	5315 ± 1012	855 ± 144
<i>C. celerecrescens</i> _61145	174,986 ± 9828	22,114 ± 740	185,590 ± 17,421	21,559 ± 1213
<i>E. faecalis</i> _56297	21,868 ± 1190*	2919 ± 131*	3290 ± 1972	1408 ± 247

^aCDS, coding sequences; IGR, intergenic regions; mEVF, milk extracellular vesicle-free media; mEVS, milk extracellular vesicle-supplemented media.

* $p < 0.05$ (mEVS vs mEVF; $n = 3$). Source data are provided as a Source Data file.

compared to other regions in mEVF cultures of *B. bacterium*_56577 and *E. faecalis*_56297 (Table 2): CDS make up 93% and 90% of the genomes in *B. bacterium*_56577 and *E. faecalis*_56297, respectively, but only 86% and 70% of the genomic variants resided in CDS. In contrast, 87% and 88% of genomic variants resided in CDS in mEVS cultures, and 86% and 70% genomic variants resided in CDS in mEVF cultures of *B. bacterium*_56577 and *E. faecalis*_56297, respectively. In *C. celerecrescens*_61145, 90% and 95% of the genomic variants in mEVF and mEVS cultures, respectively, resided in CDS. Sizes of CDS and whole genomes were accessed through the Integrated Microbial Genomes & Microbiomes system.^{25,26} In some species, the frequency of genomic variants was higher in mEVS cultures compared to mEVF cultures. For example, the number of variants was greater in mEVS cultures of *B. bacterium*_56577 and *E. faecalis*_56297 compared to mEVF cultures (Table 2). The effect of bovine mEVs was particularly

pronounced in *E. faecalis*_56297 in which the number of variants was five times greater in mEVS cultures compared to mEVF cultures. In some species, genomic variants clustered in intergenic regions (IGRs). For example, 50% of the genomic variants clustered in IGRs in *Bacteroides thetaiotaomicron*_56941 in both mEVS and mEVF cultures, despite IGRs making up only 2% of the genome in this species (Table 2).

Bovine mEVs selected genomic variants in distinct metabolic pathways. Eighteen of the 128 mEV-dependent variants are implicated in the metabolism of sugars, amino acids, and purines in both mEVS and mEVF cultures (Table 3 and Supplementary Table S2).

Bovine mEVs altered the relative copy number (RCN) of genes. At the level of high stringency, RCNs of four and six genes were higher in *B. thetaiotaomicron*_56941 and *B. bacterium*_56577, respectively, cultured in mEVF compared to mEVS media (Supplementary

Table 3. Genes with non-synonymous substitutions in *C. celerecrescens*^a

	Gene ID ^b	Pathway
mEVS ^c	29354.3.peg.1282	Starch and sucrose metabolism
	29354.3.peg.42; 29354.3.peg.643	Amino sugar and nucleotide sugar metabolism
	29354.3.peg.4350; 29354.3.peg.858	Glycine, serine and threonine metabolism
	29354.3.peg.64	Arginine and proline metabolism
	29354.3.peg.2509	D-Glutamine and D-glutamate metabolism
	29354.3.peg.2075; 29354.3.peg.2043; 29354.3.peg.3018	Porphyrin and chlorophyll metabolism
	29354.3.peg.292	Pantothenate and CoA biosynthesis
	29354.3.peg.2621	Oxidative phosphorylation
	29354.3.peg.1970	Peptidoglycan biosynthesis
	29354.3.peg.20	Purine metabolism
mEVF	29354.3.peg.1435	Tryptophan metabolism
	29354.3.peg.59	Arginine and proline metabolism
	29354.3.peg.1765	Glycerophospholipid metabolism
	29354.3.peg.1203	Glycolysis and gluconeogenesis

^aNon-synonymous substitutions were detected in all three independent cultures.

^bIDs as per the Pathosystems Resource Integration Center 3.6.8 (<https://www.patricbrc.org/>).

^cmEVF, milk extracellular vesicle-free media; mEVS, milk extracellular vesicle-supplemented media.

Source data are provided as a Source Data file.

Figure. S1a). In contrast, RCNs of three genes were lower in *E. faecalis*_56297 cultured in mEVF compared to mEVS. At the level of low stringency, additional RCNs were detected in two biological replicates. RCNs of 17 genes were higher in *Clostridium* sp_60041 cultured in mEVF compared to mEVS when using a twofold change as cutoff (Supplementary Figure. S1b). In contrast, the RCNs of 228 genes were higher in *Clostridium* sp_60041 cultured in mEVS compared to mEVF (see Supplementary Figure. S1 for representative examples). Collectively, bovine mEVs selected both bacterial communities and genetic variants in the murine gut microbiome.

Bovine mEVs select genomic variants in marker genes in bacterial strains

Variants were also identified when focusing on marker genes using the StrainPhlAn pipeline.²⁴ Eight bacterial strains were detected by using the StrainPhlAn pipeline, and seven of the strains matched those identified by using MIDAS (Supplementary Table S3). The only exception is

Oscillibacter sp. 1–3, which was unique to the StrainPhlAn pipeline. Three of the eight species satisfied the recommended minimum of detecting 50% of the marker genes in their clades in StrainPhlAn analysis. We detected a total of 6,715 genomic variants across all loci in the three species for both mEVS and mEVF cultures combined, including 6,694 variants in protein coding regions [5,182 non-synonymous (77%) and 1,512 synonymous (23%) variants]. The non-synonymous variants included 62 insertions and 75 deletions. Non-synonymous variants occurred more frequently in *Clostridium sporogenes* than in other species (Supplementary Table S4). Divergence of species was observed in phylogenomic trees in both mEVS and mEVF cultures albeit with a species-specific bias toward one of the culture conditions. For example, genomic divergence was greater in mEVF compared to mEVS cultures in *E. faecalis* and *C. sporogenes* (Figure 3; Supplementary Figure. S2a, b), whereas divergence of *Lactobacillus johnsonii* was observed only in mEVF cultures (Supplementary Figure. S2c, d). The number of genomic variants was sufficiently high to allow for a phylogenomic analysis in only four out of six samples each for *C. sporogenes* and *L. johnsonii*.

Some strains and genomic loci were more susceptible than others to accumulating genomic variants in protein-coding genes. Genomic variants were identified in 56 protein-coding genes in *C. sporogenes* (40 unique genes), *L. johnsonii* (6) and *E. faecalis* (10) (Supplementary Table S5). Five hundred-eighteen out of 1007 variants in *C. sporogenes* resided in only three genes: a gene coding for a protein of unknown function containing a DUF5050 domain (DUF5050 domain-containing protein), a gene coding for a protein containing a HAMP domain (HAMP domain-containing protein), and a gene coding for a protein containing an anti-sigma factor domain (anti-sigma factor domain-containing protein) (Figure 4a). In *L. johnsonii*, 58 genomic variants (94% of total) were identified in genes coding for a member of the superfamily of ATPases associated with diverse cellular functions (AAA family ATPase), a protein from choloylglycine hydrolase (choloylglycine hydrolase family protein) and a hypothetical protein (Figure 4b). In *E. faecalis*, genomic variants were detected in 10 protein-

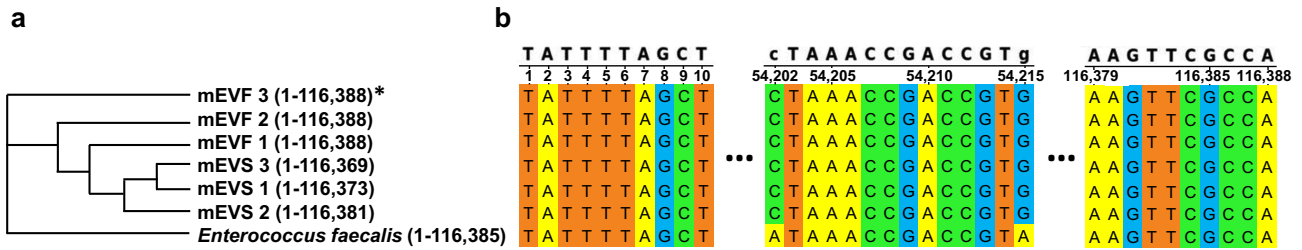


Figure 3. Genomic variants in *E. faecalis* cultured in media defined by the content of milk extracellular vesicles. a) Phylogenomic tree of marker genes in *E. faecalis*. b) Multiple sequence alignment of concatenated *E. faecalis* markers from mEVS and mEVF cultures with the *E. faecalis* reference genome (RefSeq: GCF_902161805.1). Gut content was collected from ceca of C57BL/6J mice and cultured in mEVS or mEVF under anaerobic conditions for 7 d ($n = 3$ mice per group). The phylogenomic tree was built using the RAXML algorithm. *Numbers in parentheses identify the length of the concatenated markers. Sequencing reads of marker genes were aligned with the reference sequence in the MetaPhlAn2 database. Source data are provided as a Source Data file.

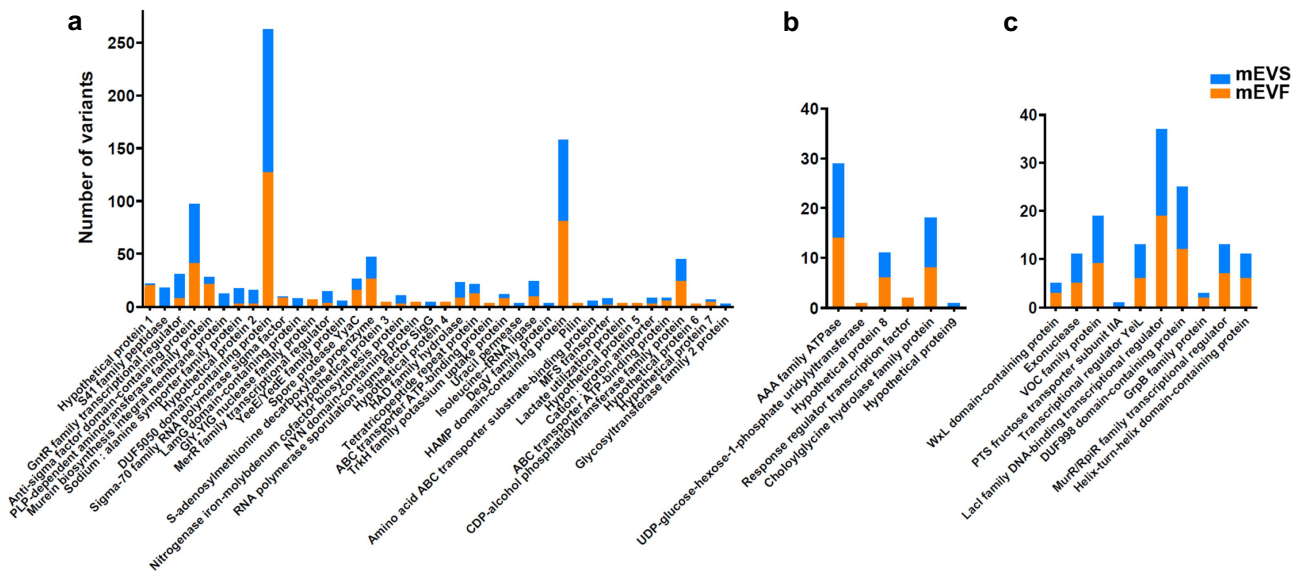


Figure 4. Number of genomic variants in strains from three bacterial species strains cultured in media defined by the content of milk extracellular vesicles. a) *Clostridium sporogenes*. b) *Lactobacillus johnsonii*. c) *Enterococcus faecalis*. Gut content was collected from ceca of C57BL/6J mice and cultured in mEVS or mEVF media under anaerobic conditions for 7 d ($n = 3$ mice per group). Blue bars represent the number of variants in mEVS; orange bars represent the number of variants in mEVF. Source data are provided as a Source Data file.

coding genes. Fifty-nine percent of the genomic variants in *E. faecalis* were identified in three genes (Figure 4c): a gene coding for a protein of the lactose operon repressor (LacI) family of transcriptional regulators (LacI family DNA-binding transcriptional regulator), a gene coding for a protein of unknown function from a family of conserved archaeal proteins containing a DUF998 domain (DUF998 domain-containing protein), and a gene coding for a protein from the superfamily of vicinal oxygen chelators (VOC family protein). Genomic variants in these genes were detected in both mEVS and mEVF cultures. Collectively, we observed

that some bacterial strains were more resistant to the selection of genetic variants than others, and insertions and deletions were rare events.

In some genes, variants were detected in only one of the two culture conditions. For example, seven and nine genes genomic variants were detected only in mEVS and mEVF cultures, respectively, in *C. sporogenes* (Supplementary Table S5). All 13 variants in the gene coding for murJ (murein biosynthesis integral membrane protein) were identified in mEVS cultures in *C. sporogenes*. Variants in genes coding for a type II secretion system protein (pilin) and a protein related to the utilization of lactate

(lactate utilization protein) only were detected in mEVF cultures in *C. sporogenes* (Figure 4a). KEGG pathway analysis implicated synonymous and non-synonymous genomic variants in *C. sporogenes* in regions coding for enzymes implicated in purine metabolism and biosynthesis of cofactors, such as nicotinamide adenine dinucleotide (NAD) and pantothenate biosynthesis in mEVS cultures and aminoacyl-tRNA biosynthesis in mEVF cultures (Supplementary Table S6).

Transcription of genomic variants

Forty-eight to 70 million raw reads were obtained for each of 12 samples (two cultures \times two sexes \times 3 replicates) by RNA sequencing analysis using a 100-bp paired-end protocol. For mEVS and mEVF cultures $55,894,867 \pm 6,335,845$ and $61,127,337 \pm 4,910,738$ reads, respectively, passed quality checks (Q score > 10 and length ≥ 15 bp) and were included in subsequent analyses (Supplementary Table S7, Table 1). When the reads were mapped using the MIDAS pipeline, we detected 63,613 and 56,385 RNA-level variants in 9,165 and 6,914 unique transcripts in 12 and 9 bacterial species mEVS and mEVF cultures, respectively. Transcripts from bacteria in female mice harbored significantly less variants in mEVF and mEVS cultures, respectively, than male mice. Three thousand six hundred and thirty-seven and 3,683 transcripts matched synonymous and non-synonymous genomic variants in mEVF and mEVS cultures, respectively, in the low stringency dataset (Supplementary Table S7, Tables 2 and 3). Six hundred-eight and 1,494 transcripts matched synonymous and non-synonymous genomic variants in mEVF and mEVS cultures, respectively, in the high stringency dataset. When only non-synonymous variants were considered, 39 and 215 variant transcripts matched genomic variants in mEVF and mEVS, respectively, in the high stringency dataset. In both mEVS and mEVF cultures, variants clustered in pathways of glucose metabolism, energy homeostasis and metabolism of cofactors (purines) in glucose and energy homeostasis, which accounted for 52% and 63% of all variants detected in mEVS and mEVF cultures, respectively ($p < 0.05$; Supplementary Figure S3, Supplementary Table S7, Tables S4 and S5). Collectively, genetic variants are transcribed and cluster in pathways implicated in major metabolic pathways.

Bacterial metabolome

The selection of genomic variants was associated with changes in the bacterial metabolome. We assessed the bacterial metabolome in both bacterial pellets and cell-free media supernatant by using non-targeted metabolomics analysis. Bacterial cultures from the ceca of both male and female mice were analyzed. We identified a total of 536 and 494 mEV-dependent metabolites in cell pellets and media supernatant, respectively, for both sexes and media combined. Ninety and 147 metabolites were significantly different in cell pellets and media supernatants, respectively, from mEVS and mEVF cultures. The abundance of metabolites was distinct for the two media, as per principal component analysis (PCA) and hierarchical cluster analysis (HCA; Figure 5a-d). Most metabolites were shared across treatment groups and input material (Figure 5e). Exceptions included 136 metabolites that were detected only in cell pellets from mEVS cultures and 94 metabolites that were detected only in media from mEVF cultures. mEV-dependent metabolites were overrepresented in distinct super pathways, including the metabolism of amino acids, lipids, carbohydrates, and nucleotides (Figure 5f, Supplementary Tables S8 and S9).

Genomic variants in the metabolism of purines and sugars

For an in-depth analysis of the association between genomic variants and changes in bacterial metabolism, we focused on pathways implicated in the metabolism of purines and sugars based on the following rationale. First, the dietary intake of milk sEVs altered the levels of purine metabolites in urine and tissues in infants, adults, and mice.²⁷ Second, purine compounds such as ATP, cAMP, GTP, NAD and NADP play crucial roles in energy metabolism and purinergic receptor signaling in the brain.^{28–32} Third, alterations in the metabolism of purines and sugars by gut bacteria under mEV-defined conditions are consistent with a report that spatial learning and memory were impaired and the severity of kainic acid-induced seizures increased in mice fed an mEV-depleted diet compared to controls fed an mEV-sufficient diet.¹⁰ Fourth, the metabolism of sugars and purines

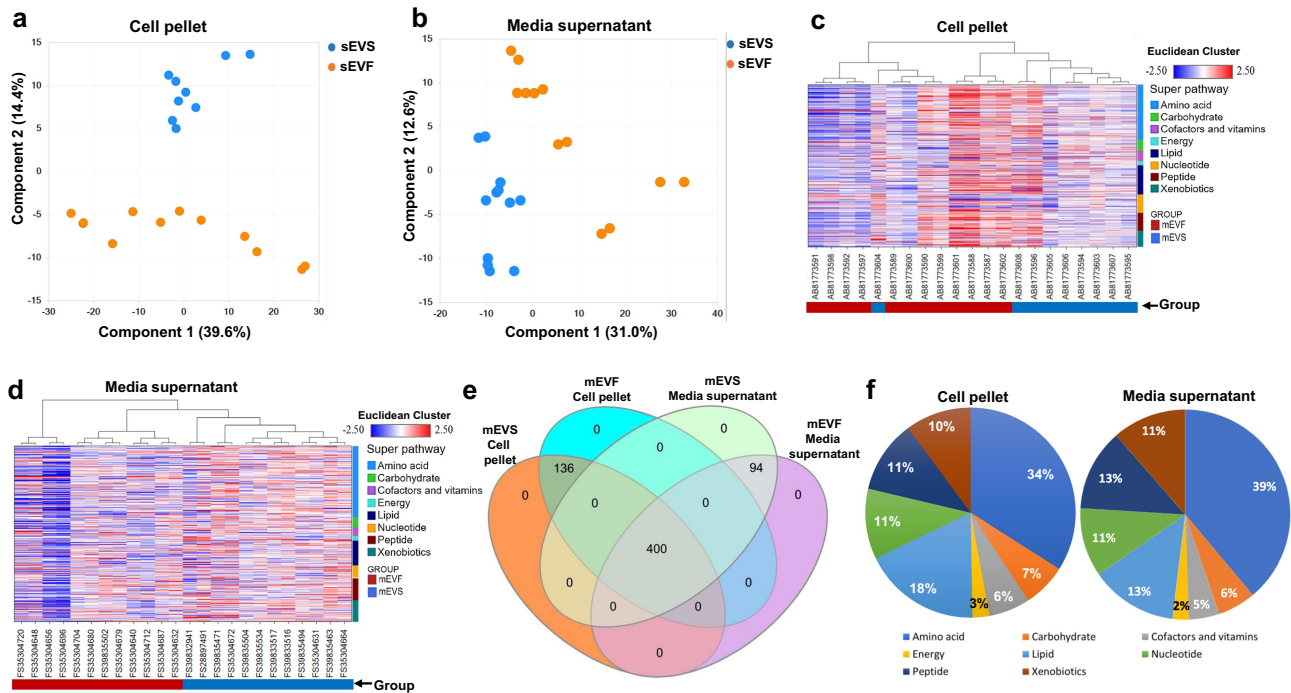


Figure 5. Bacterial metabolites in cultures defined by the content of milk extracellular vesicles. Gut content was collected from ceca of C57BL/6J mice and cultured in mEVs or mEVF media under anaerobic conditions for 7 d ($n = 3$ mice per sex and group). a) Principal component analysis of metabolites in cell pellets. Plots represent the combined total of all metabolites in both male and female mice cultured in mEVs and mEVF media ($n = 12$). b) Principal component analysis of metabolites in cell-free media supernatants. Plots represent the combined total of all metabolites in both male and female mice cultured in mEVs and mEVF media ($n = 12$). c) Hierarchical clustering analysis of metabolites in cell pellets significantly altered by the content of mEVs in bacterial cultures. d) Hierarchical clustering analysis of metabolites in cell-free media supernatant significantly altered by the content of mEVs in bacterial cultures. e) Venn diagram of metabolites significantly altered in bacterial pellets and cell-free media supernatant from mEV-defined bacterial cultures. f) Percent distribution of metabolites significantly altered by mEV content among super pathways in cell pellets and cell-free supernatant.

ranked among the top five mEV-dependent super pathways.

mEV-dependent genomic variants were overrepresented in pathways of purine metabolism. Thirty-nine of the 338 genes harboring 325 genomic variants identified by StrainPhlAn were implicated in purine and nucleic acid metabolism (Supplementary Table S7, Table S5). Genomic variants were more frequent in the following five genes than in any other gene that harbored polymorphisms: DNA polymerase I (EC 2.7.7.7), phosphoribosylformylglycinamide synthase (EC 6.3.5.3), ribonucleoside-triphosphate reductase (EC 1.17.4.2), DNA polymerase III (EC 2.7.7.7), DNA polymerase III polC-type (EC 2.7.7.7) (Supplementary Table S7, Table S5). Six variants of genes in purine metabolism were detected at the genome level, and four variants were detected at the transcript level in *Clostridium celerecrens* 152B and *C. sporogenes*. Three variants of genes

in purine metabolism [adenosine deaminase (*add*), adenine deaminase (*ade*), and N5-carboxyaminoimidazole ribonucleotide mutase (*purE*)] were detected at both the genome and transcript level and represent a particularly high level of confidence.

mEV-dependent changes in genomes and transcriptomes were associated with changes in the metabolome. We identified 26 purine metabolites in bacterial pellets and 21 in media supernatants. However, the effect of mEVs was statistically significant only for allantoin which was 2.2-fold and 2.0 higher in media and cell pellets, respectively, from mEVs cultures compared to mEVF cultures ($p < 0.05$; Figure 6a). Allantoin is an end product in bacterial purine catabolism (Supplementary Figure. S4; Supplementary Table S9). Changes in purine metabolites were associated with changes in metabolites from major pathways of sugar and energy metabolism. We identified 36 and 29 sugar

metabolites and intermediates from energy metabolism in bacterial pellets and media supernatants. The levels of 24 metabolites were up to 3.7-fold different in pellets and media from mEV-defined cultures ($p < 0.05$; Figure 6b; Supplementary Table S9). The compounds are intermediates in glycolysis, gluconeogenesis, Krebs cycle, and pentose phosphate pathway (Supplementary Figure S5).

Bacteria internalize mEVs

Bacteria internalized mEVs, which is consistent with the selection of genomic variants by mEVs. In a first round of experimentation, we added HiLyte 750-labeled mEVs at nutritionally relevant concentrations to cultures of Gram-positive *B. subtilis* and Gram-negative *E. coli*. The internalization of mEVs by bacteria reached a plateau within less than 10 min (Supplementary Figure S6a). However, the data do not allow to distinguish between adsorption of mEVs to the bacterial surface and internalization of mEVs by bacteria. This limitation was addressed in a second round of experimentation, in which we used the expression

of reporter plasmids as readout. mEVs were loaded with plasmids pMRE-Tn7-135 (encoding mScarlet I) or pBAV1K-T5-gfp (encoding EGFP) and added to cultures of *E. coli* and Gram-positive *Bifidobacterium infantis*. The expression of fluorescent proteins was assessed using confocal microscopy and fluorescence activated cell sorting (FACS). We detected mScarlet I and EGFP in both bacterial species (Figure 7; columns a₁ and a₂ denote mScarlet I, and columns a₃ and a₄ denote EGFP). The images depict protein expression with (Figure 7a) and without (Figure 7b) electronically enhanced brightness and contrast. Data from confocal microscopy were independently confirmed using FACS. In FACS analysis, control cells from both bacterial strains formed distinct populations, with forward scatter (FSC) as the discriminant to exclude background noise (Supplementary Figure S6b₁-b₂). Bacteria treated with pBAV1K-T5-gfp – loaded mEVs expressed EGFP, while mScarlet I fluorescence was not detectable (Supplementary Figure S5b₃-b₄). Bacteria treated with pMRE-Tn7-135-loaded mEVs expressed mScarlet I, while EGFP fluorescence was not detectable (Supplementary Figure S6b₅-b₆).

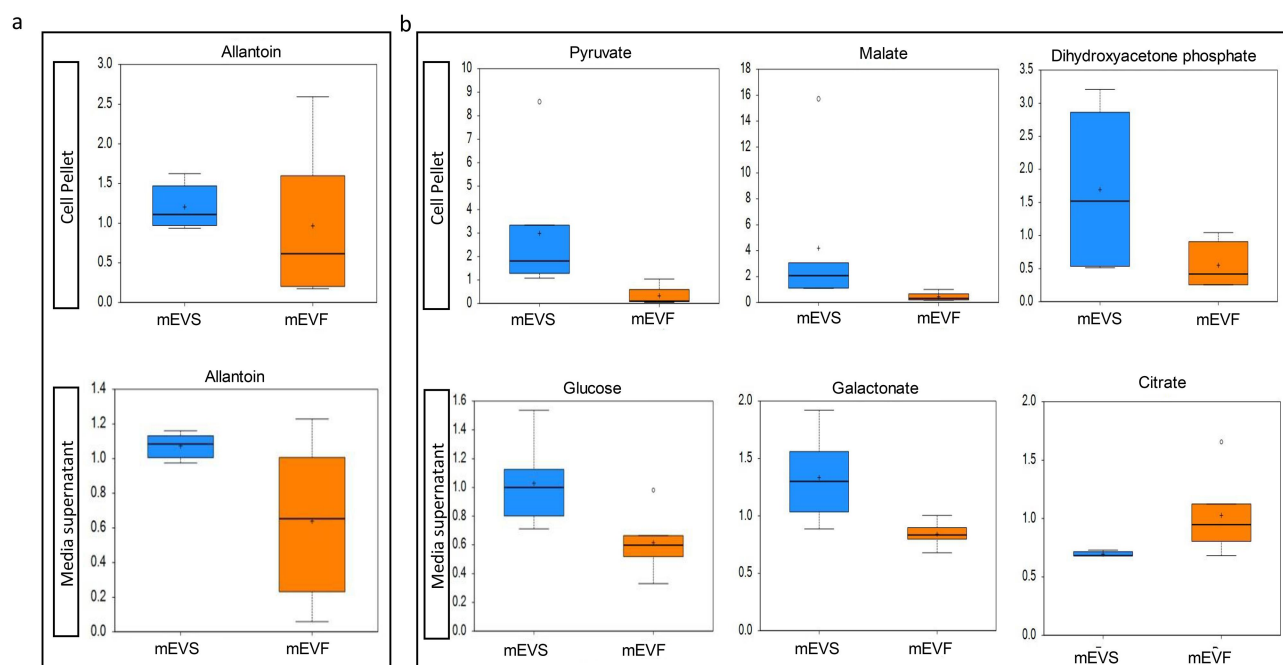


Figure 6. Levels of purine, energy and sugar metabolites in cell pellets and media supernatants from bacterial cultures in media defined by the content of milk extracellular vesicles. a) Levels of significantly altered purine metabolite in cell pellets and media supernatants ($p < 0.05$). b) Levels of significantly altered energy and sugar metabolites in cell pellets and media supernatants ($p < 0.05$). The y-axes are the relative metabolite concentrations. The central horizontal lines in boxes represent median values, and + denotes mean values; the top and bottom of boxes represent upper and lower quartiles. Error bars depict the minimum and maximum distributions, whereas ○ identifies extreme data points.

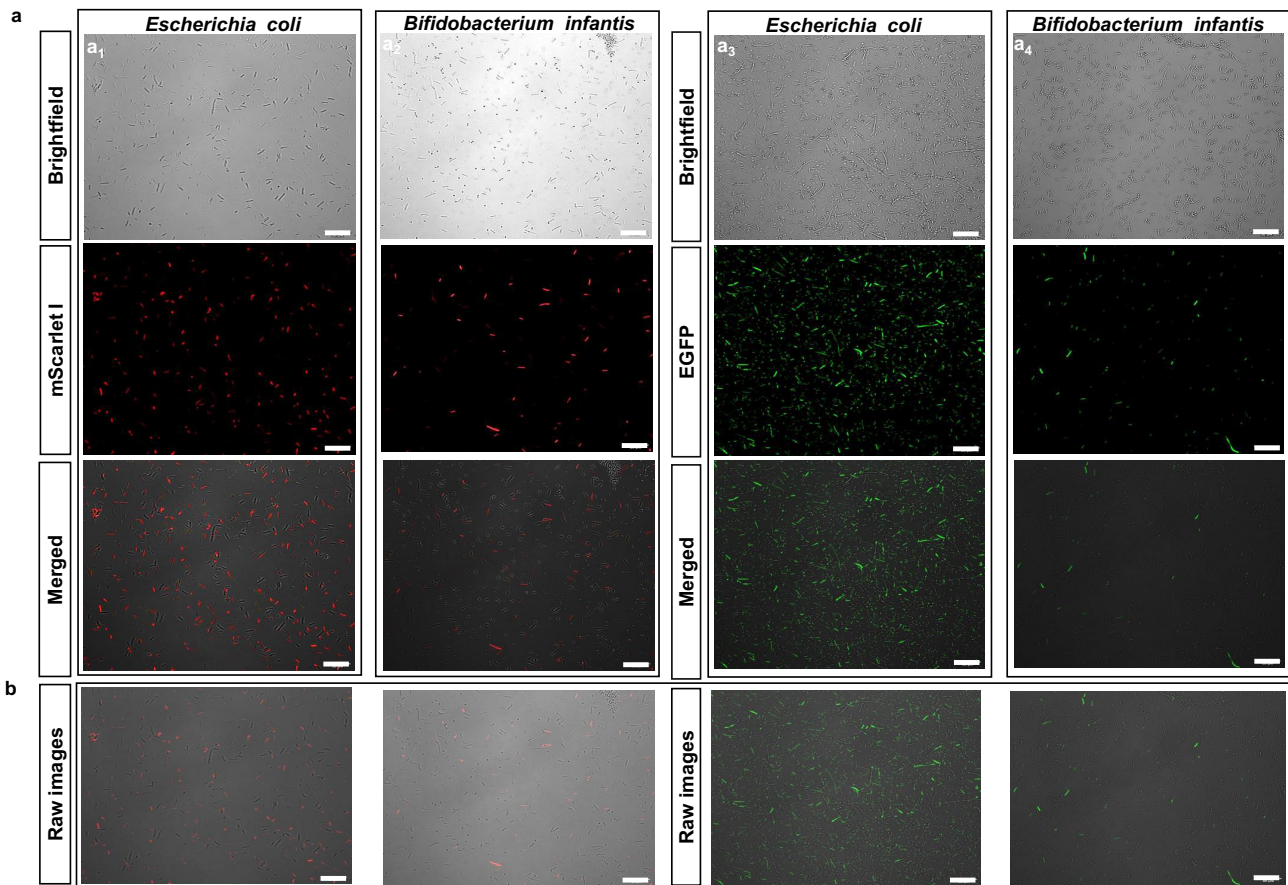


Figure 7. a) Expression of plasmids encoding EGFP and mScarlet-i, loaded into and delivered by mEVs, in *B. infantis* and *E. coli*. The images were captured using laser scanning confocal microscopy 6 h after induction by IPTG and L-arabinose. Each image was captured through a 60× magnification lens. Scale bar: 20 μm. b) Same as panel a, but without enhancing brightness and contrast. It serves as a reference point, allowing for a direct comparison with the enhanced images in panel (a1-4), which underwent specific brightness and contrast enhancements.

Discussion

This report links dietary EVs with the selection of genomic variants in gut bacteria. Previous studies, including ours, were limited in scope by focusing on EV-dependent changes in bacterial communities in the gut.^{19,33} This report is important because it implicates mEVs in the divergence of gut bacteria as a first step in speciation and altered bacterial signalome in the host.^{21,34} There is precedent for diet-driven selection of genomic variants and mutations. For example, the −13910*T mutation in the *lactase* gene leads to persistent expression of lactose in adults and conferred a growth advantage to European populations practicing dairy farming.³⁵ The selection of a mutation in the *lactase* gene by dietary lactose does not distract from this report's novelty because here we show for the first time that a *bona fide* signal transmitter from the animal kingdom, mEVs,

communicates with the bacteria kingdom leading to genomic divergence in bacteria. Both Gram-positive and Gram-negative bacteria communicate with the environment through bacterial EVs.^{36,37} Our observation that bacteria internalize bovine mEVs is consistent with EV-dependent communication in bacteria, but that does not exclude the possibility that some bacteria degrade mEVs and are auxotrophic for the degradation products. The roles of mEVs in bacterial signaling and nutrition are not mutually exclusive.

The selection of genomic variants by mEVs was achieved using a nutritionally relevant concentration of mEVs in culture media within only 7 d of culturing, *i.e.*, this report has implications for human nutrition and health. We previously reported that dietary depletion of mEVs elicits an increase in the concentration of purine metabolites in body fluids and tissues from adults, infants and

mice.²⁷ In the light of the new discoveries reported here, we speculate that altered bacterial metabolism of purines in the gut might have contributed to the increase in purine metabolites observed in humans and mice on mEV-depleted diets. For example, the c.562C>A variation in the *ade* gene (adenine deaminase, EC 3.5.4.2) in *C. celerecrescens* changes an amino acid change in the substrate binding site of the enzyme.³⁸ However, this is an untested hypothesis and requires experimental confirmation, especially when considering that the amino acid substitution constitutes a conservative change (*p. Leu188Ile*). We further speculate that bacterial metabolites such as short chain fatty acids, tryptophan, and bile acid derivatives might play a role in decreasing the severity of necrotizing enterocolitis in mEV-supplemented mice.^{14,39} Potentially rewarding lines of future investigation include changes in antibiotics resistance and virulence in pathogenic bacteria. For example, single point mutation in *Clostridium difficile* ribotype RT027, selected by trehalose, increased the virulence of this commensal gut pathogen.⁴⁰

This study was conducted using mature cow's milk from a local grocery store for preparing mEVs. Using milk from a grocery store has the advantage that milk from many cows is pooled at the producer's end, thereby eliminating intra-individual variation as a confounder. The trade-off is that one does not know breed, parturition, and age of the cow. mEVs from different breeds and lactational stages might have unique effects on the gut microbiome and host. For example, the mEV cargo differed between Holstein and Normande breeds, and colostrum had a strong protective effect against necrotizing enterocolitis than mature milk in intestinal organoids.^{41,42} We propose that mEVs from distinct mammalian species might also have unique effects on the gut microbiome, but this is an untested hypothesis. Studies are ongoing in our laboratory to assess the effects of human mEVs on the communities, genetic variants, and metabolism in bacteria from infant feces. The ongoing studies will enable to assess the relevance of the selection of genomic variants in a population at risk for mEV depletion.

Among human cohorts, this report is relevant in infant nutrition because most infants born in the United States are partially or exclusively formula-

fed, and the formula contains only trace amounts of mEVs.^{7,15,16} This is a concern because consumption of an mEV-depleted diet impaired brain health and development in neonate mice.¹⁰ The significance of this report may extend beyond mEVs and includes the selection of genomic variants by plant foods, which also contain EVs.^{33,43}

We acknowledge the following limitations and uncertainties of this report. While there is good consensus regarding the bacterial species detected by MIDAS and StrainPhlAn, the genomic variants revealed different mEV-dependent pathways when analyzed by the two pipelines. This can be attributed to differences in gene coverage in MIDAS compared to StrainPhlAn.^{23,24} We expect that additional variants will emerge as pipelines for mapping the variants improve. We did not formally assess the mechanisms leading to the divergence in evolving bacterial populations, as well as bacterial communities in distinct regions of the gut. A previous report suggests that multiple mechanisms contribute to diversity in glucose-limited cultures and the capacity of a population to diversify is a function of growth rate.⁴⁴ Ongoing studies focus on mEV-dependent checkpoints of bacterial divergence, associations of genomic variants with changes in transcriptome and metabolome, and phenotypes of bacterial divergence in the host.

Methods

Ethical statement

All animal care procedures and experimental procedures were conducted following the protocols approved (protocol 1229) by the Institutional Animal Care Program at the University of Nebraska-Lincoln.

Mice

C57BL/6J mice aged 7 weeks were purchased from the Jackson Laboratory (stock no. 000664) and acclimated for 7 d (22°C, 12-h light/dark cycle) with free access to food (2016 Teklad global 16% protein rodent diet; Teklad 2916, Envigo, Inc., Madison, WI, USA), and water prior to collection of cecal content.

Collection and culture of cecal content

Mice were euthanized by using carbon dioxide followed by cervical dislocation, and cecal content was immediately suspended in 10 mL of anaerobic sterile M9 minimal salts media (MM; Gibco Life Technologies, Carlsbad, CA, USA) supplemented with 2 mM MgSO_4 , 0.1 mM CaCl_2 and 5 mL of sterile-filtered gut content in MM media using a BD GasPakTM EZ anaerobe container system (Becton – Dickinson, Sparks, MD, USA) at room temperature. Aliquots of ceca content were cultured in mEVS and mEVF media. For preparation of mEVS media, sEVs were isolated from store-bought skim bovine milk using sequential ultracentrifugation, sterile-filtered using a 0.22- μm filter, and suspended in phosphate-buffered saline (PBS).⁴⁵ mEVs were authenticated using Nanosight NS300 nanoparticle size analysis, immunoblot analysis, and scanning and transmission electron microscopy as previously described.¹⁹ Protocols have been deposited in the open-access EV-Track database (ID EV210338). mEVS and mEVF cultures were prepared by adding mEV suspension and PBS, respectively, to media. The concentration of mEVs in mEVS cultures was the equivalent of mEVs from 0.5 L bovine milk distributed in the intestinal water space in a human adult, normalized by body weight for mice.⁴⁶

Samples intended for use in metagenomics analysis were cultured in an anaerobic chamber (Coy Laboratory Products Inc., Grass Lake, MI, USA) at 37°C for 7 d and bacteria were pelleted by centrifugation for isolation of DNA (5,000 \times g, 10 min, room temperature). Samples intended for use in metabolomics and transcriptomics analysis were cultured in *Gut microbiota medium* (GMM) (Supplementary Table 10).^{47,48} Three mice were used for each treatment, and media inoculated with PBS was used as control. Cultures were collected after 7 d, pelleted by centrifugation, and the media supernatant was filter-sterilized using 0.2 μm filter. Both the media supernatants and bacterial cell pellets were flash frozen in liquid nitrogen and stored at –80°C until analysis.

Metagenomics and mRNA transcriptomics analyses

DNA was extracted using the PowerSoil DNA Isolation Kit following the manufacturer's

instructions (Mo Bio Laboratories Inc., Carlsbad, CA, USA). The purity and integrity of DNA were assessed by using the 260-to-280 nm ratio (Nanodrop ND-1000; Nanodrop Technologies, Wilmington, DE, USA) and agarose gel electrophoresis (0.8%). Libraries were prepared by using the Nextera XT kit (Illumina, San Diego, CA, USA) and their quality was assessed using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) in the Next Generation Sequencing Core Facility at the University of Nebraska Medical Center. Samples were sequenced using the Illumina NextSeq 500 platform (Illumina, San Diego, CA, USA) using a 75-basepair (bp) single end protocol with an estimated coverage of 150 \times . We used Trimmomatic to remove reads with a Phred quality score of less than Q30, suggesting that less than 1 in 1000 base calls were incorrect in the remaining reads.⁴⁹

Two bioinformatics pipelines were used to identify genomic variants selected by mEVs. We used the MIDAS pipeline to reveal the landscape of species abundance and strain-level genomic variants by alignment against more than 30,000 bacterial reference pangenomes. In addition, MIDAS was used to assess species-level extent of population structure among the metagenomic data and quality pan-genome gene content for prevalent bacterial species.²³ We used the StrainPhlAn pipeline to identify dominant sequence variants by concatenated alignment of species-specific marker genes at the strain level, thereby leading to a higher level of resolution compared to MIDAS.²⁴ StrainPhlAn uses the same reference library as MetaPhlAn2, which includes clade-specific marker genes from ~13,500 bacterial and archaeal, ~3,500 viral and ~110 eukaryotic reference genomes.⁵⁰ Muscle was used to call variants in MetaPhlAn2 marker genes through StrainPhlAn.^{51,52} We required a minimum of 50% of markers to be present in clades for each sample. Phylogenomic trees of strains were built based on the reconstructed marker genes by using the Randomized Accelerated Maximum Likelihood (RAxML) algorithm.³⁴ KEGG pathway analysis was performed by using a hypergeometric test.⁵³ The Integrated Microbial Genomes & Microbiomes system was used to assess the sizes of CDS and whole genomes in *B. bacterium*, *C. celerecrescens* and *E. faecalis*.^{25,26}

RNA was extracted from cell pellets of three independent biological replicates for each sex and media using RNeasy Mini Kit (Qiagen), following the manufacturer's recommendation. The RNA concentration and quality were determined using Qubit analyzer, and the Agilent RNA ScreenTape assay in combination with the TapeStation system. Analysis of data generated was performed using the Agilent TapeStation software 4.1. Total RNA yield from the samples ranged from 44 to 200 ng/ul. Subsequently, the ribosomal RNA was removed using the QIAseq FastSelect –5S/16S/23S rRNA removal kit (Qiagen), designed to selectively remove 5S, 16S, and 23S rRNA from complex bacterial community samples. The rRNA-free samples (100 ng) were used for cDNA library preparation and purification using KAPA RNA Hyperprep Kit (Roche). Each resulting uniquely dual-indexed library was quantified and quality accessed by Qubit and Agilent TapeStation, and multiple libraries were pooled in equal molarity. The pooled libraries were sequenced with 100-bp paired-end configuration on an Illumina NovaSeq 6000 sequencing platform. Transcript variants were linked with metabolic pathways using the Bacterial and Viral Bioinformatics Resource Center database.⁵⁴

Metabolomics analysis

Global metabolomics profiling was conducted using ultra-high-performance liquid chromatography-tandem mass-spectrometry by Metabolon Inc. (Morrisville, NC) as previously described.^{55,56} Briefly, samples were prepared using MicroLab STAR (Hamilton; Reno, NV). Proteins were precipitated with methanol under vigorous shaking for 2 min and removed by centrifugation using GenoGrinder 2000 (Glen Mills; Clifton, NJ). Supernatants were analyzed by reverse phase ultra-high-performance liquid chromatography – tandem mass spectroscopy analysis (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), RP/UPLC-MS/MS with negative ion mode ESI, and hydrophilic interaction liquid chromatography/ultra-high-performance liquid-chromatography-tandem mass spectroscopy (HILIC/UPLC-MS/MS) with negative ion mode ESI. Samples were analyzed using a ACQUITY UPLC (Waters Corporation; Milford, MA) and Q-Exactive high-resolution/accurate mass spectrometer interfaced with a heated ESI source and

Orbitrap mass analyzer operated at 35,000 mass resolution (Thermo Fisher Scientific; Waltham, MA). The raw data was processed using Metabolon's hardware and software as previously described.^{56,57} Metabolites were identified by comparison with libraries of authenticated standards with known retention time/indices, mass to charge ratios, and chromatographic and MS/MS spectral data. Biochemical identification was based on mass match (± 10 ppm), retention index, and forward- or reverse-search matching between the experimental data and library standards. The consistency of peak identification and library matches for each compound were checked for each sample and corrected if necessary. Peaks were quantified using area-under-the-curve. Peak area values allowed the determination of relative quantification of metabolites among samples.⁵⁸

Internalization of bovine mEVs by bacteria

Bacterial strains and media

Bacillus subtilis 168 (ATCC 23,857) and *Bifidobacterium longum* subsp. *infantis* (ATCC 15,697) were obtained from the American Type Culture Collection (Manassas, VA). *Escherichia coli* DH5 α was purchased from Thermo Fisher Scientific (Waltham, MA). *Bacillus subtilis* and *E. coli* were cultivated in LB (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37°C under aerobic conditions in flasks with shaking at 200 rpm. *Bifidobacterium infantis* was grown in de Man Rogosa Sharpe broth [MRS broth (MilliporeSigma; Burlington, MA)] supplemented with 0.05% (w/v) cysteine (Sigma-Aldrich, St. Louis, MO) and 2% (w/v). In reporter plasmid studies, media were supplemented with arabinose (0.1%) and isopropyl β -D-1-thiogalactopyranoside [IPTG, (0.5 mM)]. *B. infantis* was anaerobically grown in a vinyl chamber (Coy Laboratory Products, Grass Lake, MI) at 37°C for 24 h, in an atmosphere consisting of 5% carbon dioxide, 5% hydrogen, and 90% nitrogen. Optical density was assayed using a Biotek Synergy H1m microplate reader (BioTek Instruments, Inc.; Winooski, VT). Uptake of fluorophore-labeled mEVs

Milk EVs were isolated from skim bovine milk by differential ultracentrifugation and labeled with carbonyl-reactive HiLyteTM Fluor 750 hydrazide (AnaSpec, Inc, AS-81268) as previously described.⁵⁹

We incubated *Bacillus subtilis* 168 (ATCC 23,857) and *Escherichia coli* with HiLyte 750-labeled mEVs and assessed their uptake using an Odyssey Imaging System (LI-COR Bioscience; Lincoln, NE). Bacteria were incubated using an mEV concentration equivalent to mEVs from 0.5 L bovine milk suspended in the total gastrointestinal water in a human adult (1.7×10^{10} /mL).

Uptake of plasmid-loaded mEVs

Five micrograms of plasmid [addgene #26702 (EGFP) or addgene #118553 (mScarlet-I)] were added to 50 μ L mEV suspension, corresponding to $\sim 3 \times 10^8$ mEVs, in a final volume of 50 μ L electroporation buffer (Bio-Rad #1652677). The suspensions were mixed and incubated for 5 min at room temperature. Electroporation was carried out using 0.1-cm electroporation cuvettes (Bio-Rad 1,652,083) in a GenePulser Xcell electroporator (Bio-Rad; Hercules, CA). Electroporation was carried out at 1000 V and 125 μ F with 5 pulses and 10-s intervals between pulses. Free plasmids were removed by treating samples with DNase I (1.5 Kunitz units/ μ L, Qiagen #79254), followed by washing with PBS, centrifugation at 130,000 g at 4°C, and resuspending the mEVs in 50 μ L PBS.

Plasmid-loaded mEVs were added to bacterial cultures in 1–2 mL of antibiotics-free media and incubated at 37°C for 2 h in a shaking incubator under aerobic or anaerobic conditions. Clones were selected on LBS or MRS agar plates with antibiotics [ampicillin (100 μ g/mL) or kanamycin (50 μ g/mL)]. Clones were grown in liquid media until the absorbance at 600 nm reached 0.5–0.6 units. Expression of EGFP and mScarlet-I was induced with 0.5 mM IPTG and 0.1% L-arabinose at 37°C for 6 h. For analysis by confocal laser scanning microscopy, bacterial cultures were centrifuged at $2,500 \times g$ for 5 min and washed twice with PBS. The washed cells were resuspended in 1 mL PBS for imaging using a Nikon A1R confocal system (Thermo Fisher Scientific; Waltham, MA). The EGFP and mScarlet-I fluorescence was acquired using a 488 and 561 nm laser line and detected through a 530/30 and 585/42 nm bandpass emission filter, respectively. For analysis by flow cytometry, bacterial suspensions were diluted approximately 100-fold and analyzed using a CytoFLEX LX flow cytometer (Beckman Coulter; Brea, CA). For each sample, 50,000 events were collected for up to 10 min; doublets were

excluded by using a forward scatter with 50 as the discriminator. Data was acquired using CytExpert 2.0 (Beckman Coulter; Brea, CA) and analyzed using FlowJo software [(v10.7), FlowJo LLC; Ashland, OR]. All events were collected in R1 (Supplementary Fig6d₁), and additional regions were defined for fluorescence analysis. Events in R2 and R3 (Supplementary Fig6d₂–d₆) were defined as bacteria expressing EGFP and mScarlet-I, respectively.

Statistics

For sequencing data, the non-parametric Mann–Whitney U test was used to determine if differences were statistically significant in pairwise comparisons. Values are means \pm SD. $p < 0.05$ was considered statistically significant. For metabolomics data, peak area values for metabolites were median scaled and log transformed, and any missing values were replaced with the minimal value detected in the data set. Log-transformed data were analyzed using a two-way analysis of variance (ANOVA) adjusted for multiple comparisons with the false discovery rate procedure with ArrayStudio/Jupyter Notebook and R package qvalue (version 2.20.0).⁶⁰

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Acknowledgments

The research was supported by the National Institutes of Health (P20GM104320 and P30GM154608), the National Institute of Food and Agriculture (2016-67001-25301, 2020-67017-30834, and 2023-67017-39056), and the United States Department of Agriculture (Hatch NEB-36-096 and W5002). The authors thank the staff in the Biomedical and Obesity Research Core at the University of Nebraska-Lincoln for their expert technical assistance. The core is supported by the National Institutes of Health (P20GM104320 and P30GM154608).

Disclosure statement

J.Z. is a co-founder and co-owner of Minovacca, Inc., which develops genetically and chemically engineered milk

exosomes for delivering therapeutics. No potential conflict of interest was reported by the authors.

Funding

The work was supported by the National Institute of Food and Agriculture [2016-67001-25301]; National Institutes of Health [P20GM104320 and P30GM154608]; U.S. Department of Agriculture [Hatch NEB-036-096].

Author contributions

J.Z. designed the study. F.Z., P.T.M. and J.C. contributed to the study design and designed experiments. F.Z. and P.T.M. carried out experiments. F.Z., H.D., R.S. and J.C. analyzed data. F.Z. drafted the manuscript. P.T.M. and J.C. edited the draft manuscript. J.Z. wrote the final version of the manuscript and has the primary responsibility for final content. All authors read and approved the final manuscript.

Code availability

Custom scripts for manipulating DNA sequencing outputs have been deposited on GitHub at <https://github.com/publicationsubmission/naturecommunications2020>.

Data availability statement

The data (raw reads) that support the findings of this study are openly available in NCBI BioProject at <https://www.ncbi.nlm.nih.gov/bioproject>, reference numbers PRJNA650546 (DNA) and PRJNA1086571 (RNA). The authors confirm that the metabolomics data supporting the findings of this study are available within the article and its supplementary materials.

References

- O'Brien K, Breyne K, Ughetto S, Laurent LC, Breakefield XO. RNA delivery by extracellular vesicles in mammalian cells and its applications. *Nat Rev Mol Cell Biol*. 2020;21(10):585–606. doi:10.1038/s41580-020-0251-y.
- Yang J, Kim EK, McDowell A, Kim YK. Microbe-derived extracellular vesicles as a smart drug delivery system. *Transl Clin Pharmacol*. 2018;26(3):103–110. doi:10.12793/tcp.2018.26.3.103.
- Nagakubo T, Nomura N, Toyofuku M. Cracking open bacterial membrane vesicles. *Front Microbiol*. 2019;10:3026. doi:10.3389/fmicb.2019.03026.
- Valadi H, Ekström K, Bossios A, Sjöstrand M, Lee JJ, Lötvald JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange

- between cells. *Nat Cell Biol*. 2007;9(6):654–659. doi:10.1038/ncb1596.
- Jung AL, Stoiber C, Herkt CE, Schulz C, Bertrams W, Schmeck B. Legionella pneumophila-derived outer membrane vesicles promote bacterial replication in macrophages. *PLOS Pathog*. 2016;12(4):e1005592. doi:10.1371/journal.ppat.1005592.
- Sukreet S, Braga CP, An TT, Adamec J, Cui J, Tribble B, Zemleni J. Isolation of extracellular vesicles from byproducts of cheese making by tangential flow filtration yields heterogeneous fractions of nanoparticles. *J Dairy Sci*. 2021;104(9):9478–9493. doi:10.3168/jds.2021-20300.
- Leiferman A, Shu J, Upadhyaya B, Cui J, Zemleni J. Storage of extracellular vesicles in human milk, and microRNA profiles in human milk exosomes and infant formulas. *J Pediatr Gastroenterol Nutr*. 2019;69(2):235–238. doi:10.1097/MPG.0000000000002363.
- Haisma H, Coward WA, Albernaz E, Visser GH, Wells JCK, Wright A, Victora CG. Breast milk and energy intake in exclusively, predominantly, and partially breast-fed infants. *Eur J Clin Nutr*. 2003;57(12):1633–1642. doi:10.1038/sj.ejcn.1601735.
- Izumi H, Kosaka N, Shimizu T, Sekine K, Ochiya T, Takase M. Bovine milk contains microRNA and messenger RNA that are stable under degradative conditions. *J Dairy Sci*. 2012;95(9):4831–4841. doi:10.3168/jds.2012-5489.
- Zhou F, Ebea P, Mutai E, Wang H, Sukreet S, Navazesh S, Dogan H, Li W, Cui J, Ji P, et al. Small extracellular vesicles in milk cross the blood-brain barrier in murine cerebral cortex endothelial cells and promote dendritic complexity in the hippocampus and brain function in C57BL/6J mice. *Front Nutr*. 2022;9:838543. doi:10.3389/fnut.2022.838543.
- Sadri M, Shu J, Kachman SD, Cui J, Zemleni J. Milk exosomes and microRNAs cross the placenta and promote embryo survival in mice. *Reproduction*. 2020;160(4):501–509. doi:10.1530/REP-19-0521.
- Rosa F, Mercer KE, Lin H, Sims CR, Pack LM, Goode G, Badger T, Andres A, Yeruva L. Early infant formula feeding impacts urinary metabolite profile at 3 months of age. *Nutrients*. 2020;12(11):3552. doi:10.3390/nu12113552.
- Deoni S, Dean D 3rd, Joelson S, O'Regan J, Schneider N. Early nutrition influences developmental myelination and cognition in infants and young children. *Neuroimage*. 2018;178:649–659. doi:10.1016/j.neuroimage.2017.12.056.
- Li B, Hock A, Wu RY, Minich A, Botts SR, Lee C, Antounians L, Miyake H, Koike Y, Chen Y, et al. Bovine milk-derived exosomes enhance goblet cell activity and prevent the development of experimental necrotizing enterocolitis. *PLOS ONE*. 2019;14(1):e0211431. doi:10.1371/journal.pone.0211431.
- Eidelman AI, Schanler RJ, Johnston M, Landers S, Noble L, Szucs K, Viehmann L. American Academy of

- Pediatrics, section on breastfeeding. Breastfeeding and the use of human milk. *Pediatrics*. 2012;129(3):e827–e841. doi:10.1542/peds.2011-3552.
16. Centers for Disease Control and Prevention. Breastfeeding among U.S. children born 2010–2017, CDC national immunization surveys. [accessed 2021 Aug 12]. <https://www.cdc.gov/breastfeeding/data/reportcard.htm>.
 17. Khanam A, Ngu A, Zemleni J. Bioavailability of orally administered small extracellular vesicles from bovine milk in C57BL/6J mice. *Int J Pharm*. 2023;639:122974. doi:10.1016/j.ijpharm.2023.122974.
 18. Benmoussa A, Lee CHC, Laffont B, Savard P, Laugier J, Boilard E, Gilbert C, Fliss I, Provost P. Commercial dairy cow milk microRNAs resist digestion under simulated gastrointestinal tract conditions. *J Nutr*. 2016;146(11):2206–2215. doi:10.3945/jn.116.237651.
 19. Zhou F, Paz HA, Sadri M, Cui J, Kachman SD, Fernando SC, Zemleni J. Dietary bovine milk exosomes elicit changes in bacterial communities in C57BL/6 mice. *Am J Physiol Gastrointest Liver Physiol*. 2019;317(5):G618–G624. doi:10.1152/ajpgi.00160.2019.
 20. Aguilar-Lopez M, Wetzel C, MacDonald A, Ho TTB, Donovan SM. Metagenomic profile of the fecal microbiome of preterm infants consuming mother's own milk with bovine milk-based fortifier or infant formula: a cross-sectional study. *Am J Clin Nutr*. 2022;116(2):435–445. doi:10.1093/ajcn/nqac081.
 21. Zheng X, Cai X, Hao H. Emerging targetome and signalome landscape of gut microbial metabolites. *Cell Metab*. 2022;34(1):35–58. doi:10.1016/j.cmet.2021.12.011.
 22. Wang S, Auchtung J, Zemleni J. Milk exosomes protect human microbiota associated-mice against *Clostridioides difficile* infection [peer-reviewed meeting abstract]. *Curr Dev Nutr*. 2021;5:1189. doi:10.1093/cdn/nzab054_044.
 23. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res*. 2016;26(11):1612–1625. doi:10.1101/gr.201863.115.
 24. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res*. 2017;27(4):626–638. doi:10.1101/gr.216242.116.
 25. Chen IA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. 2019;47(D1):D666–D677. doi:10.1093/nar/gky901.
 26. Mukherjee S, Stamatis D, Li C, Ovchinnikova G, Bertsch J, Sundaramurthi J, Kandimalla M, Nicolopoulos P, Favagnano A, Chen I-M, et al. Twenty-five years of genomes OnLine database (GOLD): data updates and new features in v.9. *Nucleic Acids Res*. 2023;51(D1):D957–D963. doi:10.1093/nar/gkac974.
 27. Aguilar-Lozano A, Baier S, Grove R, Shu J, Giraud D, Leiferman A, Mercer KE, Cui J, Badger TM, Adamec J, et al. Concentrations of purine metabolites are elevated in fluids from adults and infants and in livers from mice fed diets depleted of bovine milk exosomes and their RNA cargos. *J Nutr*. 2018;148(12):1886–1894. doi:10.1093/jn/nxy223.
 28. Mergenthaler P, Lindauer U, Dienel GA, Meisel A. Sugar for the brain: the role of glucose in physiological and pathological brain function. *Trends Neurosci*. 2013;36(10):587–597. doi:10.1016/j.tins.2013.07.001.
 29. Tang Z, Ye W, Chen H, Kuang X, Guo J, Xiang M, Peng C, Chen X, Liu H. Role of purines in regulation of metabolic reprogramming. *Purinergic Signal*. 2019;15(4):423–438. doi:10.1007/s11302-019-09676-z.
 30. Duster R, Prickaerts J, Blokland A. Purinergic signaling and hippocampal long-term potentiation. *Curr Neuropharmacol*. 2014;12(1):37–43. doi:10.2174/1570159X113119990045.
 31. Cieslak M, Wojtczak A, Komoszynski M. Role of the purinergic signaling in epilepsy. *Pharmacol Rep*. 2017;69(1):130–138. doi:10.1016/j.pharep.2016.09.018.
 32. Abel T, Nguyen PV. Regulation of hippocampus-dependent memory by cyclic amp-dependent protein kinase. *Prog Brain Res*. 2008;169:97–115.
 33. Teng Y, Ren Y, Sayed M, Hu X, Lei C, Kumar A, Hutchins E, Mu J, Deng Z, Luo C, et al. Plant-derived exosomal microRNAs shape the gut microbiota. *Cell Host Microbe*. 2018;24(5):637–652 e638. doi:10.1016/j.chom.2018.10.001.
 34. Gulick JT. Divergent evolution through cumulative segregation. *J Linn Soc Lond Zool*. 1888;20(120):189–274. doi:10.1111/j.1096-3642.1888.tb01445.x
 35. Gerbault P, Liebert A, Itan Y, Powell A, Currat M, Burger J, Swallow DM, Thomas MG. Evolution of lactase persistence: an example of human niche construction. *Philos Trans R Soc Lond B Biol Sci*. 2011;366(1566):863–877. doi:10.1098/rstb.2010.0268.
 36. Schwechheimer C, Kuehn MJ. Outer-membrane vesicles from gram-negative bacteria: biogenesis and functions. *Nat Rev Microbiol*. 2015;13(10):605–619. doi:10.1038/nrmicro3525.
 37. Brown L, Wolf JM, Prados-Rosales R, Casadevall A. Through the wall: extracellular vesicles in gram-positive bacteria, mycobacteria and fungi. *Nat Rev Microbiol*. 2015;13(10):620–630. doi:10.1038/nrmicro3480.
 38. Wilson DK, Rudolph FB, Quiocho FA. Atomic structure of adenosine deaminase complexed with a transition-state analog: understanding catalysis and immunodeficiency mutations. *Science*. 1991;252(5010):1278–1284. doi:10.1126/science.1925539.
 39. Wang J, Zhu N, Su X, Gao Y, Yang R. Gut-microbiota-derived metabolites maintain gut and systemic immune

- homeostasis. *Cells*. 2023;12(5):793. doi:10.3390/cells12050793.
40. Collins J, Robinson C, Danhof H, Knetsch CW, van Leeuwen HC, Lawley TD, Auchtung JM, Britton RA. Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature*. 2018;553(7688):291–294. doi:10.1038/nature25178.
 41. Le Guillou S, Leduc A, Laubier J, Barbey S, Rossignol M-N, Lefebvre R, Marthey S, Laloë D, Le Provost F. Characterization of holstein and normande whole milk miRNomes highlights breed specificities. *Sci Rep*. 2019;9(1):20345. doi:10.1038/s41598-019-56690-7.
 42. Gao R, Zhang R, Qian T, Peng X, He W, Zheng S, Cao Y, Pierro A, Shen C. A comparison of exosomes derived from different periods breast milk on protecting against intestinal organoid injury. *Pediatr Surg Int*. 2019;35(12):1363–1368. doi:10.1007/s00383-019-04562-6.
 43. Liu B, Lu Y, Chen X, Muthuraj PG, Li X, Pattabiraman M, Zemleni J, Kachman SD, Natarajan SK, Yu J, et al. Protective role of shiitake mushroom-derived exosome-like nanoparticles in D-galactosamine and lipopolysaccharide-induced acute liver injury in mice. *Nutrients*. 2020;12(2):477. doi:10.3390/nu12020477.
 44. Maharjan RP, Ferenci T. *Escherichia coli* mutation rates and spectra with combinations of environmental limitations. *Microbiology (Reading)*. 2018;164(12):1491–1502. doi:10.1099/mic.0.000727.
 45. Manca S, Upadhyaya B, Mutai E, Desaulniers AT, Cederberg RA, White BR, Zemleni J. Milk exosomes are bioavailable and distinct microRNA cargos have unique tissue distribution patterns. *Sci Rep*. 2018;8(1):11321. doi:10.1038/s41598-018-29780-1.
 46. McConnell EL, Basit AW, Murdan S. Measurements of rat and mouse gastrointestinal pH, fluid and lymphoid tissue, and implications for in-vivo experiments. *J Pharm Pharmacol*. 2008;60(1):63–70. doi:10.1211/jpp.60.1.0008.
 47. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
 48. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods*. 2012;9(8):811–814. doi:10.1038/nmeth.2066.
 49. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–1797. doi:10.1093/nar/gkh340.
 50. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–1313. doi:10.1093/bioinformatics/btu033.
 51. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, Wadi L, Meyer M, Wong J, Xu C, et al. Pathway enrichment analysis and visualization of omics data using g: profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc*. 2019;14(2):482–517. doi:10.1038/s41596-018-0103-9.
 52. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods*. 2015;12(10):902–903. doi:10.1038/nmeth.3589.
 53. Hung KS, Hsiao C-C, Pai T-W, Hu C-H, Tzou W-S, Wang W-D, Chen Y-R. Functional enrichment analysis based on long noncoding RNA associations. *BMC Syst Biol*. 2018;12(S4):45. doi:10.1186/s12918-018-0571-0.
 54. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis J, Dempsey D, Dickerman A, Dietrich E, Kenyon R, et al. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res*. 2023;51(D1):D678–D689. doi:10.1093/nar/gkac1003.
 55. Ford L, Kennedy AD, Goodman KD, Pappan KL, Evans AM, Miller LAD, Wulff JE, Wiggs BR, Lennon JJ, Elsea S, et al. Precision of a clinical metabolomics profiling platform for use in the identification of inborn errors of metabolism. *J Appl Lab Med*. 2020;5(2):342–356. doi:10.1093/jalm/jfz026.
 56. Mesnage R, Teixeira M, Mandrioli D, Falcioni L, Ducarmon QR, Zwittink RD, Mazzacuva F, Caldwell A, Halket J, Amiel C, et al. Use of shotgun metagenomics and metabolomics to evaluate the impact of glyphosate or roundup MON 52276 on the gut microbiota and serum metabolome of Sprague-Dawley rats. *Environ Health Perspect*. 2021;129(1):17005. doi:10.1289/EHP6990.
 57. Dehaven CD, Evans AM, Dai H, Lawton KA. Organization of GC/MS and LC/MS metabolomics data into chemical libraries. *J Cheminform*. 2010;2(1). doi:10.1186/1758-2946-2-9.
 58. Evans AM, DeHaven CD, Barrett T, Mitchell M, Milgram E. Integrated, nontargeted ultrahigh performance liquid chromatography/electrospray ionization tandem mass spectrometry platform for the identification and relative quantification of the small-molecule complement of biological systems. *Anal Chem*. 2009;81(16):6656–6667. doi:10.1021/ac901536h.
 59. Khanam A, Yu J, Zemleni J. Class a scavenger receptor-1/2 facilitates the uptake of bovine milk exosomes in murine bone marrow-derived macrophages and C57BL/6J mice. *Am J Physiol Cell Physiol*. 2021;321(3):C607–C614. doi:10.1152/ajpcell.00222.2021.
 60. Storey JD. False discovery rate, in international encyclopedia of statistical science. M Lovric, Editor. Berlin, Heidelberg: Springer; 2011. p. 504–508.