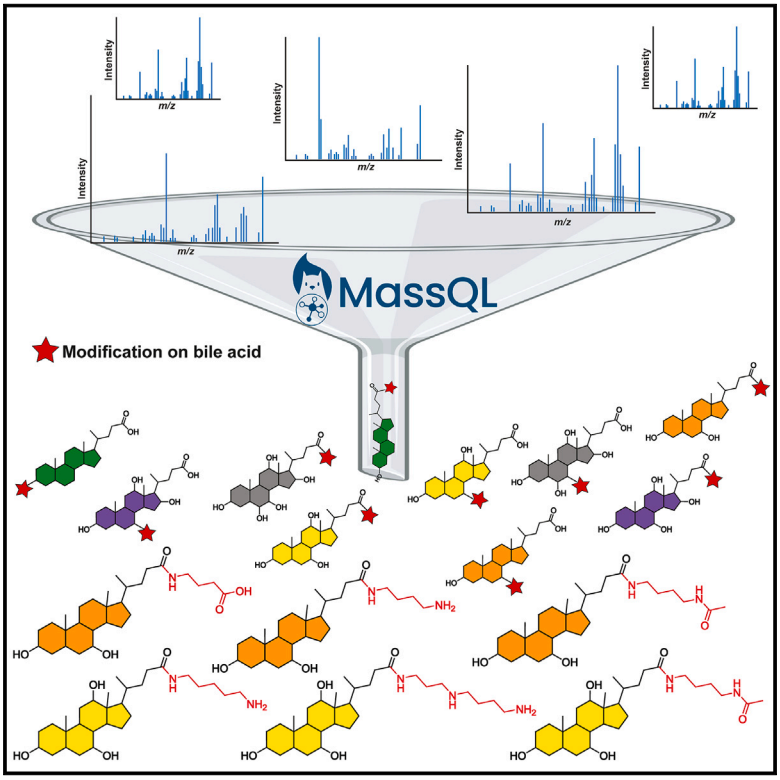
Cell

Resource

The underappreciated diversity of bile acid modifications

Graphical abstract



Highlights

- Mining of public, untargeted LC-MS/MS data reveals a large diversity of bile acids
- This work created an MS/MS spectral library resource of modified bile acids
- Polyamine biosynthetic pathway-derived metabolites amidated to bile acids discovered
- American-to-Mediterranean diet transition altered levels of polyamine bile amidates

Authors

Ipsita Mohanty, Helena Mannochio-Russo, Joshua V. Schweer, ..., Mingxun Wang, Allegra T. Aron, Pieter C. Dorrestein

Correspondence

pdorrestein@health.ucsd.edu

In brief

Using mass spectrometry-based tools to mine public untargeted MS/MS data revealed the vastly underappreciated diversity of bile acids hidden in plain sight and led to the discovery of previously unknown polyamine-conjugated bile acids.

Mohanty et al., 2024, Cell *187*, 1–18
March 28, 2024
https://doi.org/10.1016/j.cell.2024.02.019

CellPress

Cell

CellPress
OPEN ACCESS

Resource

The underappreciated diversity of bile acid modifications

Ipsita Mohanty,[1,18] Helena Mannochio-Russo,[1,18] Joshua V. Schweer,[1,17] Yasin El Abiead,[1] Wout Bittremieux,[15] Shipei Xing,[1,3] Robin Schmid,[1,2] Simone Zuffa,[1] Felipe Vasquez,[1] Valentina B. Muti,[7,10] Jasmine Zemlin,[1,9] Omar E. Tovar-Herrera,[13,14] Sarah Moraïs,[13,14] Dhimant Desai,[4] Shantu Amin,[4] Imhoi Koo,[5] Christoph W. Turck,[11,12] Itzhak Mizrahi,[13,14] Penny M. Kris-Etherton,[16] Kristina S. Petersen,[16] Jennifer A. Fleming,[16] Tao Huan,[3] Andrew D. Patterson,[5] Dionicio Siegel,[1] Lee R. Hagey,[6] Mingxun Wang,[7] Allegra T. Aron,[10] and Pieter C. Dorrestein[1,2,8,9,19,*]

[1]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA, USA
[2]Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, USA
[3]Department of Chemistry, Faculty of Science, University of British Columbia, Vancouver Campus, Vancouver, BC, Canada
[4]Department of Pharmacology, Penn State University College of Medicine, Hershey, PA, USA
[5]Center for Molecular Toxicology and Carcinogenesis, Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA
[6]Department of Medicine, University of California, San Diego, San Diego, CA, USA
[7]Department of Computer Science and Engineering, University of California, Riverside, Riverside, CA, USA
[8]Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA
[9]Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA 92093, USA
[10]Department of Chemistry and Biochemistry, University of Denver, Denver, CO 80210, USA
[11]Max Planck Institute of Psychiatry, Proteomics and Biomarkers, Kraepelinstrasse 2-10, Munich 80804, Germany
[12]Key Laboratory of Animal Models and Human Disease Mechanisms of Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650201, China
[13]Department of Life Sciences, Ben-Gurion University of the Negev, Be'er Sheva, Israel
[14]Goldman Sonnenfeldt School of Sustainability and Climate Change, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel
[15]Department of Computer Science, University of Antwerp, 2020 Antwerpen, Belgium
[16]Department of Nutritional Sciences, The Pennsylvania State University, University Park, PA, USA
[17]Department of Chemistry and Biochemistry, University of California, San Diego, San Diego, CA, USA
[18]These authors contributed equally
[19]Lead contact
*Correspondence: pdorrestein@health.ucsd.edu
https://doi.org/10.1016/j.cell.2024.02.019

SUMMARY

The repertoire of modifications to bile acids and related steroidal lipids by host and microbial metabolism remains incompletely characterized. To address this knowledge gap, we created a reusable resource of tandem mass spectrometry (MS/MS) spectra by filtering 1.2 billion publicly available MS/MS spectra for bile-acid-selective ion patterns. Thousands of modifications are distributed throughout animal and human bodies as well as microbial cultures. We employed this MS/MS library to identify polyamine bile amidates, prevalent in carnivores. They are present in humans, and their levels alter with a diet change from a Mediterranean to a typical American diet. This work highlights the existence of many more bile acid modifications than previously recognized and the value of leveraging public large-scale untargeted metabolomics data to discover metabolites. The availability of a modification-centric bile acid MS/MS library will inform future studies investigating bile acid roles in health and disease.

INTRODUCTION

Bile acids and related steroids play critical roles in animals, including human, biology. They are involved in nutrient transport, organelle function, such as in mitochondria, and signaling via a broad range of G protein-coupled receptors, nuclear receptors, and ion channels where they are active in the μM to pM concentrations.[1,2] Bile acids also regulate the immune system through T cell differentiation[3–7] and play key roles in various diseases, including norovirus infections, inflammatory bowel disease (IBD), gastric cancers, Alzheimer's disease, diabetes, and others.[8,9] Bile acids, bile acid derivatives, or bile acid mimics are currently being tested in clinical trials for the treatment of multiple sclerosis, IBD, carcinoma, type 2 diabetes (T2D),

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

gastroesophageal reflux disease, sepsis in malnourished children, peroxisomal disorders, asthma, and alcoholic hepatitis. They are also currently prescribed to treat primary biliary cirrhosis, neonatal bilirubin accumulation, itching, gallstones, and bile acid synthesis disorders.[10] Most recently, the Food and Drug Administration (FDA) approved the use of tauroursodeoxycholic acid for extending the life of patients with amyotrophic lateral sclerosis.[11]

Bile acids, such as taurocholate, are synthesized from cholesterol through oxidation and amidation with taurine in the liver, followed by storage in the gallbladder.[12,13] Once a meal is consumed, bile acids from the gallbladder are released into the duodenum, where they can undergo microbial biotransformations including deconjugation, oxidation, isomerization, reduction, acylation, glucuronidation, other saccharide additions, sulfation, and dehydration as they move through the digestive tract before they are absorbed into the bloodstream, facilitating the absorption of nutrients. The total bile acid concentration in the human gallbladder can be as high as 300 mM.[14] Once the gallbladder is emptied into the duodenum, the place where microbes first act on the bile acids, their concentration reaches over 10 mM.[14] The concentrations are subject to the timing of food intake and diurnal patterns,[15,16] as they vary significantly. As bile acids move down the digestive tract, the concentrations of bile acids further decrease to 0.2–3 mM.[17] Thus, even a 0.1% conversion by microorganisms would yield 200 nM to 10 μM concentration of modified bile acids. This concentration is sufficient for receptor binding. Therefore, microbially modified bile acids should be considered abundant biological molecules with plausible physiological implications. Recently, we reported the discovery of bile acids where cholic acid was amidated with phenylalanine, leucine, tyrosine,[18] alanine, serine, and deoxycholic acid with phenylalanine and histidine.[19–21] The term amidate or amidation used in this paper is based on the 1992 proposed nomenclature for bile acids, wherein the aminoacyl amidate refers to an amine linked to the carboxylic acid of the bile acid—the acyl group—forming an amide.[22] And this term is used to distinguish other non-nitrogen conjugations to bile acids, such as glycosylations, or acylations generated via esters. The discovery of these amidates builds on other reports that have identified unusual bile acid conjugates with ornithine in a mammal[23] and in humans with cholestasis,[24] dipeptides in rabbits,[25] arginine from rat liver,[26] fluoro-β-alanine that originated from the drug fluorouracil,[27] *N*-methyltaurine,[28,29] cysteinolic acid in fish,[30] and cysteamine in mice.[31] The abundance of microbial bile amidates can change under different health conditions, such as IBD,[32] non-alcoholic fatty acid liver disease,[33] and T2D in humans.[34] Although still in the early phases of our understanding, microbial bile amidates have many roles—they function as ligands for receptors,[35] assist in T cell differentiation, have antimicrobial properties, and alter the germination of *Clostridioides difficile*.[36,37] Sixteen cholate and four deoxycholate amidates are produced by the microbial enzymes bile salt hydrolases (BSHs) *in vitro*, *ex vivo*, and *in vivo*.[35,38] The BSHs, from *Bacteroides fragilis*, *Clostridium perfringens*, and *Lactobacillus* species, known to catalyze the hydrolysis of taurocholate and glycocholate, have now also been shown to possess aminotransferase activities.[35,36,38]

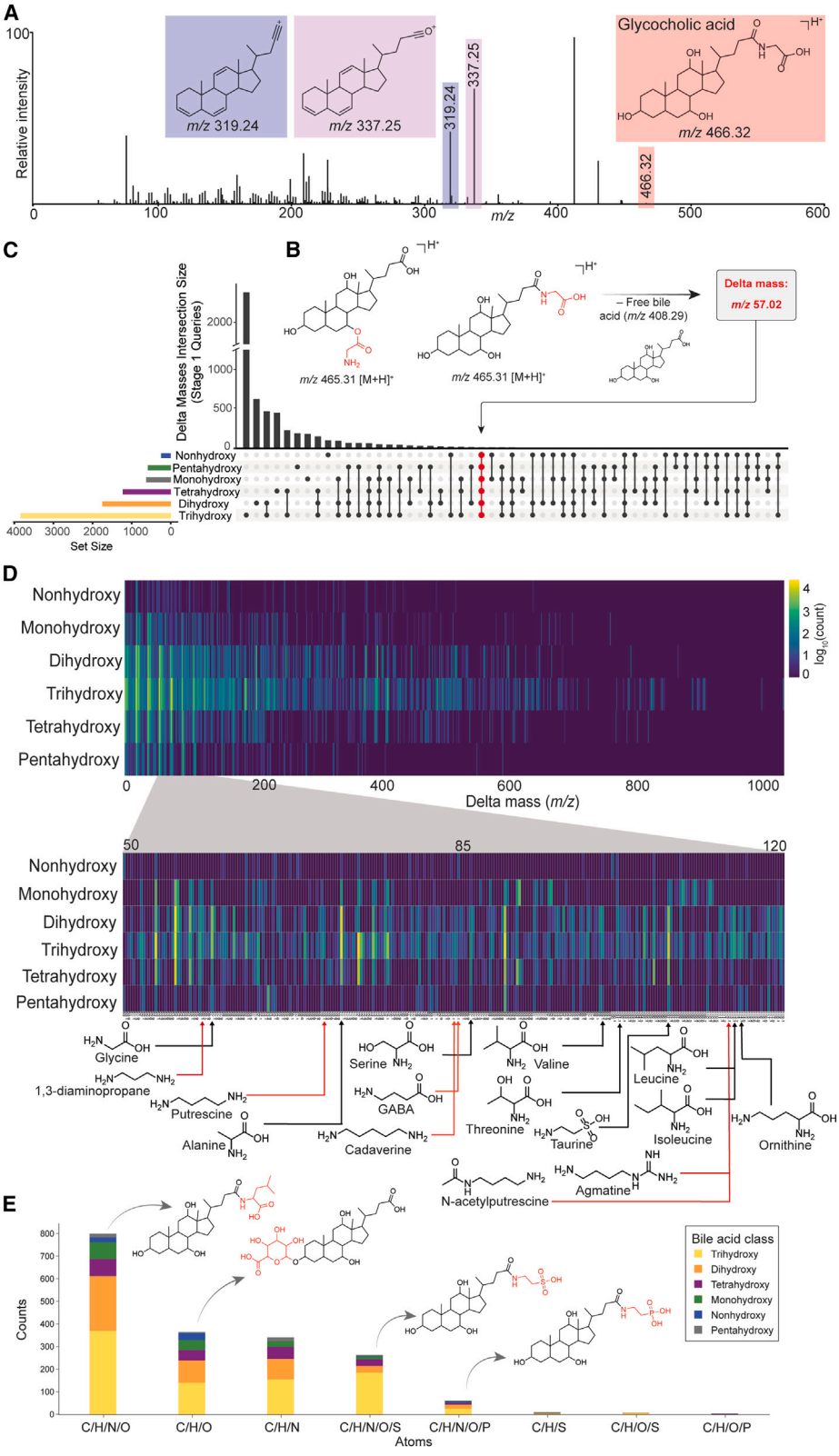
Since the discovery of microbial bile amidates, techniques like culturing of fecal bacteria with bile acids,[39] innovative mass spectrometry (MS) informatic approaches such as molecular networking,[18] reverse metabolomics,[40] *in silico* compound classification,[20] polarity switching,[41] and ion mobility MS,[42] have been applied to discover bile amidates, with both proteinogenic and non-proteinogenic amino acids. From 2020 to 2023, just under 200 of such amidates have been discovered; many of which were detected in public untargeted metabolomics data but had previously remained unidentified. These observations suggest that existing public data can be better leveraged to define the diversity of modifications that bile acids and related lipids might undergo.

Herein, we describe how mining public tandem mass spectrometry (MS/MS) spectra in GNPS/MassIVE (Global Natural Product Social Molecular Networking/Mass Spectrometry Interactive Virtual Environment)[43] facilitated the generation of a reusable candidate bile acid MS/MS library. Although molecular networking can be used to discover modified bile acids,[18,44] as much as 80% of all MS/MS comparisons of bile acids have modified cosine similarities below the scoring threshold of 0.7 and can, therefore, evade discovery by molecular networking.[45] This is especially true when the modification of the bile acids results in large changes in the MS/MS spectra. Therefore, we employed our recently developed mass query language (MassQL)[46] to expand the discovery of unknown bile acids. This highlights a unique approach for metabolite discovery at the public data repository scale. We also demonstrate the use of our bile acid library as a resource, leading to the discovery of bile acids amidated with polyamines. These bile amidates were found in fecal samples recovered from public studies and validated with synthetic standards. We provide this bile acid MS/MS spectral library as a public resource to promote and accelerate research delineating the role of bile acids in human health, metabolism, and diseases.

RESULTS

To build the MS/MS spectral library, all the MS/MS spectra possibly originating from bile acids were retrieved from the GNPS/MassIVE public repository (~1.2 billion spectra, 304,415 liquid chromatography-tandem mass spectrometry [LC-MS/MS] files, 2,706 projects). This was achieved using MassQL,[46] a recently developed tool, which allows filtering public mass spectral data for selective data patterns. Since more than 90% of the data in GNPS/MassIVE is acquired in positive ionization mode, and the recent discoveries of bile acids from positive ion mode data[18,40,41] coupled with the rich MS/MS fragmentation of both unconjugated and conjugated bile acids (Figures S1A and S1B), we only mined positive ionization data in the public domain. The bile acid diagnostic ions, which led to the development of the reusable MassQL queries, were obtained via expert manual interpretation of the bile acid fragmentation spectra available in the GNPS bile acid spectral library (4,533 bile acid reference spectra). The queries developed leverage the expert domain knowledge of our co-authors. Next, we checked if other bile acids with same hydroxylation states had same fragment ions and assessed matches obtained against all 582,966 non-bile acid reference MS/MS spectra as decoy for false discovery estimation.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019



(legend on next page)

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

As an example, for the tri-hydroxylated bile acids, via the above expert curation, we identified two diagnostic fragment ions at *m/z* 319.24 and *m/z* 337.25, which represent a completely dehydrated and deconjugated bile acid (Figure 1A), and designed a MassQL query selective to tri-hydroxylated bile acids (Figure S1C). The two MS/MS fragment ions selected were also observed at low and high normalized collision energies (NCEs) for fragmentation (Figures S1A and S1B) and with both conjugated and unconjugated bile acids. However, our recommendation would be to collect data at energies lower than NCE45 on an Orbitrap, as the diagnostic ion intensities decrease after that. The query can be used to search for bile acids with a tri-hydroxylated sterol core, irrespective of the location of the hydroxyl groups on the core or on the 5-carbon side chain. Correspondingly, characteristic fragment ions for di-hydroxylated bile acids are at *m/z* 321.26 and *m/z* 339.27, 2 Da more than a tri-hydroxylated core. Given that the number of hydroxyl groups on a bile acid core can vary, we created MassQL queries for non- (0), mono- (1), di- (2), tri- (3), tetra- (4), and penta-hydroxylated (5) bile acids (henceforth referred to as classes of bile acids) to obtain an exhaustive list of potential bile acid MS/MS spectra (Table S1). To assess the selectivity of our queries, MS/MS spectra obtained from the MassQL queries were searched against all publicly accessible reference spectra with known structural annotations (587,449 spectra of which 4,533 are bile acid reference spectra). The search was performed using the GNPS-based cosine spectral alignment score with settings that typically result in <1% false discovery rate (FDR).[43,47] From the library search of the total 594,365 spectra that were retrieved from the queries, 270,437 MS/MS reference library matches were obtained of which 269,711 matched to MS/MS of known bile acids, whereas 726 MS/MS spectra matched to non-bile acids (Table S2), resulting in an FDR of 0.27%. These results indicate the high selectivity of our MassQL queries for bile acid spectra.

With the selectivity of our queries tested, we utilized MassQL to find putative bile acid MS/MS spectra in public GNPS/MassIVE data. Due to computational limitations, full-scale repository MassQL queries could not be accomplished, and therefore, we limited the MassQL query-based filtering to only data acquired on Orbitrap instruments. After allocating 400 central processing units cores, each filtering step still took between 6 and 10.5 h for each query. We identified 137,381 and 320,370 MS/MS spectra with di-hydroxylated and tri-hydroxylated bile acid queries. We also performed similar searches for non-, mono-, tetra-, and penta-hydroxylated steroid cores (Table S1). MassQL-filtered data were subjected to molecular networking, which revealed many distinct networks. This spectral alignment analysis resulted in networks clustered by bile acid classes, with different hydroxylation generally separated into their own subnetworks (Figure S1D), with the rare exception of bile acids with glycosylation, where the glycan fragmentation was driving the spectral similarity, indicating the clustering in the molecular network is driven by the steroid core (Figure S1E). This is consistent with the previous observation that modified cosine-based networking would align about 20% of the bile acid spectra[45] and highlights the value of complementing molecular networking with MassQL queries to find bile acid analogs. Next, to identify the modifications to the bile acid core structures, we calculated the delta masses of the modified bile acids compared with the corresponding unconjugated or free bile acids. As an example, the mass shift of glycocholic acid (*m/z* 465.31, $[M+H]^+$) from cholic acid (*m/z* 408.29, $[M+H]^+$) yields a delta mass of *m/z* 57.07, which corresponds to glycine. However, the position of modification on the bile acid, either on the carboxyl or hydroxyl groups, remains ambiguous (Figure 1B). The resulting list of MassQL-filtered bile acid MS/MS spectra revealed a total of 5,576 delta masses upon binning (see STAR Methods), which was used to generate the MS/MS spectral library. We expect these MassQL searches to include not only modifications or multiple modifications but also other ion forms, including different adducts, in-source fragments, and multimers. The distribution of delta masses across different bile acid classes is visualized as an UpSet plot (Figure 1C; Tables S1 and S2), which reveals the number of delta masses (represented as vertical columns called intersection size) that are detected in the different bile acid classes (colored horizontal columns called set size, Figure 1C). Delta masses overlapping across multiple bile acid classes are connected by a vertical line. Delta mass corresponding to glycine is one among the 15 shared delta masses detected across all 6 bile acid classes.

The highest number of delta masses (3,839 out of 5,576) were observed in the tri-hydroxylated query, followed by the second highest (1,746 out of 5,576) observed in the di-hydroxyl query. This observation mirrors the predominance of cholic (tri-hydroxylated) and chenodeoxycholic (di-hydroxylated) acids in biological specimens.[14] Out of the 5,576, fifteen delta masses were observed in all classes of bile acids. Subsequently, the

Figure 1. MassQL query and results from repository search

(A) Representative MS/MS spectrum of glycocholic acid (tri-hydroxylated bile acid amidate, USI:mzspec:GNPS:BILELIB19:accession:CCMSLIB00005435513) with the diagnostic fragment ions indicated that were used to design the MassQL query.

(B) Schematic representation of delta mass calculation using glycocholic acid as an example. Glycine can be appended to the bile acid either as an ester or amidate, and subtracting the mass of the free bile acid—in this case, cholic acid—yields the delta mass corresponding to a mass of glycine (*m/z* 57.02).

(C) Distribution of delta masses across bile acid classes, from non-, mono-, di-, tri-, tetra-, and penta-hydroxylated bile acids, visualized as an UpSet plot of delta masses. Each vertical column represents the number of delta masses, called an intersection size, which is connected to the corresponding bile acid classes, represented by the horizontal columns on the left and called set size, where the intersection of delta masses is detected. Overlapping delta masses are connected to multiple set sizes. As an example, the delta mass of glycine, *m/z* 57.02, is one of 15 delta masses that are observed in bile acids of all classes (the full list is provided in Table S2).

(D) Heatmap of observed delta masses based on their frequency of occurrence in log scale within different classes of bile acids. A zoomed-in heatmap for delta masses between 50 and 120 *m/z* is also shown with the delta masses corresponding to MS1 matches to amino acids and amines demonstrated with structures. Arrows in red show amines that are discovered in this study and discussed later in the paper.

(E) Summary of the atomic compositions of the delta masses recovered from formula prediction tools. See also Figure S1 and Tables S1 and S2.

occurrences of delta masses in different bile acid classes were investigated, and a higher dominance of delta masses lower than *m/z* 400 was highlighted for di- and tri-hydroxylated bile acids (Figure 1D; Table S2). In addition, consistent with the analysis of only positive mode data, the majority of atomic compositions that could be assigned to the delta masses included nitrogen (N) (Figure 1E; Table S2).[48,49] Delta masses with atomic composition comprising a carbon, hydrogen, nitrogen, and oxygen only (denoted by C/H/N/O) were the most frequently observed, which includes all the bile amidates with amino acids[39–41] and N-acetylglucosamine conjugates,[50–52] followed by C/H/O, which includes free bile acids, acetylated, lipid acylated,[18,53,54] or non-amine containing glucuronidated bile acids (Figure 1E).[55–57] It is important to note here that the atomic compositions belong to the modification only and not the bile acid itself. Interestingly, modifications corresponding to the atomic composition consisting of only C/H/N and devoid of O were unexpectedly common and included many of the polyamine amidates described later in the text.

To the best of our knowledge, there are no reports in the scientific literature on modified bile acids with conjugations containing only C/H/Ns (excluding the bile acid itself), though most decarboxylated amino acids would fall in this category. Incorporation of the sulfur atom (S), commonly part of either the taurine amidates or as sulfated bile acids,[58–60] was also frequently encountered, but with a lesser count than the non-sulfur counterparts. For the next set of analyses, we compiled all the MassQL-derived MS/MS spectra and merged identical spectra using MS-cluster[61] to reduce duplications and accelerate downstream computation (Table S1).

To search the GNPS repository and link spectra to tissues/biofluids, organisms, and health conditions, we conducted repository-scale searches using fastMASST,[62] a faster version of MASST (Mass Spectrometry Search Tool),[19] and searched consensus MS/MS spectra of the MassQL results. This enabled us to identify 3,144,357 MS/MS spectra from 150,577 files in 1,168 datasets, extending our search beyond Orbitrap data to include more public datasets. Using the MassQL-filtered Orbitrap spectra, fastMASST recovers MS/MS spectra acquired on many different mass spectrometers (Table S3). To characterize the distribution of these spectra in the body, we used the multi-species anatomy ontology, called Uber-anatomy ontology (UBERON),[63] curated in ReDU,[64] to identify associations between MS/MS spectra and organs or biofluids. The analysis revealed a wide distribution of candidate bile acid modifications from MassQL queries in rodents and humans (Figures 2A and 2B). This broad distribution of bile acids is consistent with detecting bile acid-associated proteins/transcripts and transporters in protein and transcriptomic atlases.[65,66] Furthermore, the unique mass shifts of bile acids (delta masses) are found in different organs and biofluids, as illustrated in the UpSet plots (Figures 2C and 2D; Table S3). The overlap between the delta masses across different body parts is represented as dots connected by a vertical line. Total counts of delta masses across different organs and biofluids in both rodents and humans are shown in Figures S2A and S2B, highlighting the abundance of these modified bile acids in organs like the brain. It should be noted that this does not represent a complete picture, as not every experimental condition, including nutritional state, interventional condition, or biological condition, is available in the GNPS/MassIVE public repository.

To highlight the utility of the bile acid MS/MS library, we illustrate two applications. Given the presence of bile amidates in microbial cultures,[39,40] we investigated whether our MassQL-derived library contained MS/MS spectra matching to microbially modified bile amidates. To accomplish this, we utilized a domain-specific MASST search tool, called microbeMASST.[67] In microbeMASST, query MS/MS spectra can be defined by universal spectrum identifier (USI)[68] lists (list of unique virtual paths to MS/MS spectra) or as MGF (Mascot Generic Format) spectra files to search against a curated reference database of LC-MS/MS data acquired from more than 60,000 monocultures of microorganisms, comprising 1,858 unique taxonomic lineages. Prior to the microbeMASST search, we wanted to refine our stage 1 MassQL queries further to only capture MS/MS spectra for bile amidates which undergo a specific loss of a charged ion and can be easily detected. To identify likely amidates, the stage 1 MassQL results were further filtered for an MS/MS fragment ion corresponding to the protonated ion loss of the amine/amino acid (Figures S3A and S3B). We called this a stage 2 MassQL query, which was generated by adding the detection of ion mass of the candidate amidate and calculated by subtracting the precursor mass from the mass of the unmodified bile acid. Redundant MS/MS spectra were then clustered using MS-cluster to produce consensus MS/MS spectra to reduce computation for the microbeMASST queries (Table S1).

A representative microbeMASST output from a single MS/MS search corresponding to a phenylalanine tri-hydroxylated bile amidate match is shown in Figure 3A. This MS/MS spectrum was found in Actinomycetia, Bacilli, Clostridia, and Gammaproteobacteria, bacterial classes abundant in the gut microbiota.[69] In total, 15,160 MS/MS spectra, corresponding to 92 potential modifications to bile acids, were matched to roughly one-fourth of the microbial cultures in the microbeMASST knowledge base, as shown in the taxonomic tree (Figure 3B). About two-thirds of the successful microbeMASST searches did not have MS/MS library matches to previously characterized microbial bile amidates, highlighting that many microbial bile acids remain to be discovered (Figure 3B). Matches to bacterial phyla frequently observed in the human gut microbiota—including Actinobacteria, Firmicutes, and Proteobacteria—were commonly observed. Nineteen and ten matches to phylum Bacteroidota and Fusobacteria (respectively) were also found, and less than ten matches were detected in the fungal phylum of Ascomycota (Figure 3C). Although amidation of bile acids with proteinogenic amino acids in bacteria has been demonstrated in multiple studies,[39,40] glycine conjugated bile acids have been shown to have been produced by the *Penicillium* genus.[70] To further support the presence of undiscovered microbial bile acids in animals, we reanalyzed a public dataset (GNPS/MassIVE:MSV000080918, https://massive.ucsd.edu/) from mouse fecal pellets publicly available since 2017.[71] This revealed that diet altered the types of bile acid modifications, with four delta masses more prevalent in normal chow (NC) and eighteen, including the taurine conjugates, being more prevalent in high-fat diet (HFD) (Figure 3D). Furthermore, all but

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

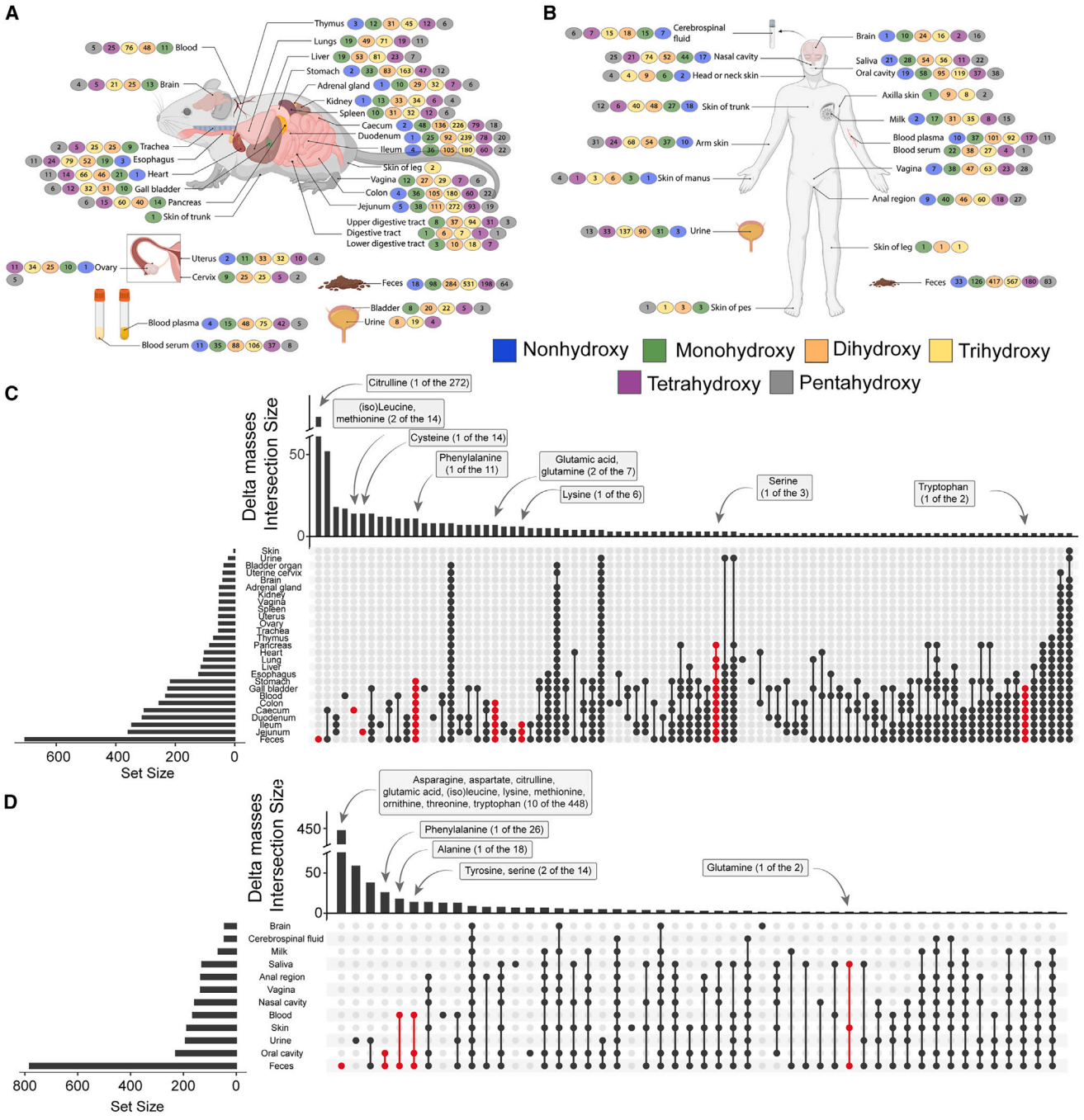
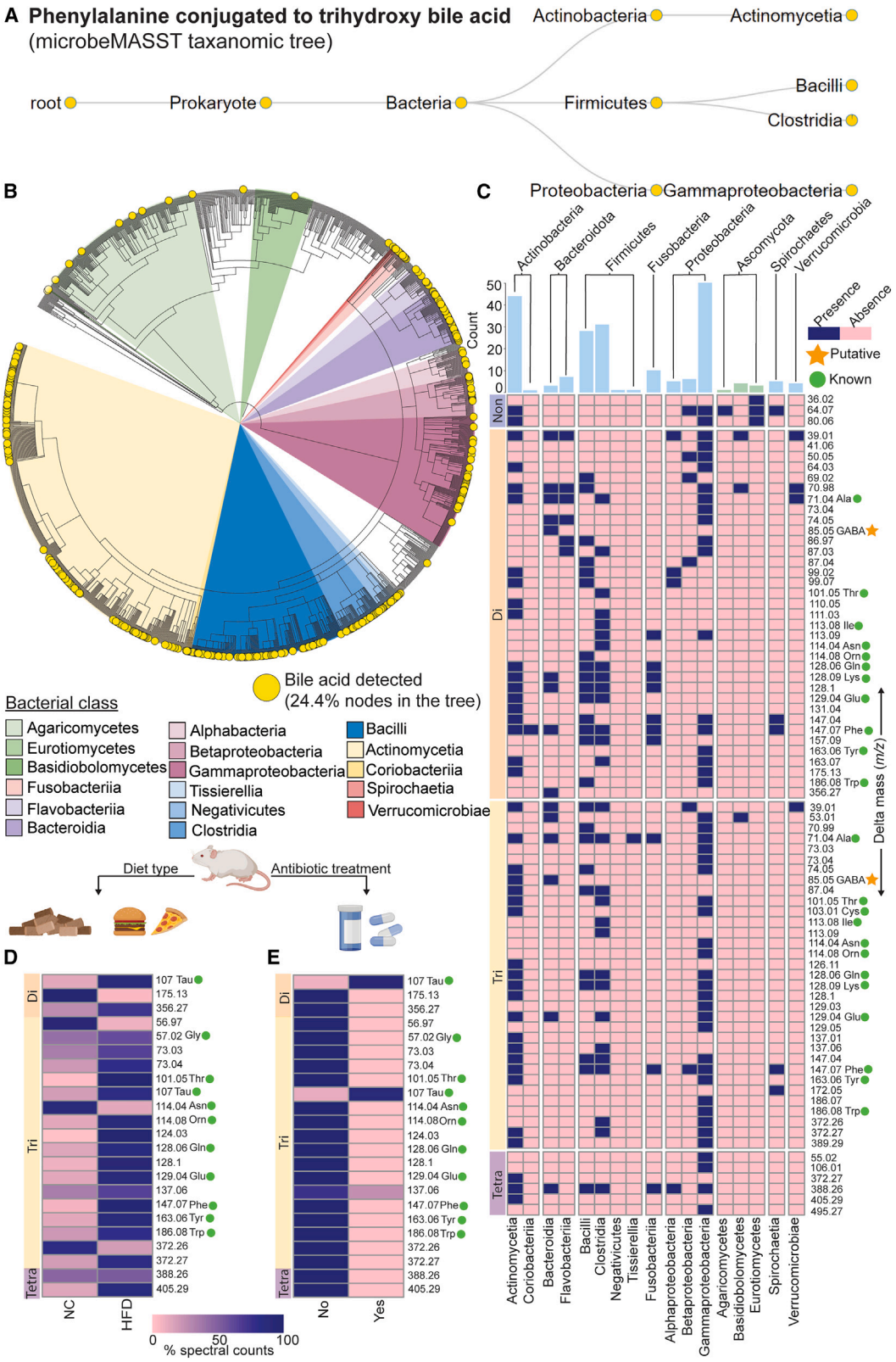


Figure 2. Distributions of delta masses obtained from fastMASST searches among different biofluids and organs

Summary of the delta mass occurrences in different body parts (organs, tissues, and biofluids) in (A) rodents and (B) humans. The numbers in each oval indicate the number of different delta masses, and colors represent the different bile acid classes. UpSet plots demonstrate the distribution of the delta masses within multiple body parts in (C) rodents and in (D) humans. Vertical columns in the UpSet plots represent the number of unique delta masses (intersection size), with their distribution in different body parts (horizontal columns on the left, called set size) shown as black dot(s). The intersections highlighted in red correspond to delta masses corresponding to modifications for which we have annotations, including proteinogenic amino acids. For example, in (C), the citrulline delta mass is 1 out of the 272 delta masses observed and is only observed in rodent feces and no other sample type, while the phenylalanine delta mass is one of 11 delta masses found in seven sample types, including stomach, gallbladder, blood, cecum, ileum, jejunum, and feces. (A) and (B) were created using BioRender.com. See Figure S2, Table S3, and Batsoyol et al, 2022.[62]

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

the host-derived taurine conjugates[72] were decreased after administration of an antibiotic cocktail (Figure 3E). It is possible for the bile amidates detected to be formed because of a spontaneous non-enzymatic cross-reactive reaction of amino acids and amines with bile acid coenzymes A (CoAs), and not catalyzed by the bile acid aminotransferase enzyme. The cross reaction of the amines during sample preparation is possible. If non-enzymatic addition was driving most of the reactions, then one would expect that most, if not all, amines and amino acids observed in the samples would be conjugated to bile acids. However, from our analysis of other amino acids and amines from the non-antibiotic and antibiotic-treated mice fecal samples, we observed that many amines, such as amino acids and peptides that are present in non-antibiotic-treated mice, are not conjugated to bile acids (Figure S3C). This observation is consistent with a selectivity in the amidation reaction and inconsistent with the non-enzymatic interpretation. Nonetheless, it is important for users to consider cross-reactivity as a possibility while interpreting results using the spectral resource.

The second example highlighting the relevance of our bile acid resource focuses on the discovery of previously unknown bile acids. Inspecting plausible annotations of the delta masses by matching to multiple databases (see STAR Methods), revealed several candidates from the polyamine biosynthetic pathway (Figure 4A; Table S4). Consistent with our observations that polyamines and their biosynthetic precursors are conjugated to bile acids, while this paper was in review, the polyamine, γ-aminobutyric acid (GABA), was shown in two preprints to be conjugated to deoxycholic acid by microbes belonging to species *Bacteroides fragilis*, *Mediterraneibacter gnavus*, *Bifidobacterium longum*, and *Bacteroides ovatus* isolated from the J-pouch of a patient suffering from ulcerative colitis pouchitis[74] and lithocholic GABA amidates and as ester when chenodeoxycholic acid or lithocholic acid were added to human stool, *ex vivo*.[75] Given that polyamines are involved in several cellular processes, including transcription, translation, stress protection, metabolism, growth, and aging, and that they are involved in many diseases, such as cancer and neurodegeneration,[76,77] we were interested in pursuing our investigation of these putative polyamine bile amidates. We further reanalyzed metabolite feature intensities for the candidate polyamine bile amidates, namely ornithine, putrescine, cadaverine, acetyl-ornithine, acetyl-cadaverine, and aminovalerate conjugates in an observational study dataset deposited in 2020 (GNPS/MassIVE: MSV000086131)[78] to reveal higher counts of spectral matches in carnivores compared with omnivores and a complete absence in herbivores (Figure 4B). Higher levels of taurine bile amidates were observed in carnivores, but there was no specific trend from carnivores to herbivores for glycine conjugates (Figure 4B). It is important to note that, at this stage, these are level 3 annotations,[79] and therefore, isomers cannot be distinguished, and only the use of synthetic standards and their retention time matching will provide additional insight to the type of bile acid. Although the possibility exists that due to differences in microbiome and genetics, these animals make different bile acids, the results above are also consistent with the formulation of a hypothesis that when more meat is consumed, the bile acids are elevated, or alternatively, that they are decreased with more plant consumption.

Given the higher number of polyamine amidates observed with di- and tri-hydroxylated bile acids, we first obtained their synthetic standards by conducting multiplexed organic synthesis[40] of cholic acid and chenodeoxycholic acid with the polyamine substrates and intermediates observed in the polyamine biosynthetic pathway (Figure 4A). MS/MS spectral matching satisfies level 2 annotation according to the 2007 metabolomics standards initiative,[79] and also level 2 probable structure, according to the Schymanski metabolomics annotation confidence levels[80] (Figure S4). By plotting MS/MS mirror plots, we observed concordance between fragment ions and their intensities in spectra recovered from MassQL searches in public data with the synthetic standard, providing additional support to our annotations of polyamine bile amidates. However, it should be noted that with MS/MS matching alone, we cannot distinguish between isomers with the same number of hydroxyl groups on the bile acid but at different positions and stereochemistry or a chimeric spectrum that contains data from more than one bile acid with the same molecular weight. An example would be the spectral overlap observed in tri-hydroxylated GABA amidate where the ion *m/z* 222.6593 in the spectra from public data has no mirrored ion in the standard (Figure S4). Therefore, at this point, we refer to them as di- or tri-hydroxylated bile amidates and not cholic or chenodeoxycholic conjugates, even though those were used in synthesizing the amidates. This is the highest level of annotation currently possible when analyzing public data with synthetic standards but without access to the physical specimen.

Level 1 identification of representative polyamine bile amidates requires matching retention times of candidate amidates in synthesized standards and samples across multiple chromatographic conditions. Hence, we contacted authors of datasets deposited in the public domain with spectral matches to the unknown bile acids

Figure 3. Evidence for microbial-derived bile acids

(A) A representative example of microbeMASST results with phenylalanine tri-hydroxylated bile amidate (MS/MS spectra ID: CCMSLIB00006582001).

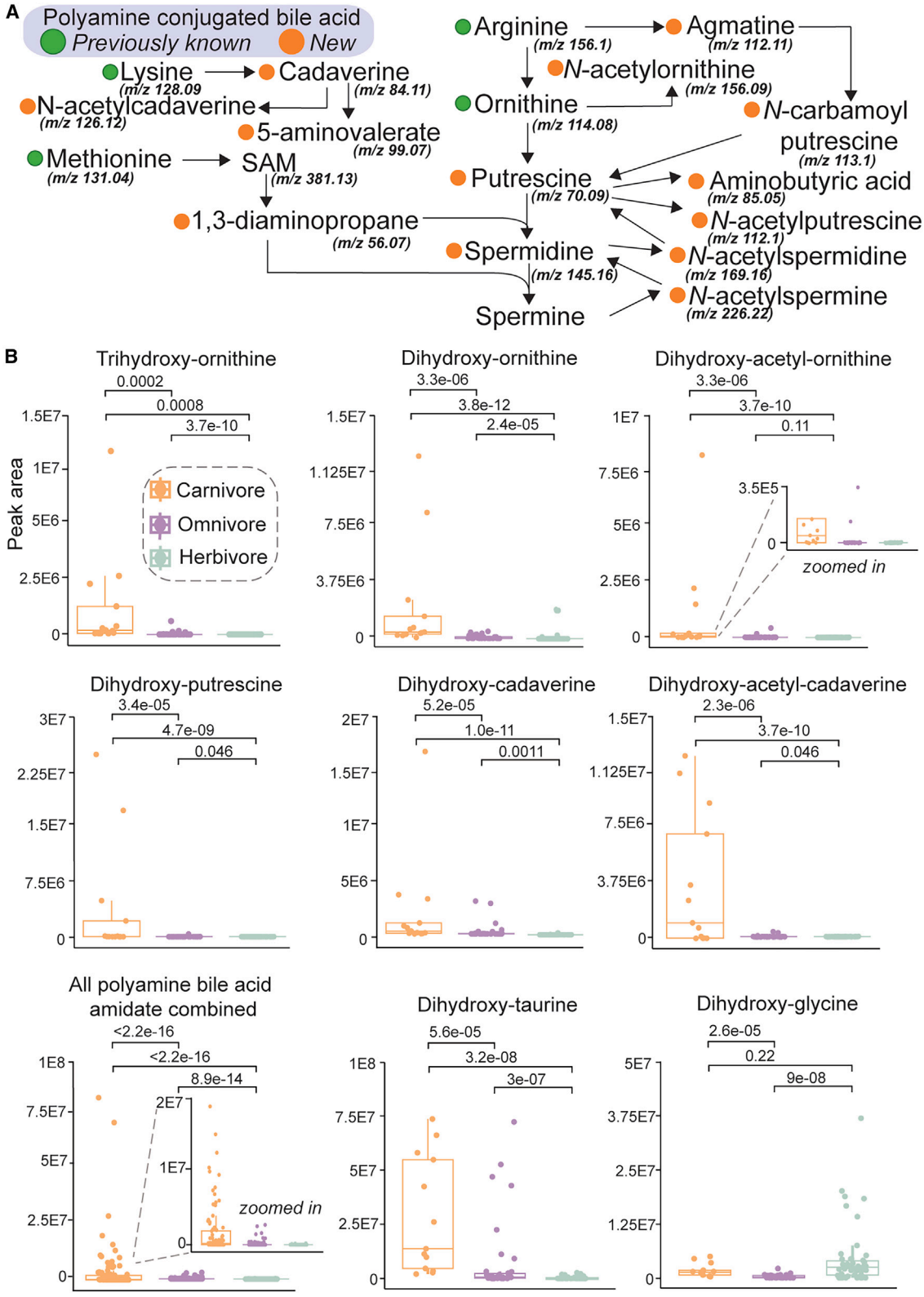
(B) Taxonomic distribution, at the class level, of the candidate bile acid amidate MS/MS spectra with matches in microbeMASST represented as a taxonomic tree generated in Interactive Tree of Life (iTOL).[73] The yellow circles on the nodes of the taxonomic tree highlight bacterial taxa where bile acid spectral matches were detected.

(C) Presence-/absence-based visualization of the bile amidates in axenic microbial/fungal cultures is highlighted for the delta masses from non-, di-, tri-, and tetra-hydroxylated bile acids. No matches were recovered from mono- and penta-hydroxylated MS/MS spectra. The count of delta masses is summarized at the phylum level using bar plots.

(D and E) Relative levels of delta masses represented by normalized counts of MS/MS spectra obtained from classical molecular networking with respect to (D) high-fat diet (HFD) or normal chow (NC) diet and (E) antibiotic use in the public dataset GNSP/MassIVE: MSV000080918. Bile acid MS/MS spectral matches to the GNPS reference library are represented with a green circle next to the delta mass. Icons were obtained from BioRender.com. See Figure S3.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

and requested the biological samples from the original study for validation. The labs we contacted had preserved samples in −80°C freezers. These samples were mice feces from a study conducted on social defeat stress (data deposited in 2017),[81] and feces from carnivores like lions, wolves, leopards, hyenas, and jungle cats (data deposited in 2020).[78] With synthetic standards and biological samples in hand, we performed retention time and MS/MS spectral matching on these animal samples under two chromatographic conditions and obtained additional validation by spiking the fecal extracts with the standards (Figures 5 and S5–S7; Table S4). In total, in a multiplexed fashion,[40] ten different cholic acid amidates with polyamines, namely agmatine, putrescine, *N*-acetyl-putrescine, *N*-carbamoyl-putrescine, spermidine, *N*-acetyl-spermidine, *N*-acetyl-spermine, cadaverine, *N*-acetyl-cadaverine, and 1,3-diaminopropane were synthesized (Figure 5A). Four of the ten cholic amidates synthesized—cholyl-putrescine, cholyl-spermidine, cholyl-cadaverine, and choyl-*N*-acetyl-cadaverine—were observed in both the mice and lion fecal samples (Figure 5B). The remaining six cholic amidates were only observed in the mice samples (Figure S5A). In addition, we multiplexed synthesized seven chenodeoxycholic amidates with putrescine, *N*-acetyl-putrescine, spermidine, *N*-acetyl-spermidine, cadaverine, and *N*-acetyl-cadaverine (Figure 5C), from which three amidates to *N*-acetyl-spermidine, *N*-acetyl-putrescine, and cadaverine were detected in both mice and lion feces (Figure 5D). Both the retention time and MS/MS spectral matched for *N*-acetyl-putrescine and *N*-acetyl-spermidine chenodeoxycholic amidates. There are additional peaks that are consistent with the presence of isomers for *N*-acetyl-spermidine and *N*-acetyl-putrescine conjugates (Figure 5D). Retention time for neither of the three cadaverine amidate peaks align with that of the synthetic standard, implying a different bile acid isomer. Three other amidates were detected in lion feces, but not in the mouse, with retention times matching the standards observed for two of the synthetic standards, putrescine and *N*-acetyl-cadaverine bile amidates (Figure S5B). The retention time for the cadaverine, spermidine, and 1,3-diaminopropane amidate with chenodeoxycholic acid did not match with the standard with two chromatographic conditions along with the ions in the MS/MS having different intensities, pointing toward a related but different isomer to be detected in the lion samples (Figures 5D, S5B, and S6C). In addition to lion feces, the polyamine bile amidates were detected in other carnivores such as tigers, sand cats, jungle cats, wolves, and hyenas, highlighting a broader presence of these polyamine bile amidates in animals with more meat in their diet (Table S4).

We next quantified the polyamine bile amidates in obligate carnivore samples, as they showed the highest signal intensities for the polyamine bile amidates of all the animals. Therefore, we quantified the amidates from the fecal samples accessible to us, which were obtained from an African lion, an African serval, a cheetah, two jaguars, three jaguarundis, and a tiger. To quantify, pure synthetic standards of individual polyamine bile amidates were obtained for cholyl-1,3-diaminopropane, chenodeoxycholyl-1,3-diaminopropane, cholyl-putrescine, chenodeoxycholyl-putrescine, cholyl-*N*-acetyl-putrescine, chenodeoxycholyl-*N*-acetyl-putrescine, cholyl-cadaverine, chenodeoxycholyl-cadaverine, cholyl-*N*-acetyl-cadaverine, chenodeoxycholyl-*N*-acetyl-cadaverine, and a mixture of two inseparable isomers of both cholyl-spermidine and chenodeoxycholyl-spermidine, as there are two non-identical primary amines on spermidine. Of these, putrescine, *N*-acetyl-putrescine, and cadaverine bile amidates were quantified in relation to the canonical glycine and taurine conjugates (Figure 6A; Table S5).

Quantities of the three polyamine bile amidates varied among the different animals studied, with their concentrations ranging from 0.05 to 59.71 ng/mg (dried weight). These ranges were 10-fold less or up to 8-fold more than what was observed for glycocholic acid that ranged from 3.4 to 8.5 ng/mg, but generally lower than what was observed for taurocholic acid, which ranged from 88.8 to 1,705 ng/mg. In the cheetah, which exhibited the highest quantities of the polyamine conjugates, the amount of cholyl-putrescine was within a third of the amount found for taurocholic acid at 58.3 ng/mg of feces, as opposed to 154.6 ng/mg for the taurocholic acid, but was more abundant compared with cholyl-glycine by approximately 10-fold.

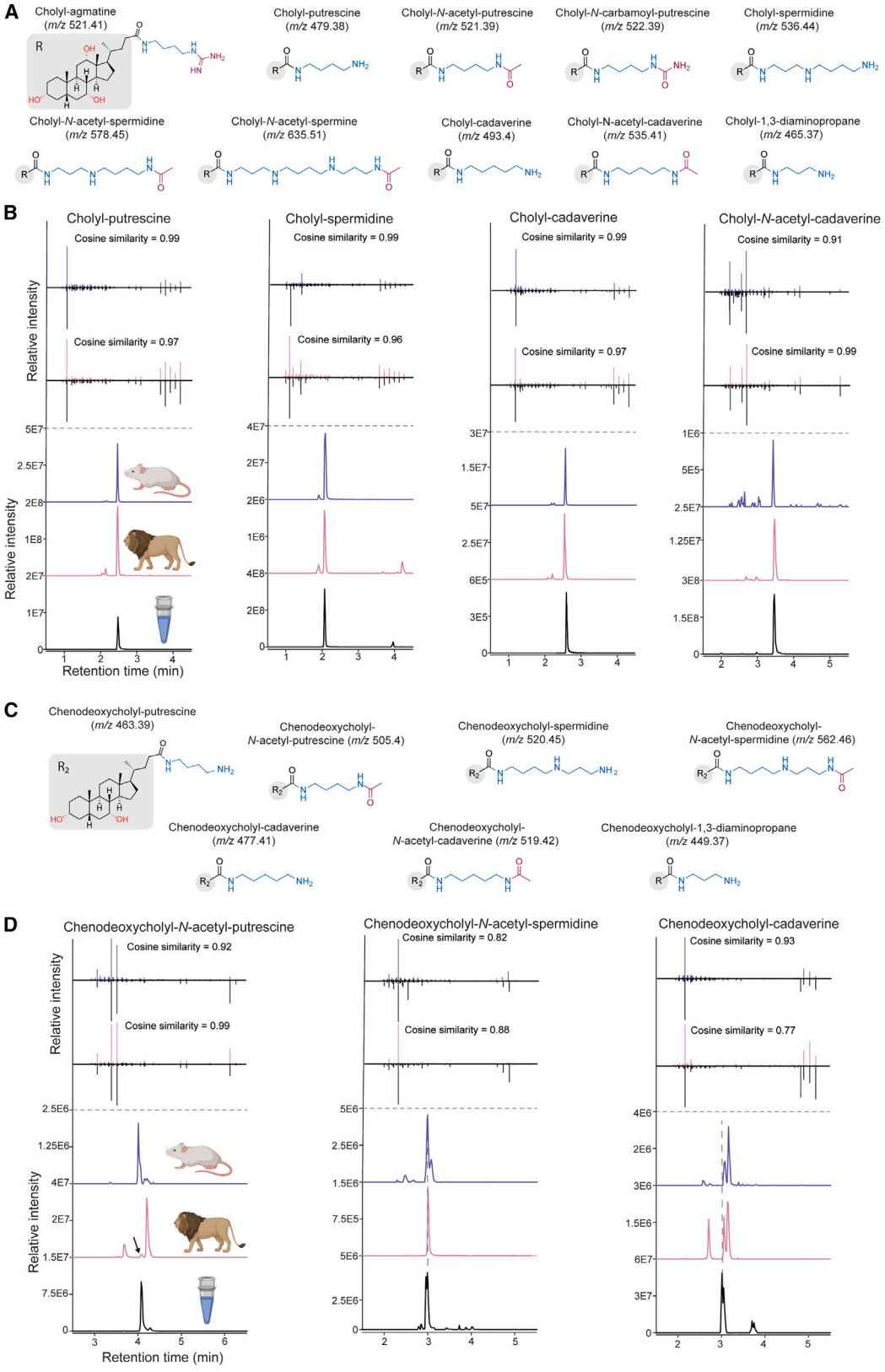
Considering the potential impact of diet on the levels of polyamine bile amidates in the animal studies, we proceeded to profile polyamine bile amidates in a human fecal dataset (GNPS/MassIVE: MSV000093005) obtained from a randomized, crossover, controlled feeding study. In this study, 36 volunteers were provided, in random order, with four different diets for 4 weeks each. This included an average American diet (AAD) and three Mediterranean (MED) diets with 0.5, 2.5, and 5.5 oz of lean beef[82] (Figure 6B) compared with the baseline diet (i.e., the participants' habitual diet) (Figures 6C and 6D). In these samples, the retention times of the amidates of Trp, Phe, Ser, Lys, Ala, Ser, and three polyamine bile amidates, aminovalerate, *N*-acetyl-putrescine, and the polyamine precursor lysine were matched. We compared their relative levels to the commonly studied taurine and glycine conjugates. The polyamine bile amidates, chenodeoxycholyl-aminovalerate and cholyl-*N*-acetyl-putrescine, decreased during the MED diets, irrespective of the amount of lean beef consumption. Similarly, other microbial bile acids, including Ala, Phe, Trp, and Ser-conjugated bile acids

Figure 4. Discovery of candidate polyamine bile amidates

(A) Polyamine biosynthetic pathway with an orange circle highlighting candidate polyamine and other intermediates and precursors in the polyamine biosynthetic pathway amidated to bile acids as recovered from matching precursor masses with various databases. Delta masses are shown in parentheses for each of the amidates known previously or recovered from public data.

(B) Sample-to-sample peak areas of the subset of polyamine bile amidates and two commonly known bile acids, taurine, and glycine amidates detected in a public dataset (GNPS/MassIVE: MSV000086131)[78] containing different animals with a carnivorous, omnivorous, and herbivorous diet. A higher abundance of the amidates was observed in carnivores. A non-parametric Kruskal-Wallis test followed by Wilcoxon was performed, and all p values were corrected for multiple comparisons using the Benjamini-Hochberg correction. Horizontal lines indicate the median value, the first (lower) and third (upper) quartiles are represented by the box edges, and vertical lines (whiskers) indicate the error range which is 1.5 times the interquartile range. Green circles in the figure represent library matches of the resource to MS/MS spectra in the GNPS reference library. See Figure S4 and Table S4.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019



(legend on next page)

decreased with the MED diets. The levels of chenodeoxycholyl-glycine decreased with the MED diet, irrespective of the amount of beef added to the diet, but this is not observed for cholyl-glycine. Taurine conjugated to both chenodeoxycholic and cholic acid did not change with diet alterations. These observations not only confirm the presence of polyamine bile amidates in humans but also highlight that they change with diet, with MED diets leading to an alteration in the levels of these bile amidates.

Although the polyamine amidates have not been documented previously to exist in biology, and the function of these polyamine bile amidates is currently unclear, there is evidence in the scientific literature that related synthetic polyamine bile amidates can be used for efficient and targeted delivery of drug molecules,[83–85] could potentially serve as treatment for metabolic syndromes[86] or function as inhibitors of bile acid transporters.[87]

DISCUSSION

In conclusion, this work uniquely focused on resource creation that allows the discovery of modifications to bile acids and related cholane structures. This highlights the underappreciated chemical diversity of bile acid modifications and provides the research community with a reusable reference MS/MS spectral library for future untargeted metabolomics experiments. Our study reveals that repository-scale searches can be used for the discovery of molecules. The resource will accelerate the discovery of unknown bile acids and their biological associations. Also, the matches to MS/MS spectra using this resource only enable the formulation of a hypothesis with respect to the modifications of the bile acids, and validation requires the use of either synthetic standards to conduct MS/MS spectral and retention time matching as was done for the polyamine bile amidates in this study, or isolation and structure elucidation using NMR, X-ray, or other techniques, and creative use of labeling methods.

Given the increase from the currently 97 known bile acid modifications that result in a detectable mass shift to thousands of candidate modifications highlighted in this study (without consideration of their stereochemistry or regiochemistry), we advocate for developing improved analytical technologies for use in bile acid research and for revising our understanding of the biological roles of bile acids.[88] Microbes can carry out a variety of modifications, including oxidation, reduction, dehydration, and isomerization, with the most recent addition being the amidation by microbes leading to diversification of bile acids.[35,38,89] Both fungi and bacteria can amidate bile acids with glycine (e.g., glycocholic acid), and now bacteria can amidate bile acids with other non-glycine and taurine- amines and amino acids.[38,40,70] The recent dual functionality of some of the BSHs lends support to the putative microbial origin of a subset of candidate bile amidates, and it is as yet unknown if the amino-transferase activity exists for the majority of the 2,456 BSH enzymes reported from human microbiome sequencing.[90] Although there is growing support to the microbial origin for many of the previously unknown candidate bile amidates, and other modifications such as hydrolysis, dehydrations, and oxidation have been well documented,[88,89] there is the possibility that a portion of these modifications associated with this resource is of non-enzymatic origin. This could be a chemical reaction during sample preparation or MS artifacts such as in-source fragments, adducts, etc. Therefore, it is critical to consider such possibilities when using this resource, especially if the modification has not yet been identified.[18,35,36,38–40,75,91–95] The biological functions of recently discovered microbial bile amidates still need investigation, which is being actively pursued by the larger scientific community. Studies are showing that the bile acids, including the amidates, interact with host ligand receptors such as farnesoid X receptor (FXR),[18,35] pregnane X receptor (PXR),[40,94,96] and play a role in T cell differentiation and dendritic signaling.[35,40,96,97] They also interact with aryl hydrocarbon receptor (AhR),[35] G protein-coupled bile acid receptor TGR5, and promote Wnt signaling and intestinal stem cell proliferation.[94] Microbial bile amidates can inhibit *C. difficile* spore germination and prevent its colonization.[36]

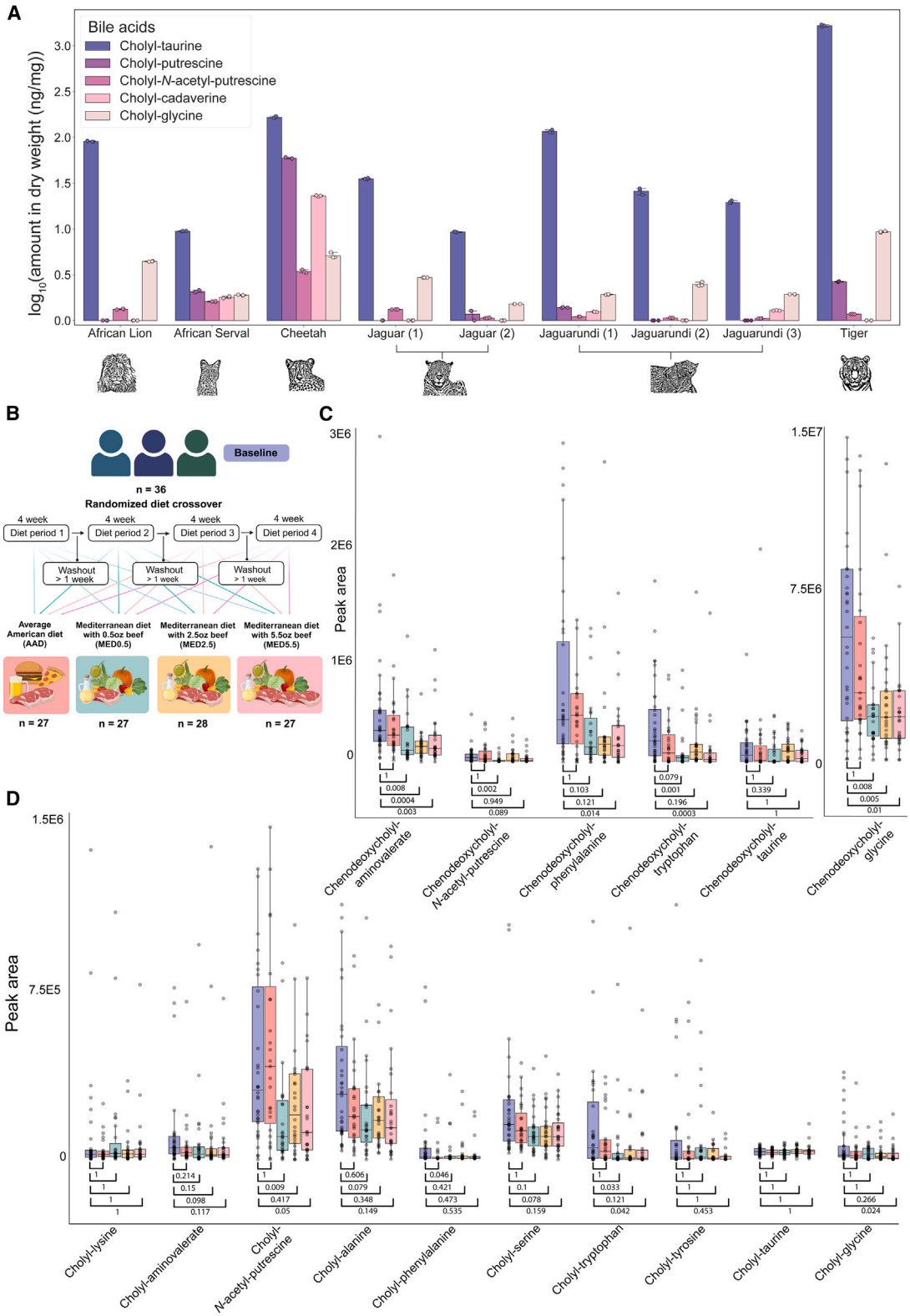
We hypothesize that the large diversity of modified bile acids uncovered in this work is part of signaling pathways from the gut to other organs. This hypothesis is supported by the presence of the bile acid receptors in different organs and bile acids throughout the body, but it also aligns with the proposed bile acid-gut-organ axis in the liver,[98] brain,[99] and muscles.[100,101] This exciting field of research in bile acid biology requires a holistic approach that warrants the analysis of bile acid modifications, as well as the conditions under which they are produced and the organisms that produce them. We anticipate that the MS/MS spectral resource generated in this study will assist in the discovery and continued exploration of the diverse functional roles of bile acid modifications.

To properly use our modified bile acid MS/MS library (also see data and code availability): matching of MS/MS spectra in experimental data with our modified bile acid MS/MS library will uniquely enable discovery of altered bile acids (and other bile-acid-like lipids), with the modification patterns providing a largely underappreciated view of bile acid biology. Matches to our spectral library can differentiate among non-, mono-,

Figure 5. Matching of polyamine bile amidates using synthetic standards

(A) Structures of all polyamine cholic amidates, (B) MS/MS spectra mirror matches and retention time matches to a subset of polyamine cholic acid amidates that were detected in both the lion and mouse studies, GNPS/MassIVE: MSV000080574 (blue) and GNPS/MassIVE: MSV000086131[78] (pink), with authentic synthetic standards. Similarly, for di-hydroxylated bile acids—(C) structures of all polyamine chenodeoxycholic amidates, (D) MS/MS spectra mirror matches, and retention time matches of the subset of polyamine chenodeoxycholic acid amidates detected in both the studies, MSV000080574 (blue) and MSV000086131 (pink), with synthetic standards. The red cross in the figure represents no detection of those polyamine bile amidates in lion or mice fecal samples. MS/MS mirror plots can be interactively visualized from the raw data in the GNPS dashboard the with the following links for the plots in the order they appear in the figure: cholyl-putrescine (top, bottom), cholyl-spermidine (top, bottom), cholyl-cadaverine (top, bottom), cholyl-*N*-acetyl-cadaverine (top, bottom), chenodeoxycholyl-*N*-acetyl-putrescine (top, bottom), chenodeoxycholyl-*N*-acetyl-spermidine (top, bottom), chenodeoxycholyl-cadaverine (top, bottom). Icons were obtained from BioRender.com. See Figures S5, S6, and S7 and Table S4.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019



(legend on next page)

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

di-, tri-, tetra-, and penta-hydroxylated bile acids, but distinguishing between regiospecific and stereospecific isomers will only be possible by retention time matching with standards. When considering the matches against this spectral library, it is important to understand that the MassQL query will not only find bile acids but also other cholane-containing structures, irrespective of their modifications on hydroxyl groups on the steroid core or the 5-carbon side chain, including the carboxylate. Atomic composition for approximately 60% of the MS/MS spectra in our library could be calculated using state-of-the-art molecular formula prediction tools.[48,49] The remaining 40% are multiple modifications (e.g., the addition of NH_3 and subtraction of OH results in a delta mass of 0.984 and cannot be explained by a single atomic composition), modifications with non-C, H, N, O, S, and P for which molecular formula prediction tools do not yet work well.[48,49] To validate atomic composition assignments predicted by the tools, end-users interested in using our spectral library for structure elucidation, are encouraged to carefully inspect the MS/MS spectrum, perhaps in comparison to the MS/MS spectrum of the conjugate portion proposed to be appended onto the bile acids (if available), to match reporter ions between the two spectra. If used judiciously, the spectral resource can provide unprecedented insights into the undiscovered biology of bile acids. Although shown here for bile acids, we anticipate that a similar workflow can be extended to other metabolite classes of interest; however, expert prior knowledge of MS/MS fragmentation patterns and an understanding of MassQL query design will be necessary. As a note to the readers, both our MassQL queries and the MS/MS spectral library can be reused to find bile acids from single LC-MS/MS data files or entire metabolomics projects. All the bile acid modifications recovered from MassQL queries are provided as an inclusion list for the users to include while acquiring data (Table S2).

Limitations of the study

The MS/MS resource likely does not capture all potential bile acid modifications and may contain other structurally similar cholane structures that can be captured by our MassQL queries due to their structurally similar core. For instance, if the modifications to the bile acids are either extensive or labile, the diagnostic bile acid MS/MS fragment ions we identified and used in the generation of our MassQL queries may not be present. This library is an underestimation of the diversity of bile acids that exist in biology. First, we did not search for all bile acids. Second, there are likely many modifications that have not yet been observed under any of the experimental conditions (e.g., sample preparation, extraction conditions, etc.) routinely used in the publicly accessible domain data. Leveraging current MassQL queries and expanding them to apply to future datasets can help reduce these limitations.

Although it has now been shown that microbial bile acids can be more abundant than their glycine counterparts,[40] bile acid concentrations are known to be regiospecific and dynamic. They can span a wide range in concentrations from over a hundred mM to pM or less.[88] Depending on the timing for detection and/or how commonly the bile acid is observed in biology, there is potential of low abundance for any bile acid candidates, and therefore it may not possible to obtain MS/MS spectrum and is missed as it is not triggered by the user's data acquisition settings. Although it is challenging to define a single set of parameters for acquiring MS/MS data, we encourage readers to refer to STAR Methods to use the acquisition parameters that led to the detection of polyamine-conjugated bile acids in this study as a starting point. Lastly, depending on the scale of the MassQL search queries, higher computational resources may also be required.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Development of MassQL queries
  - Calculation of delta mass and related analysis
  - Proposed annotation of delta masses at MS1 level
  - MS/MS spectral clustering
  - *In silico* annotations of candidate modifications
  - Repository scale MS/MS spectral analysis with fastMASST
  - Repository scale MS/MS spectral analysis with microbeMASST
  - Reanalysis of public data from GNPS/MassIVE
  - Multiplexed reaction for polyamine bile amidates
  - Co-migration analysis with multiplexed synthetic standard reaction mixtures
  - Extraction and LC-MS/MS profiling of human feces from diet cross over study
  - Synthesis of pure polyamine bile amidates
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Quantification of polyamines in feline feces
  - LC-MS/MS method validation for polyamines bile amidates quantification
  - Statistical analyses
- ADDITIONAL RESOURCES

Figure 6. Quantification of polyamine bile amidates in felines and observation of amidates in a randomized human diet study

(A) Absolute abundance of polyamine bile amidates in ng/mg dry weight of fecal material as quantified in different feline species.

(B) Study design for randomized diet crossover study in 36 human participants. Color scheme from (B) for the diet types is carried forward to (C) and (D).

(C and D) Peak area abundances of different polyamine amidated to (C) chenodeoxycholic acid and (D) cholic acid as observed in humans with four different diet treatments compared with the baseline information. The boxplots show first (lower) quartile, median, and third (upper) quartile, and whiskers are 1.5 times the interquartile range.The non-parametric Wilcoxon test was performed, and p values were Bonferroni corrected. Icons were obtained from BioRender.com. See Table S5.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.02.019.

ACKNOWLEDGMENTS

We thank the support by NIH for the NIH collaborative microbial metabolite center U24DK133658; the development of tools for structure elucidation R01GM107550; harmonization of metabolomics metadata across repositories R03OD034493; and Alzheimer's gut microbiome project U19AG063744 and NIH U01DK119702, S10OD021750, and BBSRC/NSF 2152526. A.T.A. was supported by the Gordon & Betty Moore Foundation. S.X. and T.H. were supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN-2020-04895). C.W.T. was supported by the Max Planck Society. This work was supported by the DIP (2476/2-1) to I. Mizrahi, the ERC (866530) to I. Mizrahi, and the ISF (1947/19) to I. Mizrahi and S.M. The human study was funded by the Beef Checkoff. This study was also supported by the USDA, ARS, and the Penn State Clinical and Translational Research Institute, the Pennsylvania State University Clinical and Translational Science Award, and the NIH/National Center for Advancing Translational Sciences grant no. UL1TR000127.

AUTHOR CONTRIBUTIONS

Conceptualization, P.C.D.; methodology, P.C.D., I. Mohanty, H.M.-R., A.T.A., W.B., M.W., and V.B.M.; formal analysis, I. Mohanty, H.M.-R., Y.E.A., S.X., T.H., and P.C.D.; investigation, I. Mohanty, H.M.-R., J.V.S., D.S., D.D., S.A., I.K., A.D.P., R.S., S.Z., F.V., and J.Z.; resources, P.M.K.-E., K.S.P., J.A.F., C.W.T., O.E.T.-H., S.M., I. Mizrahi, M.W., S.Z., and R.S.; writing – original draft, P.C.D., I. Mohanty, H.M.-R., and L.R.H.; writing – review & editing, all authors; supervision, P.C.D., funding acquisition, P.C.D.

DECLARATION OF INTERESTS

P.C.D. is an advisor and holds equity in Cybele and Sirenas and a scientific co-founder, advisor, and holds equity in Ometa, Enveda, and Arome with prior approval by UC San Diego. P.C.D. also consulted for DSM animal health in 2023. M.W. is a co-founder of Ometa Labs LLC.

Received: April 26, 2023
Revised: November 30, 2023
Accepted: February 15, 2024
Published: March 11, 2024

REFERENCES

1. Parks, D.J., Blanchard, S.G., Bledsoe, R.K., Chandra, G., Consler, T.G., Kliewer, S.A., Stimmel, J.B., Willson, T.M., Zavacki, A.M., Moore, D.D., et al. (1999). Bile acids: natural ligands for an orphan nuclear receptor. Science *284*, 1365–1368.
2. Makishima, M., Okamoto, A.Y., Repa, J.J., Tu, H., Learned, R.M., Luk, A., Hull, M.V., Lustig, K.D., Mangelsdorf, D.J., and Shan, B. (1999). Identification of a nuclear receptor for bile acids. Science *284*, 1362–1365.
3. Hang, S., Paik, D., Yao, L., Kim, E., Trinath, J., Lu, J., Ha, S., Nelson, B.N., Kelly, S.P., Wu, L., et al. (2019). Bile acid metabolites control TH17 and Treg cell differentiation. Nature *576*, 143–148.
4. Paik, D., Yao, L., Zhang, Y., Bae, S., D'Agostino, G.D., Zhang, M., Kim, E., Franzosa, E.A., Avila-Pacheco, J., Bisanz, J.E., et al. (2022). Human gut bacteria produce TH17-modulating bile acid metabolites. Nature *603*, 907–912.
5. Song, X., Sun, X., Oh, S.F., Wu, M., Zhang, Y., Zheng, W., Geva-Zatorsky, N., Jupp, R., Mathis, D., Benoist, C., et al. (2020). Microbial bile acid metabolites modulate gut RORγ+ regulatory T cell homeostasis. Nature *577*, 410–415.
6. Campbell, C., McKenney, P.T., Konstantinovsky, D., Isaeva, O.I., Schizas, M., Verter, J., Mai, C., Jin, W.-B., Guo, C.-J., Violante, S., et al. (2020). Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. Nature *581*, 475–479.
7. Li, W., Hang, S., Fang, Y., Bae, S., Zhang, Y., Zhang, M., Wang, G., McCurry, M.D., Bae, M., Paik, D., et al. (2021). A bacterial bile acid metabolite modulates Treg activity through the nuclear hormone receptor NR4A1. Cell Host Microbe *29*, 1366–1377.e9.
8. Shapiro, H., Kolodziejczyk, A.A., Halstuch, D., and Elinav, E. (2018). Bile acids in glucose metabolism in health and disease. J. Exp. Med. *215*, 383–396.
9. Zhao, X., Liu, Z., Sun, F., Yao, L., Yang, G., and Wang, K. (2022). Bile acid detection techniques and bile acid-related diseases. Front. Physiol. *13*, 826740.
10. Đanić, M., Stanimirov, B., Pavlović, N., Goločorbin-Kon, S., Al-Salami, H., Stankov, K., and Mikov, M. (2018). Pharmacological applications of bile acids and their derivatives in the treatment of metabolic syndrome. Front. Pharmacol. *9*, 1382.
11. Office of the Commissioner (2022). FDA Approves New Treatment Option for Patients with ALS (U.S. Food and Drug Administration). https://www.fda.gov/news-events/press-announcements/fda-approves-new-treatment-option-patients-als.
12. Russell, D.W., and Setchell, K.D. (1992). Bile acid biosynthesis. Biochemistry *31*, 4737–4749.
13. Carey, M.C. (1984). Bile acids and bile salts: ionization and solubility properties. Hepatology *4*, 66S–71S.
14. Hofmann, A.F. (1999). The continuing importance of bile acids in liver and intestinal disease. Arch. Intern. Med. *159*, 2647–2658.
15. McRae, M., Rezk, N.L., Bridges, A.S., Corbett, A.H., Tien, H.-C., Brouwer, K.L.R., and Kashuba, A.D.M. (2010). Plasma bile acid concentrations in patients with human immunodeficiency virus infection receiving protease inhibitor therapy: possible implications for hepatotoxicity. Pharmacotherapy *30*, 17–24.
16. Steiner, C., Othman, A., Saely, C.H., Rein, P., Drexel, H., von Eckardstein, A., and Rentsch, K.M. (2011). Bile acid metabolites in serum: intraindividual variation and associations with coronary heart disease, metabolic syndrome and diabetes mellitus. PLoS One *6*, e25006.
17. Hamilton, J.P., Xie, G., Raufman, J.-P., Hogan, S., Griffin, T.L., Packard, C.A., Chatfield, D.A., Hagey, L.R., Steinbach, J.H., and Hofmann, A.F. (2007). Human cecal bile acids: concentration and spectrum. Am. J. Physiol. Gastrointest. Liver Physiol. *293*, G256–G263.
18. Quinn, R.A., Melnik, A.V., Vrbanac, A., Fu, T., Patras, K.A., Christy, M.P., Bodai, Z., Belda-Ferre, P., Tripathi, A., Chung, L.K., et al. (2020). Global chemical effects of the microbiome include new bile-acid conjugations. Nature *579*, 123–129.
19. Wang, M., Jarmusch, A.K., Vargas, F., Aksenov, A.A., Gauglitz, J.M., Weldon, K., Petras, D., da Silva, R., Quinn, R., Melnik, A.V., et al. (2020). Mass spectrometry searches using MASST. Nat. Biotechnol. *38*, 23–26.
20. Hoffmann, M.A., Nothias, L.-F., Ludwig, M., Fleischauer, M., Gentry, E.C., Witting, M., Dorrestein, P.C., Dührkop, K., and Böcker, S. (2022). High-confidence structural annotation of metabolites absent from spectral libraries. Nat. Biotechnol. *40*, 411–421.
21. Petras, D., Caraballo-Rodríguez, A.M., Jarmusch, A.K., Molina-Santiago, C., Gauglitz, J.M., Gentry, E.C., Belda-Ferre, P., Romero, D., Tsunoda, S.M., Dorrestein, P.C., et al. (2021). Chemical proportionality within molecular networks. Anal. Chem. *93*, 12833–12839.
22. Hofmann, A.F., Sjövall, J., Kurz, G., Radominska, A., Schteingart, C.D., Tint, G.S., Vlahcevic, Z.R., and Setchell, K.D. (1992). A proposed nomenclature for bile acids. J. Lipid Res. *33*, 599–604.
23. Peric-Golia, L., and Jones, R.S. (1962). Ornithocholanic acids–abnormal conjugates of bile acids. Proc. Soc. Exp. Biol. Med. *110*, 327–331.

24. Peric-Golia, L., and Jones, R.S. (1963). Ornithocholanic acids and cholelithiasis in man. Science *142*, 245–246.

25. Hagey, L.R., Schteingart, C.D., Rossi, S.S., Ton-Nu, H.T., and Hofmann, A.F. (1998). An N-acyl glycyltaurine conjugate of deoxycholic acid in the biliary bile acids of the rabbit. J. Lipid Res. *39*, 2119–2124.

26. Myher, J.J., Marai, L., Kuksis, A., Yousef, I.M., and Fisher, M.M. (1975). Identification of ornithine and arginine conjugates of cholic acid by mass spectrometry. Can. J. Biochem. *53*, 583–590.

27. Sweeny, D.J., Barnes, S., Heggie, G.D., and Diasio, R.B. (1987). Metabolism of 5-fluorouracil to an N-cholyl-2-fluoro-beta-alanine conjugate: previously unrecognized role for bile acids in drug conjugation. Proc. Natl. Acad. Sci. USA *84*, 5439–5443.

28. Hagey, L.R., Møller, P.R., Hofmann, A.F., and Krasowski, M.D. (2010). Diversity of bile salts in fish and amphibians: evolution of a complex biochemical pathway. Physiol. Biochem. Zool. *83*, 308–321.

29. Satoh Née Okihara, R., Saito, T., Ogata, H., Ohsaki, A., Iida, T., Asahina, K., Mitamura, K., Ikegawa, S., Hofmann, A.F., and Hagey, L.R. (2014). N-Methyltaurine N-acyl amidated bile acids and deoxycholic acid in the bile of angelfish (Pomacanthidae): a novel bile acid profile in Perciform fish. Steroids *80*, 15–23.

30. Une, M., Goto, T., Kihira, K., Kuramoto, T., Hagiwara, K., Nakajima, T., and Hoshita, T. (1991). Isolation and identification of bile salts conjugated with cysteinolic acid from bile of the red seabream, Pagrosomus major. J. Lipid Res. *32*, 1619–1623.

31. Neugebauer, K.A., Okros, M., Guzior, D.V., Feiner, J., Chargo, N.J., Rzepka, M., Schilmiller, A.L., O'Reilly, S., Jones, A.D., Watson, V.E., et al. (2022). Baat gene knockout alters post-natal development, the gut microbiome, and reveals unusual bile acids in mice. J. Lipid Res. *63*, 100297.

32. Yang, Z.-H., Liu, F., Zhu, X.-R., Suo, F.-Y., Jia, Z.-J., and Yao, S.-K. (2021). Altered profiles of fecal bile acids correlate with gut microbiota and inflammatory responses in patients with ulcerative colitis. World J. Gastroenterol. *27*, 3609–3629.

33. Mouzaki, M., Wang, A.Y., Bandsma, R., Comelli, E.M., Arendt, B.M., Zhang, L., Fung, S., Fischer, S.E., McGilvray, I.G., and Allard, J.P. (2016). Bile acids and dysbiosis in non-alcoholic fatty liver disease. PLoS One *11*, e0151829.

34. Lamichhane, S., Sen, P., Dickens, A.M., Alves, M.A., Härkönen, T., Honkanen, J., Vatanen, T., Xavier, R.J., Hyötyläinen, T., Knip, M., et al. (2022). Dysregulation of secondary bile acid metabolism precedes islet autoimmunity and type 1 diabetes. Cell Rep. Med. *3*, 100762.

35. Rimal, B., Collins, S., Tanes, C., Rocha, E., Granda, M., Solanka, S., Hoque, N., Gentry, E., Koo, I., Reilly, E., et al. (2024). Bile salt hydrolase catalyses formation of amine conjugated bile acids. Nature *626*, 859–863.

36. Foley, M.H., Walker, M.E., Stewart, A.K., O'Flaherty, S., Gentry, E.C., Patel, S., Beaty, V.V., Allen, G., Pan, M., Simpson, J.B., et al. (2023). Bile salt hydrolases shape the bile acid landscape and restrict Clostridioides difficile growth in the murine gut. Nat. Microbiol. *8*, 611–628.

37. Tam, J., Icho, S., Utama, E., Orrell, K.E., Gómez-Biagi, R.F., Theriot, C.M., Kroh, H.K., Rutherford, S.A., Lacy, D.B., and Melnyk, R.A. (2020). Intestinal bile acids directly modulate the structure and function of C. difficile TcdB toxin. Proc. Natl. Acad. Sci. USA *117*, 6792–6800.

38. Guzior, D., Okros, M., Shivel, M., Armwald, B., Bridges, C., Fu, Y., Martin, C., Schilmiller, A., Miller, R., et al. (2024). Bile salt hydrolase acyltransferase activity expands bile acid diversity. Nature *626*, 852–858.

39. Lucas, L.N., Barrett, K., Kerby, R.L., Zhang, Q., Cattaneo, L.E., Stevenson, D., Rey, F.E., and Amador-Noguez, D. (2021). Dominant bacterial phyla from the human gut show widespread ability to transform and conjugate bile acids. mSystems, e0080521.

40. Gentry, E., Collins, S., Panitchpakdi, M., Belda-Ferre, P., Stewart, A., Terrazas, M., Lu, H., Zuffa, S., Yan, T., et al. (2024). Reverse metabolomics for the discovery of chemical structures from humans. Nature *626*, 419–426.

41. Wang, Y.-Z., Mei, P.-C., Bai, P.-R., An, N., He, J.-G., Wang, J., Zhu, Q.-F., and Feng, Y.-Q. (2022). Discovery and identification of new amino acid-conjugated bile acids by polarity-switching multiple reaction monitoring mass spectrometry. Preprint at ChemRxiv. https://doi.org/10.26434/chemrxiv-2022-rdjcx.

42. Zheng, X., Smith, F.B., Aly, N.A., Cai, J., Smith, R.D., Patterson, A.D., and Baker, E.S. (2019). Evaluating the structural complexity of isomeric bile acids with ion mobility spectrometry. Anal. Bioanal. Chem. *411*, 4673–4682.

43. Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. *34*, 828–837.

44. Wang, Y.-Z., Chen, Y.-Y., Wu, X.-Z., Bai, P.-R., An, N., Liu, X.-L., Zhu, Q.-F., and Feng, Y.-Q. (2023). Uncovering the carboxylated metabolome in gut microbiota-host co-metabolism: A chemical derivatization-molecular networking approach. Anal. Chem. *95*, 11550–11557.

45. Bittremieux, W., Schmid, R., Huber, F., van der Hooft, J.J.J., Wang, M., and Dorrestein, P.C. (2022). Comparison of cosine, modified cosine, and neutral loss based spectrum alignment for discovery of structurally related molecules. J. Am. Soc. Mass Spectrom. *33*, 1733–1744.

46. Jarmusch, A.K., Aron, A.T., Petras, D., Phelan, V.V., Bittremieux, W., Acharya, D.D., Ahmed, M.M.A., Bauermeister, A., Bertin, M.J., Boudreau, P.D., et al. (2022). A universal language for finding mass spectrometry data patterns. Preprint at bioRxiv. https://doi.org/10.1101/2022.08.06.503000.

47. Scheubert, K., Hufsky, F., Petras, D., Wang, M., Nothias, L.-F., Dührkop, K., Bandeira, N., Dorrestein, P.C., and Böcker, S. (2017). Significance estimation for large scale metabolomics annotations by spectral matching. Nat. Commun. *8*, 1494.

48. Dührkop, K., Fleischauer, M., Ludwig, M., Aksenov, A.A., Melnik, A.V., Meusel, M., Dorrestein, P.C., Rousu, J., and Böcker, S. (2019). SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat. Methods *16*, 299–302.

49. Xing, S., Shen, S., Xu, B., and Li X. , Huan, T. (2023). BUDDY: molecular formula discovery via bottom-up MS/MS interrogation. Nat. Methods. *20*, 881–890.

50. Marschall, H.U., Griffiths, W.J., Götze, U., Zhang, J., Wietholtz, H., Busch, N., Sjövall, J., and Matern, S. (1994). The major metabolites of ursodeoxycholic acid in human urine are conjugated with N-acetylglucosamine. Hepatology *20*, 845–853.

51. Perreault, M., Białek, A., Trottier, J., Verreault, M., Caron, P., Milkiewicz, P., and Barbier, O. (2013). Role of glucuronidation for hepatic detoxification and urinary elimination of toxic bile acids during biliary obstruction. PLoS One *8*, e80994.

52. Marschall, H.U., Matern, H., Wietholtz, H., Egestad, B., Matern, S., and Sjövall, J. (1992). Bile acid N-acetylglucosaminidation. In vivo and in vitro evidence for a selective conjugation reaction of 7 beta-hydroxylated bile acids in humans. J. Clin. Invest. *89*, 1981–1987.

53. Takei, H., Narushima, S., Suzuki, M., Kakiyama, G., Sasaki, T., Murai, T., Yamashiro, Y., and Nittono, H. (2022). Characterization of long-chain fatty acid-linked bile acids: a major conjugation form of 3β-hydroxy bile acids in feces. J. Lipid Res. *63*, 100275.

54. Baldessari, A., and García Liñares, G. (2018). Lipase-catalyzed acetylation and esterification of bile acids. Methods Mol. Biol. *1835*, 337–350.

55. Mackenzie, P., Little, J.M., and Radominska-Pandya, A. (2003). Glucosidation of hyodeoxycholic acid by UDP-glucuronosyltransferase 2B7. Biochem. Pharmacol. *65*, 417–421.

56. Matern, S., Matern, H., Farthmann, E.H., and Gerok, W. (1984). Hepatic and extrahepatic glucuronidation of bile acids in man. Characterization of bile acid uridine 5′-diphosphate-glucuronosyltransferase in hepatic, renal, and intestinal microsomes. J. Clin. Invest. *74*, 402–410.

57. Trottier, J., Perreault, M., Rudkowska, I., Levy, C., Dallaire-Theroux, A., Verreault, M., Caron, P., Staels, B., Vohl, M.-C., Straka, R.J., et al. (2013). Profiling serum bile acid glucuronides in humans: gender divergences, genetic determinants, and response to fenofibrate. Clin. Pharmacol. Ther. *94*, 533–543.

58. Alnouti, Y. (2009). Bile Acid sulfation: a pathway of bile acid elimination and detoxification. Toxicol. Sci. *108*, 225–246.

59. Alvelius, G., Hjalmarson, O., Griffiths, W.J., Björkhem, I., and Sjövall, J. (2001). Identification of unusual 7-oxygenated bile acid sulfates in a patient with Niemann-Pick disease, type C. J. Lipid Res. *42*, 1571–1577.

60. Abdel-Khalik, J., Hearn, T., Dickson, A.L., Crick, P.J., Yutuc, E., Austin-Muttitt, K., Bigger, B.W., Morris, A.A., Shackleton, C.H., Clayton, P.T., et al. (2021). Bile acid biosynthesis in Smith-Lemli-Opitz syndrome bypassing cholesterol: potential importance of pathway intermediates. J. Steroid Biochem. Mol. Biol. *206*, 105794.

61. Frank, A.M., Monroe, M.E., Shah, A.R., Carver, J.J., Bandeira, N., Moore, R.J., Anderson, G.A., Smith, R.D., and Pevzner, P.A. (2011). Spectral archives: extending spectral libraries to analyze both identified and unidentified spectra. Nat. Methods *8*, 587–591.

62. Batsoyol, N., Pullman, B., Wang, M., Bandeira, N., and Swanson, S. (2022). P-massive: A real-time search engine for a multi-terabyte mass spectrometry database. In SC22: International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1–15.

63. Mungall, C.J., Torniai, C., Gkoutos, G.V., Lewis, S.E., and Haendel, M.A. (2012). Uberon, an integrative multi-species anatomy ontology. Genome Biol. *13*, R5.

64. Jarmusch, A.K., Wang, M., Aceves, C.M., Advani, R.S., Aguirre, S., Aksenov, A.A., Aleti, G., Aron, A.T., Bauermeister, A., Bolleddu, S., et al. (2020). ReDU: a framework to find and reanalyze public mass spectrometry data. Nat. Methods *17*, 901–904.

65. Baldarelli, R.M., Smith, C.M., Finger, J.H., Hayamizu, T.F., McCright, I.J., Xu, J., Shaw, D.R., Beal, J.S., Blodgett, O., Campbell, J., et al. (2021). The mouse Gene Expression Database (GXD): 2021 update. Nucleic Acids Res. *49*, D924–D931.

66. Sjöstedt, E., Zhong, W., Fagerberg, L., Karlsson, M., Mitsios, N., Adori, C., Oksvold, P., Edfors, F., Limiszewska, A., Hikmet, F., et al. (2020). An atlas of the protein-coding genes in the human, pig, and mouse brain. Science *367*, eaay5947.

67. Zuffa, S., Schmid, R., Bauermeister, A., Gomes, P.W.P., Caraballo-Rodriguez, A.M., El Abiead, Y., Aron, A.T., Gentry, E.C., Zemlin, J., Meehan, M.J., et al. (2024). microbeMASST: a taxonomically informed mass spectrometry search tool for microbial metabolomics data. Nat. Microbiol. *9*, 336–345.

68. Deutsch, E.W., Perez-Riverol, Y., Carver, J., Kawano, S., Mendoza, L., Van Den Bossche, T., Gabriels, R., Binz, P.-A., Pullman, B., Sun, Z., et al. (2021). Universal Spectrum Identifier for mass spectra. Nat. Methods *18*, 768–770.

69. Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G.A.D., Gasbarrini, A., and Mele, M.C. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms *7*, 14.

70. Ohashi, K., Miyagawa, Y., Nakamura, Y., and Shibuya, H. (2008). Bioproduction of bile acids and the glycine conjugates by Penicillium fungus. J. Nat. Med. *62*, 83–86.

71. Shalapour, S., Lin, X.-J., Bastian, I.N., Brain, J., Burt, A.D., Aksenov, A.A., Vrbanac, A.F., Li, W., Perkins, A., Matsutani, T., et al. (2017). Inflammation-induced IgA+ cells dismantle anti-liver cancer immunity. Nature *551*, 340–345.

72. Pellicoro, A., van den Heuvel, F.A.J., Geuken, M., Moshage, H., Jansen, P.L.M., and Faber, K.N. (2007). Human and rat bile acid-CoA:amino acid N-acyltransferase are liver-specific peroxisomal enzymes: implications for intracellular bile salt transport. Hepatology *45*, 340–348.

73. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296.

74. Mullowney, M.W., Fiebig, A., Schnizlein, M.K., McMillin, M., Koval, J., Rubin, D., Dalal, S., Sogin, M.L., Chang, E.B., Sidebottom, A.M., et al. (2023). Microbially-catalyzed conjugation of GABA and tyramine to bile acids. Preprint at bioRxiv. https://doi.org/10.1101/2023.09.25.559407.

75. Garcia, C.J., García-Villalba, R., Beltrán, D., Frutos-Lisón, M.D., and Tomás-Barberán, F.A. (2023). MS/MS fragmentation pattern analysis confirms the production of the new esterified bile acids by the human gut microbiota. Preprint at bioRxiv. https://doi.org/10.1101/2023.11.07.564921.

76. Miller-Fleming, L., Olin-Sandoval, V., Campbell, K., and Ralser, M. (2015). Remaining mysteries of molecular biology: the role of polyamines in the cell. J. Mol. Biol. *427*, 3389–3406.

77. Lenis, Y.Y., Elmetwally, M.A., Maldonado-Estrada, J.G., and Bazer, F.W. (2017). Physiological importance of polyamines. Zygote *25*, 244–255.

78. Gregor, R., Probst, M., Eyal, S., Aksenov, A., Sasson, G., Horovitz, I., Dorrestein, P.C., Meijler, M.M., and Mizrahi, I. (2022). Mammalian gut metabolomes mirror microbiome composition and host phylogeny. ISME J. *16*, 1262–1274.

79. Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W.-M., Fiehn, O., Goodacre, R., Griffin, J.L., et al. (2007). Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics *3*, 211–221.

80. Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., and Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. *48*, 2097–2098.

81. Misiewicz, Z., Iurato, S., Kulesskaya, N., Salminen, L., Rodrigues, L., Maccarrone, G., Martins, J., Czamara, D., Laine, M.A., Sokolowska, E., et al. (2019). Multi-omics analysis identifies mitochondrial pathways associated with anxiety-related behavior. PLoS Genet. *15*, e1008358.

82. Fleming, J.A., Kris-Etherton, P.M., Petersen, K.S., and Baer, D.J. (2021). Effect of varying quantities of lean beef as part of a Mediterranean-style dietary pattern on lipids and lipoproteins: a randomized crossover controlled feeding trial. Am. J. Clin. Nutr. *113*, 1126–1136.

83. Kazakova, O., Giniyatullina, G., Babkov, D., and Wimmer, Z. (2022). From marine metabolites to the drugs of the future: squalamine, trodusquemine, their steroid and triterpene analogues. Int. J. Mol. Sci. *23*, 1075.

84. Lei, K., Yuan, M., Zhou, T., Ye, Q., Zeng, B., Zhou, Q., Wei, A., and Guo, L. (2021). Research progress in the application of bile acid-drug conjugates: A "Trojan horse" strategy. Steroids *173*, 108879.

85. Wu, D., Ji, S., Wu, Y., Ju, Y., and Zhao, Y. (2007). Design, synthesis, and antitumor activity of bile acid-polyamine-nucleoside conjugates. Bioorg. Med. Chem. Lett. *17*, 2983–2986.

86. Tirosh, O., Voloshin, I., Sasson, S., Obercyger, M., and Madar, Z. (2018). Bile acid-basic amino acid conjugates and uses thereof. US Patent.

87. Vicens, M., Medarde, M., Macias, R.I.R., Larena, M.G., Villafaina, A., Serrano, M.A., and Marin, J.J.G. (2007). Novel cationic and neutral glycocholic acid and polyamine conjugates able to inhibit transporters involved in hepatic and intestinal bile acid uptake. Bioorg. Med. Chem. *15*, 2359–2367.

88. Mohanty, I., Allaband, C., Mannochio-Russo, H., El Abiead, Y., Hagey, L.R., Knight, R., and Dorrestein, P.C. (2024 in press). A perspective: the changing metabolic landscape of bile acids, keys to metabolism and immune regulation. Nat. Rev. Gastroenterol. Hepatol.

89. Guzior, D.V., and Quinn, R.A. (2021). Review: microbial transformations of human bile acids. Microbiome *9*, 140.

90. Doden, H.L., Wolf, P.G., Gaskins, H.R., Anantharaman, K., Alves, J.M.P., and Ridlon, J.M. (2021). Completion of the gut microbial epi-bile acid pathway. Gut Microbes *13*, 1–20.

91. Jones, B.V., Begley, M., Hill, C., Gahan, C.G.M., and Marchesi, J.R. (2008). Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc. Natl. Acad. Sci. USA *105*, 13580–13585.

92. Song, Z., Cai, Y., Lao, X., Wang, X., Lin, X., Cui, Y., Kalavagunta, P.K., Liao, J., Jin, L., Shang, J., et al. (2019). Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. Microbiome *7*, 9.

93. Daly, J.W., Keely, S.J., and Gahan, C.G.M. (2021). Functional and phylogenetic diversity of BSH and PVA enzymes. Microorganisms *9*, 732.

94. Fu, T., Huan, T., Rahman, G., Zhi, H., Xu, Z., Oh, T.G., Guo, J., Coulter, S., Tripathi, A., Martino, C., et al. (2023). Paired microbiome and metabolome analyses associate bile acid changes with colorectal cancer progression. Cell Rep. *42*, 112997.

95. Garcia, C.J., Kosek, V., Beltrán, D., Tomás-Barberán, F.A., and Hajslova, J. (2022). Production of new microbially conjugated bile acids by human gut microbiota. Biomolecules *12*, 687.

96. Dubrac, S., Elentner, A., Ebner, S., Horejs-Hoeck, J., and Schmuth, M. (2010). Modulation of T lymphocyte function by the pregnane X receptor. J. Immunol. *184*, 2949–2957.

97. Gutiérrez-Vázquez, C., and Quintana, F.J. (2018). Regulation of the immune response by the aryl hydrocarbon receptor. Immunity *48*, 19–33.

98. Chiang, J.Y.L., and Ferrell, J.M. (2018). Bile acid metabolism in liver pathobiology. Gene Expr. *18*, 71–87.

99. Monteiro-Cardoso, V.F., Corlianò, M., and Singaraja, R.R. (2021). Bile acids: A communication channel in the gut-brain axis. NeuroMolecular Med. *23*, 99–117.

100. Qiu, Y., Yu, J., Li, Y., Yang, F., Yu, H., Xue, M., Zhang, F., Jiang, X., Ji, X., and Bao, Z. (2021). Depletion of gut microbiota induces skeletal muscle atrophy by FXR-FGF15/19 signalling. Ann. Med. *53*, 508–522.

101. Aoi, W., Inoue, R., Mizushima, K., Honda, A., Björnholm, M., Takagi, T., and Naito, Y. (2023). Exercise-acclimated microbiota improves skeletal muscle metabolism via circulating bile acid deconjugation. iScience *26*, 106251.

102. Conway, J.R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics *33*, 2938–2940.

103. Wishart, D.S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., Dizon, R., Sayeeda, Z., Tian, S., Lee, B.L., et al. (2022). HMDB 5.0: the human metabolome database for 2022. Nucleic Acids Res. *50*, D622–D631.

104. Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A.C., Djoumbou, Y., Knox, C., Wilson, M., Liang, Y., Grant, J., et al. (2015). T3DB: the toxic exposome database. Nucleic Acids Res. *43*, D928–D934.

105. Rothwell, J.A., Perez-Jimenez, J., Neveu, V., Medina-Remón, A., M'hiri, N., García-Lobato, P., Manach, C., Knox, C., Eisner, R., Wishart, D.S., et al. (2013). Phenol-Explorer 3.0: a major update of the phenol-Explorer database to incorporate data on the effects of food processing on polyphenol content. Database *2013*, bat070.

106. Mohammed Taha, H., Aalizadeh, R., Alygizakis, N., Antignac, J.-P., Arp, H.P.H., Bade, R., Baker, N., Belova, L., Bijlsma, L., Bolton, E.E., et al. (2022). The Norman Suspect List Exchange (Norman-SLE): facilitating European and worldwide collaboration on suspect screening in high resolution mass spectrometry. Environ. Sci. Eur. *34*, 104.

107. Aron, A.T., Gentry, E.C., McPhail, K.L., Nothias, L.-F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J.M., Sikora, N., Vargas, F., et al. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. Nat. Protoc. *15*, 1954–1991.

108. Ludwig, M., Nothias, L.-F., Dührkop, K., Koester, I., Fleischauer, M., Hoffmann, M.A., Petras, D., Vargas, F., Morsy, M., Aluwihare, L., et al. (2020). Database-independent molecular formula annotation using Gibbs sampling through Zodiac. Nat. Mach. Intell. *2*, 629–641.

109. Dührkop, K., Shen, H., Meusel, M., Rousu, J., and Böcker, S. (2015). Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc. Natl. Acad. Sci. USA *112*, 12580–12585.

110. Plotly Technologies Inc. (2015). Collaborative Data Science (Plotly Technologies Inc.).

111. Bittremieux, W., Avalon, N.E., Thomas, S.P., Kakhkhorov, S.A., Aksenov, A.A., Gomes, P.W.P., Aceves, C.M., Rodríguez, A.M.C., Gauglitz, J.M., Gerwick, W.H., et al. (2017). Open access repository-scale propagated nearest neighbor suspect spectral library for untargeted metabolomics. Nat. Commun *14*, 8488.

112. Bittremieux, W., Chen, C., Dorrestein, P.C., Schymanski, E.L., Schulze, T., Neumann, S., Meier, R., Rogers, S., and Wang, M. (2020). Universal MS/MS visualization and retrieval with the metabolomics spectrum resolver web service. Preprint at bioRxiv. https://doi.org/10.1101/2020.05.09.086066.

113. Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis (Springer Science & Business Media).

114. Adams, K.J., Pratt, B., Bose, N., Dubois, L.G., St John-Williams, L., Perrott, K.M., Ky, K., Kapahi, P., Sharma, V., MacCoss, M.J., et al. (2020). Skyline for small molecules: A unifying software package for quantitative metabolomics. J. Proteome Res. *19*, 1447–1458.

115. Tian, Y., Cai, J., Allman, E.L., Smith, P.B., and Patterson, A.D. (2021). Quantitative analysis of bile acid with UHPLC-MS/MS. Methods Mol. Biol. *2194*, 291–300.

116. Koelmel, J.P., Kroeger, N.M., Gill, E.L., Ulmer, C.Z., Bowden, J.A., Patterson, R.E., Yost, R.A., and Garrett, T.J. (2017). Expanding lipidome coverage using LC-MS/MS data-dependent acquisition with automated exclusion list generation. J. Am. Soc. Mass Spectrom. *28*, 908–917.

117. ICH (2005). International Conference on Harmonization (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use. J. Pharmacol. Pharmacother. *6*, 185–187.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human fecal samples	Penny M Kris-Etherton, Pennsylvania State University	clinicaltrials.gov: NCT02723617
Feline fecal samples	Lee R. Hagey, University of California San Diego	N/A
Deposited data		
GNPS-BILE-ACID-MODIFICATIONS library	GNPS/MassIVE REPOSITORY	GNPS/MassIVE: MSV000091795; https://doi.org/10.5281/zenodo.7864031
Untargeted LC-MS/MS data of mice feces	Christoph W. Turck	GNPS/MassIVE: MSV000091663
Untargeted LC-MS/MS data of feline feces	Lee R. Hagey; Pieter C. Dorrestein	GNPS/MassIVE: MSV000093618
Untargeted LC-MS/MS data of human feces	Penny M Kris-Etherton	GNPS/MassIVE: MSV000093005 ftp://MSV000093005@massive.ucsd.edu
Untargeted LC-MS/MS data of animal feces	Itzhak Mizrahi	GNPS/MassIVE: MSV000086131; MSV000091663
Untargeted LC-MS/MS data of mice intestine	Michael Karin	GNPS/MassIVE: MSV000080918
NMR (raw data)	This paper	https://doi.org/10.5281/zenodo.10570399
Software and algorithms		
fastMASST	Robin Schmid, https://fasst.gnps2.org/fastsearch/	https://doi.org/10.5281/zenodo.7828220
microbeMASST	Zuffa et al.[67]	https://masst.gnps2.org/microbemasst/
MassQL	Jarmusch et al.[46]	https://massql.gnps2.org/

RESOURCE AVAILABILITY

Lead contact

Further queries and reagent requests may be directed and will be fulfilled by the lead contact, Pieter C. Dorrestein (pdorrestein@health.ucsd.edu).

Materials availability

This study did not generate new unique reagents. All the reagents in this study were included in the key resources table.

Data and code availability

- The MS/MS spectral library is available in GNPS (https://gnps.ucsd.edu/ProteoSAFe/gnpslibrary.jsp?library=GNPS-BILE-ACID-MODIFICATIONS; https://external.gnps2.org/gnpslibrary) and archived in Zenodo (https://doi.org/10.5281/zenodo.7864031). All untargeted metabolomics LC-MS/MS data is deposited to GNPS/MassIVE under the accession numbers MSV000091663, MSV000093618, MSV000093005, MSV000086131 and MSV000080918. Scripts to generate the figures in this study and NMR data can be accessed from GitHub (https://github.com/helenamrusso/Bile_acids_modifications). NMR data has also been archived at Zenodo (https://doi.org/10.5281/zenodo.10570399). A step-by-step guide to use our spectral library is provided in the GNPS documentation (https://wang-bioinformatics-lab.github.io/GNPS2_Documentation/libraries/)
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The Institutional Review Board at the Pennsylvania State University and MedStar Health Research Institute (for Beltsville Human Nutrition Research Center) approved the study protocol before the initiation of the study and all participants provided written informed consent. The trial is registered at clinicaltrials.gov (identifier: NCT02723617). All details about participants are available in the original published article on this study.[82]

METHOD DETAILS

Development of MassQL queries

The Mass Spec Query Language (MassQL) is an open-source mass spectrometry data mining tool that enables us to search, or query, mass spectrometry raw data for selective data patterns. In our case, we used MassQL for searching LC-MS/MS raw data in the public domain for bile-acid-selective patterns. Both "Stage 1" and "Stage 2" MassQL queries were designed to search for key fragment ions in the MS/MS spectra of non-, mono-, di-, tri-, tetra-, and penta-hydroxylated bile acids (Table S1). Stage 1 queries consisted of MS/MS fragment ions originating from the bile acid steroidal core only to ensure we could capture more modified bile acids, irrespective of the site of modification—which is either the steroidal hydroxyl or the side chain carboxyl group. The query was used to search for candidate bile acid MS/MS spectra in the high resolution and accurate mass Orbitrap data in GNPS/MassIVE. In our attempt to subset for only bile amidates (at the carboxyl), we designed Stage 2 queries by adding the detection of the modification (which can be any nucleophilic nitrogen-containing molecule like amino acids, amines, etc.) in the MS/MS fragmentation spectra as a requirement in the query. MassQL queries were run on the GNPS interface and the job links for each bile acid class are provided below:

Stage 1 MassQL query links

Non-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=cfda47ab389744f89972cd8ce2c6ace4
Mono-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=a6e6d4b22a2543ec94d381c2a2034ff8
Di-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=c62f29789f954b738c5d62f76cc35b0f
Tri-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=0a8d98af013b481786604e6021cab004
Tetra-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=04a19695e1a24fbfa7be956e9b142be4
Penta-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=7552066dfc904e7290cd30780eec5a02

Stage 2 MassQL query links

Non-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=dd5c3268383a4b9e8122acdec4a8152d
Mono-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7ed09cde325146e08acf6442e261c047
Di-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=912a4428469244a598da7f1920a96848
Tri-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e326a42131d04979b917b17ccf178be9
Tetra-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=1b81ef8acd024c69bdb7e5f9302579fd
Penta-hydroxy: https://proteomics2.ucsd.edu/ProteoSAFe/status.jsp?task=fc936525641341eab8a058f112fe2ae2

For testing the specificity of the Stage 1 MassQL queries we searched public spectral libraries using a precursor ion filter at 0.02 as we are searching for an entire repository, minimum of 6 matched ions and cosine > 0.7. The links of the library searches are provided below:

Stage 1 MassQL query library searches links

Non-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=88e7bc8b87434f0e9f7bda5006dd7319
Mono-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7cd8b114c87447428e8183e4cafa78e1
Di-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a6d3696b93774e1096bb092015af636f
Tri-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=7ac2036d971e4211882223c59726db96
Tetra-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=6bde3266cb0c4e1aad48e6ea9206a96e
Penta-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ee683fb8798a4d1395999e22b1986a7d

Calculation of delta mass and related analysis

Delta masses were calculated by subtracting the mass of the unconjugated bile acid steroidal core from the precursor *m/z* of the MS/MS spectra recovered from MassQL queries. After calculation of the delta masses from the MassQL query derived MS/MS spectra, the distribution of these delta masses in different bile acid classes was investigated. In order to determine how selective to the bile acids cores the delta masses were, UpSet plots were generated in R (https://www.R-project.org/, version 4.0.0) using the 'UpSetR' package (version 1.4.0).[102] Additionally, to represent the delta mass *m/z* distribution, a heatmap (Figure 1E) of the delta mass counts in the six bile acid classes was generated using the 'pheatmap' package (version 1.0.12). The delta mass binning on the x-axis of the heatmap was kept at 1 Da and delta mass counts were converted to a logarithmic scale.

Proposed annotation of delta masses at MS1 level

Delta masses were matched against the public molecular databases HMDB,[103] FoodDB (https://foodb.ca/), T3DB,[104] Phenol-Explorer,[105] and Norman.[106] Only molecules containing at least one hydrogen, as well as one nitrogen or one oxygen atom were considered. Matching was performed using the 'data.table' package (version 1.14.8). The mass tolerance was set to 5 ppm relative to the precursor *m/z* of the suspected conjugated bile acid.

MS/MS spectral clustering

MS/MS Spectra from the extracted results of the MassQL queries (which we accessed from the "Extract Results" tab on the status page of the completed MassQL job) were clustered using MS-Cluster[61] to reduce the number of spectra with redundant and nearly identical information. For this, we used the molecular networking workflow in the GNPS environment with precursor ion mass tolerance of 0.02 Da and MS/MS fragment ion mass tolerance of 0.05.[107] Minimum cosine score was set to 0.7 and minimum matched fragment peaks to 6. The job links are provided below:

Clustered spectra from Stage 1 query

Non-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=75a705dd5c3c4ea7ab7135920cad9dc2
Mono-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=88b52c909e54497f9fbfdcc4ccaa9cb3
Di-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=11424cb43b4c493ba1740912f6d121a2
Tri-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=85b060ec523f4955820ceed436b62885
Tetra-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b51812c1bc8e4d6bbf9ca20cbf9c13c7
Penta-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b8404e5531594ae4bf40a8dd37d5774a

Clustered spectra from Stage 2 query

Non-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=31f0317fa1144f148201cf13ee16ff29
Mono-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=63c1dc213d8647778b0c2d9e83262f2d
Di-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=4f4c6d59d5dc450799df9a47421ad767
Tri-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=40b2ead01c644b48bd2a2ef78f5b3ff5
Tetra-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=843a5e48c5e84e3a8684907e6542cd02
Penta-hydroxy: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ce294c37b77a46acbc34d6d695478776

The resulting clustered spectral files in MGF format were retrieved from the jobs above and used for downstream analyses. In addition, the most intense MS/MS spectrum from each of the clustered spectra were selected and used to create the GNPS-BILE-ACID-MODIFCATIONS library.

In silico annotations of candidate modifications

The molecular formulas of the clustered spectra were predicted using SIRIUS (version 5.6.2).[48] Precursor and fragment ion tolerance were set at 10 ppm, selecting Orbitrap profile, and considering only $[M+H]^+$ ions as the training models for SIRIUS was largely based on protonated MS/MS spectra and the MassQL searches we performed were designed based on $[M+H]^+$ spectra. The top 50 molecular formulas were calculated for each MS/MS spectrum, which were re-ranked using Zodiac[108] by setting the thresholdFilter at 0.98. Molecular fingerprints were generated via CSI:FingerID[109] (all available databases were selected). The molecular formulas obtained in this final step were considered for determining the molecular formulas of the delta masses observed for the bile acids.

The molecular formulas of the clustered spectra were also predicted using BUDDY (version 1.3).[49] In this case, the bottom-up MS/MS interrogation was selected, and the experiment-specific global peak annotation was not considered. MS1 and MS/MS tolerance were set to 10 and 20 ppm, respectively. Meta-scores were included for annotation. Chemical element counts were limited as N, 0-6; O, 1-10; P, 0-4; S, 0-2; F, 0-4; Cl, 0-4; Br, 0-4; I, 0-4. Adduct forms were defaulted to $[M+H]^+$. All the other settings were set as default.

From the predicted molecular formulas, the formulas of the delta masses were obtained using the chemparse package (https://pypi.org/project/chemparse/) in Python (version 3.7.6). The delta masses atomic difference (calculated from the SIRIUS predictions) which contained the groups of atoms C/H/N/P, C/H/O, C/H/N, C/H/N/O/S, C/H/N/O/P, C/H/S, C/H/O/S, and C/H/O/P were plotted as bar plots using the 'plotly' package (version 5.5.0).[110]

Considering that the atomic composition assignments are only proposed based on automated molecular formula prediction tools SIRIUS and BUDDY, they are not absolute identifications. When end-users are interested in using the spectral library resource for structure elucidation, they are encouraged to carefully inspect the MS/MS spectrum, verify the proposed atomic compositions as mass differences between modifications as small differences such as the mass of sulfates or phosphates can be difficult to calculate and therefore one needs to take extra care.[111]

Repository scale MS/MS spectral analysis with fastMASST

FastMASST, an evolution of the MASST concept, provides repository-scale spectral matches within seconds, matching an input spectrum (provided as an USI or list of ions in the MGF file) against the whole GNPS/MassIVE repository or the GNPS library of reference spectra.[62] Options to use this tool include a web dashboard (https://fasst.gnps2.org/fastsearch) and a REST web API. This tool is also leveraged by a novel tool that is in development that merges fastMASST spectral matches against a highly curated database

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

of > 60,000 LC-MS/MS samples obtained from monocultures of microorganisms with taxonomically-informed metadata (microbeMASST; https://masst.gnps2.org/microbemasst). Both fastMASST and microbeMASST are currently in the open test phase.

The clustered MGF spectra files were imported into the fastMASST pipeline and searched against both the GNPS/MassIVE repository and GNPS spectral libraries using the REST web API and the following parameters: 0.05 for both the precursor *m/z* tolerance and fragment *m/z* tolerance, 0.7 cosine threshold, 4 minimum matched signals, and analog search off. The wider *m/z* tolerances were chosen to also include datasets of instruments with lower resolving powers, present within public data. From these searches, fastMASST will provide information of similar MS/MS spectra that were observed in public dataset files. The matches were filtered to only keep the best match (highest cosine similarity) for each raw data file (sample). The microbeMASST pipeline produced tables of spectral matches to specific data files in the public domain which were then mapped to the curated list of taxonomically informed metadata. These tabular results are then visualized onto taxonomic trees according to NCBI taxonomy. Interactive HTML/javascript files summarizing the results also provide links to other GNPS/MassIVE tools and services (*e.g.*, matching GNPS library spectra, matching MassIVE datasets, interactive visualization and analysis of MS data files in the GNPS dashboard, and spectral mirror plots with the USI and the metabolomics spectrum resolver[112]). Outputs of microbeMASST indicate that the input spectrum was observed in a specific NCBI taxon for which MS data was previously acquired and deposited in GNPS/MassIVE

In order to have an overview of the distribution of the bile acid modifications (delta masses) within humans and rodents datasets, the results obtained from fastMASST[62] were merged with ReDU controlled vocabulary metadata. ReDU, which stands for "Reanalysis of Data User Interface," is an interface linking metadata with curated controlled vocabularies to public untargeted metabolomics data files.[64] This merged table was then filtered to contain only rows relative to rodent ("10088|Mus", "10090|Mus musculus", "10105|Mus minutoides", "10114|Rattus", "10116|Rattus norvegicus") or human-related datasets ("9606|Homo sapiens") in the NCBITaxonomy column. After filtering, the unique delta masses count for each bile acid core were mapped to the different body parts and tissues based on the UBERONBodyPartName column. UpSet plots were generated to get an estimation of delta masses that are shared among different body parts and tissues. To get a better visualization of the group intersections, the different skin body parts were combined into "skin", as well as blood plasma and blood serum were combined into "blood". In addition, intersections which occurred with a frequency of only 1 were removed from the visualization, but tables that summarize the body part distribution of each delta mass shown in these plots are also provided in Table S3. Code for generating this figure is available in the code availability section.

Repository scale MS/MS spectral analysis with microbeMASST

The clustered MGF files were imported into microbeMASST with the same search parameters mentioned previously for fastMASST. The resulting list of NCBI taxonomy IDs for bacterial cultures in microbeMASST were filtered for MS/MS spectral matches with modified bile acids obtained from MassQL queries. A taxonomic tree was generated using the Interactive Tree of Life (iTOL).[73] Association of the microbeMASST matched MS/MS spectra with bacterial class was realized by generating a heatmap of the corresponding delta masses in different bacterial classes. The heatmap was created using the 'pheatmap' package. The count of spectral matches was condensed at the phyla level as well and highlighted as barplots created in R using the 'ggplot2' package (version 3.4.2).[113] Delta masses from the mice study (MSV000080918) obtained from the MassQL queries with matches in microbeMASST were collected and the distribution with the type of diet and administration of antibiotic cocktail was visualized with the help of a heatmap for the sum normalized spectral counts in mice with high fat diet and normal chow diet and mice with and without antibiotic treatment. In addition, a similar method was adopted for generating the heatmap for amines from this dataset in Figure S3C. The spectral counts were obtained from the classical molecular networking job using the GNPS workflow (job link: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=a270b60e3d264097b5a603c8a93fc4f2).

Reanalysis of public data from GNPS/MassIVE

For reanalysis of the public dataset (MSV000086131), peak areas of polyamine bile amidates in this dataset were obtained by manual integration using the open-source quantitative software tool Skyline.[114] Boxplots representing the peak areas for carnivore, omnivore, and herbivore animals were generated using the 'ggpubr' package (version 0.5.0). The R script for this figure is provided in the code availability section.

Multiplexed reaction for polyamine bile amidates

All reactions were performed in flame- or oven-dried glassware sealed with rubber septa and under a nitrogen atmosphere unless otherwise indicated. Air- and/or moisture-sensitive liquids or solutions were transferred by cannula or syringe. Organic solutions were concentrated by rotary evaporator at 30 millibars with the water bath heated to not more than 50°C unless specified otherwise. Tetrahydrofuran (THF) was purified with a Pure-Solve MD-5 Solvent Purification System (Innovative Technology).

Solid cholic acid (0.08 mmol, 30 mg, 2.5 eq), solid chenodeoxycholic acid (0.08 mmol, 30 mg, 2.5 eq), and THF (3 mL) were added to a 20 mL scintillation vial with a magnetic stir bar. To the solution solid EDC (0.2 mmol, 30 mg, 5 eq), neat DIPEA (0.3 mmol, 0.05 mL, 10 eq), and solid DMAP (0.2 mmol, 20 mg, 5 eq) were added sequentially and the solution was stirred at 23 °C. After 15 min, agmatine, putrescine, spermidine, cadaverine, and 1,3-diaminopropane (1 eq each) were added simultaneously and the reaction was stirred for 14 h. The mixture was concentrated *in vacuo* and used directly for analysis.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

To obtain acetylated polyamine bile amides, solid cholic acid (0.08 mmol, 30 mg, 2.5 eq), solid chenodeoxycholic acid (0.08 mmol, 30 mg, 2.5 eq), and THF (3 mL) were added to a 20 mL scintillation vial with a magnetic stir bar. To the solution solid EDC (0.2 mmol, 30 mg, 5 eq), neat DIPEA (0.3 mmol, 0.05 mL, 10 eq), and solid DMAP (0.2 mmol, 20 mg, 5 eq) were added sequentially and the solution was stirred at 23 °C. After 15 minutes agmatine, putrescine, spermidine, cadaverine, and 1,3-diaminopropane (1 eq each) were added simultaneously and the reaction was stirred for 14 h. Neat acetic anhydride (0.3 mmol, 0.03 mL, 10 eq) was added and the reaction was stirred for 2 h. The mixture was concentrated *in vacuo* and used directly for analysis.

Co-migration analysis with multiplexed synthetic standard reaction mixtures

Samples from two studies were obtained for retention time and MS/MS spectral matching of polyamine bile amidates between the samples and the synthetic standards. Therefore, the samples, synthetic standards, and samples spiked with the synthetic standards were subjected to LC-MS/MS analyses. The LC-MS/MS analyses were carried out with a Vanquish UHPLC system coupled to a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The chromatographic separation was performed on a Polar C18 column (Kinetex C18, 100 x 2.1 mm, 2.6 μm particle size, 100A pore size – Phenomenex, Torrance, USA), and the mobile phase consisted of H_2O (solvent A), and ACN (solvent B), both acidified with 0.1% formic acid. Two different gradients were employed to evaluate retention time matching between synthetic standards and the compounds present in the samples: the first method (LC1) consisted of 0-0.5 min 5% B, 0.5-1.1 min 5-30% B, 1.1-5.0 min 30-60% B, 5.0-9.0 min 60-100% followed by a 1.5 min washout phase at 100% B, and a 1.5 min re-equilibration phase at 5% B. The second method (LC2) consisted of 0-0.5 min 5% B, 0.5-1.1 min 5-20% B, 1.1-5.0 min 20-40% B, 5.0-9.0 min 40-100% followed by a 1.5 min washout phase at 100% B, and a 1.5 min re-equilibration phase at 5% B. The flow rate was set at 0.5 mL/min, the injection volume was fixed at 3 μL, and the column temperature was set at 40 °C. Data-dependent acquisition (DDA) of MS/MS spectra was performed in the positive ionization mode. Electrospray ionization (ESI) parameters were set as: 52.5 AU sheath gas flow, 13.75 AU auxiliary gas flow, 2.7 AU spare gas flow, and 400 °C auxiliary gas temperature; the spray voltage was set to 3.5 kV and the inlet capillary to 320 °C and 50 V S-lens level was applied. MS scan range was set to 300-800 *m/z* with a resolution of 35,000 with one micro-scan. The maximum ion injection time was set to 100 ms with an automated gain control (AGC) target of 1.0E6. Up to 5 MS/MS spectra per MS1 survey scan were recorded in DDA mode with a resolution of 17,500 with one micro-scan. The maximum ion injection time for MS/MS scans was set to 150 ms with an AGC target of 5E5 ions. The MS/MS precursor isolation window was set to 1 *m/z*. The normalized collision energy was set to a stepwise increase from 30 to 40 to 60 with $z = 1$ as the default charge state. MS/MS scans were triggered at the apex of chromatographic peaks within 2 to 5 s from their first occurrence. The quality and reproducibility of the analyses were evaluated considering the retention time and the *m/z* of a standard solution containing a mixture of six standards, which was analyzed every 5 samples.

Extraction and LC-MS/MS profiling of human feces from diet cross over study

The Institutional Review Board at the Pennsylvania State University and MedStar Health Research Institute (for Beltsville Human Nutrition Research Center) approved the study protocol before the initiation of the study and all participants provided written informed consent. The trial is registered at clinicaltrials.gov (identifier: NCT02723617). The stool samples were prepared following the protocol established by Tian et al.[115] The methods for this study and preliminary results have been previously published.[82] A 25 mg stool sample was measured and placed in a 1.5 mL screw-cap homogenization tube, to which approximately 50 μL of 0.1 mm zirconia beads were added. Subsequently, 1 mL of ice-cold methanol containing 0.5 μM deuterated internal bile acid standards was added to the sample. The samples were thoroughly homogenized using a Precellys 24 tissue homogenizer, operating at 6500 rpm for two 20-second intervals. Following homogenization, the samples underwent a freeze-thaw cycle three times in liquid nitrogen and were centrifuged at 15,000 rpm at a temperature of 4 °C. Finally, a 200 μL of the supernatant was transferred to autosampler vials in preparation for LC-MS/MS analysis.

LC-MS/MS analyses were performed using a Thermo Scientific™ Orbitrap Exploris™ 120 Mass Spectrometer system coupled with a Thermo Scientific™ Vanquish™ UHPLC System (Thermo Scientific, Waltham, MA). The chromatographic separation was carried out using an ACQUITY BEH C18 UPLC column (2.1 x 100 mm, 1.7 μm; Waters, Milford, MA). The column was maintained at a temperature of 60 °C. The mobile phase solvent A was a mixture of 9 % acetonitrile and 1 mM ammonium acetate with a pH of 4.15, and the solvent B was a 1:1 ratio of acetonitrile and isopropanol. Samples (3 μL) were injected into the column and an initial mobile phase was 10 % B and kept for 0.1 min. Mobile phase B was increased in consecutive linear gradients from 10 % to 35 % over 9.15 min, 35 % to 85 % B over 2.25 min, and 85 % to 100 % B over 0.3 min. 100 % B was then held for 0.6 min followed by a linear gradient to 10 % B over 0.1 min and held for 2.5 min, for a total of 15 min. The flow rate was modified throughout the run starting with 300 μL/min for 9.25 min, then increasing to 325 μL/min over 2.25 min and 500 μL/min over 0.6 min. The flow rate was held at 500 μL/min for 0.3 min, then reduced to 300 μL/min over 0.4 min and held for 2.2 min. MS analyses for MS1 full scan were carried out using heated electrospray ionization in polarity switching mode using the following settings: positive and negative spray voltages of 3.5 kV and 2.5 kV respectively, sheath gas 50 Arb, auxiliary gas 10 Arb, sweep gas 1 Arb, ion transfer tube temperature 325 °C, vaporize temperature 350 °C and Orbitrap resolution of 120,000. The scan range (m/z) was from 100 to 1000. MS/MS spectra were separately acquired using IEomics method[116] with a QC pool sample.

Peak areas for polyamine bile amidates were extracted using Skyline.[114] Boxplots representing the peak areas for different diet treatments were generated using the 'ggplot2' package (version 3.4.2). The R script for this figure is provided in the code availability section.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

Synthesis of pure polyamine bile amidates

NMR spectra were collected at 298 K on a 600 MHz Bruker Avance III spectrometer fitted with a 1.7 mm triple resonance cryoprobe with z-axis gradients. (^{1}H NMR: MeOD (3.31) at 600 MHz; ^{13}C NMR: $CDCl_3$ (49.00) at 151 MHz). All spectra were taken in MeOD with shifts reported in parts per million (ppm) referenced to the proton or Carbon of the solvent (3.31 or 49.00, respectively). Coupling constants are reported in Hertz (Hz). Data for ^{1}H-NMR are reported as follows: chemical shift (ppm, reference to protium; s = single, d = doublet, t = triplet, q = quartet, dd = doublet of doublets, m = multiplet, coupling constant (Hz), and integration).

Cholyl-1,3-diaminopropane

Solid methyl cholate (284 μmol, 120 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in methanol (3 mL), then neat 1,3-diaminopropane (1.99 mmol, 0.17 mL, 7 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded cholyl-1,3-diaminopropane as a white solid. ^{1}H NMR (600 MHz, MeOD) δ 3.95 (s, 1H), 3.80 (d, J = 2.5 Hz, 1H), 3.41 – 3.35 (m, 1H), 3.24 (t, J = 6.7 Hz, 2H), 2.73 (t, J = 7.3 Hz, 2H), 2.33 – 2.22 (m, 3H), 2.11 (ddd, J = 13.8, 9.5, 6.7 Hz, 1H), 2.03 – 1.94 (m, 2H), 1.93 – 1.83 (m, 2H), 1.83 – 1.73 (m, 3H), 1.70 (m, 2H), 1.65 (m 1H), 1.63-1.50 (m, 5H), 1.47 – 1.33 (m, 4H), 1.29 (m, 1H), 1.18 – 1.07 (m, 1H), 1.03 (d, J = 6.4 Hz, 3H), 0.99 (td, J = 14.3, 3.2 Hz, 1H), 0.92 (s, 3H), 0.71 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 177.1, 74.0, 72.8, 67.0, 48.0, 47.5, 43.2, 43.0, 41.0, 40.4, 39.2, 37.4, 36.9, 36.5, 35.9, 34.1, 33.3, 31.9, 31.2, 29.6, 28.7, 27.8, 24.2, 23.2, 17.7, 13.0. HRESIMS m/z 465.3690 $[M+H]^+$ (calcd for C27H49N2O4, 465.3687, Δ ppm = 0.6).

Chenodeoxycholyl-1,3-diaminopropane

Solid chenodeoxycholic acid methyl ester (271 μmol, 110 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in methanol (3 mL), then neat 1,3-diaminopropane (1.89 mmol, 0.16 mL, 7 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded chenodeoxycholyl-1,3-diaminopropane. Subsequent purification by CombiFlash Nextgen300+ (5.5 g Redisep Gold C18 reverse phase column, evaporative light scanning detector, 13mL/min, gradient eluent system 0.1% FA H2O (A)/ 0.1% FA ACN (B)), eluted title compound at 25% B. ^{1}H NMR (600 MHz, MeOD) δ 8.32 (s, 2H), 3.80 (s, 1H), 3.38 (m,1H), 3.27 (t, J = 6.5 Hz, 2H), 2.93 (t, J = 7.2 Hz, 2H), 2.31 – 2.22 (m, 2H), 2.17 – 2.08 (m, 1H), 2.04 – 1.94 (m, 2H), 1.94 – 1.90 (m, 1H), 1.90 – 1.81 (m, 5H), 1.81 – 1.70 (m, 2H), 1.63 (dd, J = 26.5, 12.5 Hz, 2H), 1.56 – 1.47 (m, 4H), 1.44 (m, 1H), 1.40-1.26 (m, 6H), 1.25 – 1.15 (m, 2H), 1.15-1.07 (m, 1H), 1.05 – 1.00 (m, 1H), 0.98 (d, J = 6.5 Hz, 4H), 0.93 (s, 3H), 0.69 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 177.7, 72.8, 69.0, 57.3, 51.6, 43.7, 43.1, 41.1, 40.7, 40.5, 38.2, 37.0, 36.8, 36.5, 36.2, 35.9, 34.1, 33.9, 33.3, 31.3, 29.3, 28.8, 24.6, 23.4, 21.8, 18.9, 12.2. HRESIMS m/z 449.3741 $[M+H]^+$ (calcd for C27H49N2O3, 449.3738, Δ ppm = 0.7).

Cholyl-putrescine

Solid methyl cholate (284 μmol, 120 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in methanol (3 mL), then neat putrescine (1.42 mmol, 0.14 mL, 5 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded cholyl-putrescine. ^{1}H NMR (600 MHz, MeOD) δ 3.95 (s, 1H), 3.80 (d, J = 2.2 Hz, 1H), 3.41 – 3.36 (m, 1H), 3.25 – 3.15 (m, 2H), 2.95 (t, J = 7.4 Hz, 2H), 2.33 – 2.21 (m, 3H), 2.15 – 2.07 (m, 1H), 2.03 – 1.94 (m, 2H), 1.92-1.83 (m, 2H), 1.83 – 1.73 (m, 3H), 1.71-1.63 (m, 3H), 1.62 – 1.49 (m, 6H), 1.45 – 1.34 (m, 4H), 1.34 – 1.24 (m, 3H), 1.16 – 1.06 (m, 1H), 1.03 (d, J = 6.3 Hz, 3H), 0.99 (td, J = 14.1, 2.9 Hz, 1H), 0.92 (s, 3H), 0.71 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 177.0, 74.0, 72.8, 69.0, 47.9, 47.4, 43.0, 43.0, 40.9, 40.4, 40.1, 39.5, 36.9, 36.4, 35.8, 34.0, 33.2, 31.1, 29.5, 28.7, 27.8, 27.3, 25.8, 25.5, 24.2, 23.1, 17.8, 13.0. HRESIMS m/z 479.3844 $[M+H]^+$ (calcd for C28H51N2O4, 479.3843, Δ ppm = 0.2).

Chenodeoxycholyl-putrescine

Solid chenodeoxycholic acid methyl ester (246 μmol, 100 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid dissolved in methanol (3 mL), then neat putrescine (1.23 mmol, 0.12 mL, 5 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded chenodeoxycholyl-putrescine. ^{1}H NMR (600 MHz, MeOD) δ 3.79 (s, 1H), 3.43 – 3.36 (m, 1H), 3.20 (t, J = 6.4 Hz, 2H), 2.96 (t, J = 15.0, 7.5 Hz, 2H), 2.33-2.21 (m, 2H), 2.18 – 2.06 (m, 1H), 2.04 – 1.90 (m, 3H), 1.89-1.81 (m, 2H), 1.80 – 1.64 (m, 5H), 1.63-1.56 (m, 3H), 1.55 – 1.46 (m, 4H), 1.45-1.40 (m, 1H), 1.39 – 1.24 (m, 6H), 1.23-1.14 (m, 2H), 1.14 – 1.04 (m, 1H), 0.98 (d, J = 6.5 Hz, 4H), 0.92 (s, 3H), 0.68 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.8, 72.7, 68.9, 57.2, 51.5, 43.6, 43.1, 41.0, 40.7, 40.3, 39.5, 36.9, 36.5, 36.1, 35.9, 34.1, 34.0, 33.3, 31.3, 29.3, 27.3, 25.8, 24.6, 23.4, 21.8, 19.0, 12.2. HRESIMS m/z 463.3895 $[M+H]^+$ (calcd for C28H51N2O3, 463.3894, Δ ppm = 0.2).

Cholyl-N-acetyl-putrescine

Solid cholic acid (245 μmol, 100 mg) and solid EDC (367 μmol, 70.4 mg, 1.5 eq.) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in DMF (3 mL). Neat DIPEA (734 μmol, 128 μL, 3 eq.) was added and the reaction mixture stirred for 15 min at 23 °C. *N*-(4-aminobutyl)acetamide (367μmol, 47.8 mg, 1.5 eq) dissolved in DMF (0.5 mL) was added and stirred overnight. The reaction was concentrated in vacuo, triturated with diethyl ether, and purified by silica gel column chromatography. Elution with DCM:MeOH (7: 3) afforded cholyl-*N*-acetyl-putrescine. Subsequent purification by CombiFlash Nextgen300+ (5.5 g Redisep Gold C18 reverse phase column, evaporative light scanning detector, 13 mL/min, gradient eluent system 0.1% FA H2O (A)/ 0.1% FA ACN (B)), eluted title compound at 48% B. ^{1}H NMR (600 MHz, MeOD) δ 3.95 (s, 1H), 3.79 (d, J = 2.4 Hz, 1H), 3.40 – 3.36 (m, 1H), 3.23-3.15 (m, 4H), 2.33 – 2.22 (m, 3H), 2.16 – 2.08 (m, 1H), 2.04 – 1.98 (m, 1H), 1.97 – 1.93 (m, 4H), 1.92 – 1.83 (m, 2H), 1.83 -1.70 (m, 3H), 1.65 (d, J = 13.1 Hz, 1H), 1.62 – 1.55 (m, 3H), 1.54 – 1.48 (m, 6H), 1.47 – 1.32 (m, 4H), 1.27 (dd, J = 18.3,

9.3 Hz, 1H), 1.16 – 1.06 (m, 1H), 1.03 (d, J = 6.5 Hz, 3H), 0.98 (td, J = 14.2, 3.0 Hz, 1H), 0.91 (s, 3H), 0.71 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.9, 173.3, 74.0, 72.8, 69.0, 48.0, 47.4, 43.2, 43.0, 41.0, 40.4, 39.9, 39.8, 36.9, 36.5, 35.9, 34.2, 33.4, 31.2, 29.6, 28.7, 27.8, 27.7, 27.6, 24.2, 23.2, 22.7, 17.7, 13.0. HRESIMS m/z 521.3950 $[M+H]^+$ (calcd for C28H51N2O3, 521.3949, Δ ppm = 0.2).

Chenodeoxycholyl-N-acetyl-putrescine

Solid chenodeoxycholic acid (255 μmol, 100 mg) and solid EDC (382 μmol, 73.2 mg, 1.5 eq.) were added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in DMF (3 mL). Neat DIPEA (764 μmol, 133 μL, 3 eq.) was added and the reaction mixture was stirred for 15 min at 23 °C. *N*-(4-aminobutyl)acetamide (382 μmol, 49.7 mg, 1.5 eq) dissolved in DMF (0.5 mL) was added and stirred overnight. The reaction was concentrated in vacuo, triturated with diethyl ether, and purified by silica gel column chromatography. Elution with DCM:MeOH (7:3) afforded chenodeoxycholyl-*N*-acetyl-putrescine. Subsequent purification by CombiFlash Nextgen300+ (5.5 g Redisep Gold C18 reverse phase column, evaporative light scanning detector, 13 mL/min, gradient eluent system 0.1% FA H2O (A)/ 0.1% FA ACN (B)), eluted title compound at 50% B. ^{1}H NMR (600 MHz, MeOD) δ 3.79 (s, 1H), 3.40 – 3.34 (m, 1H), 3.20 – 3.13 (m, 4H), 2.30 – 2.20 (m, 2H), 2.13 – 2.05 (m, 1H), 2.04 – 1.94 (m, 2H), 1.92 (s, 3H), 1.91-1.85 (m, 2H), 1.84 – 1.58 (m, 5H), 1.56 – 1.47 (m, 7H), 1.46-1.41 (m, 1H), 1.40 – 1.24 (m, 5H), 1.23 – 1.14 (m, 2H), 1.13 – 1.00 (m, 2H), 0.98 (d, J = 5.8 Hz, 4H), 0.93 (s, 3H), 0.69 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.6, 173.1, 69.2, 69.1, 57.3, 51.5, 43.7, 43.1, 41.0, 40.7, 40.4, 40.1, 40.0, 36.9, 36.5, 36.2, 35.9, 34.1, 34.0, 33.4, 31.3, 29.3, 27.8, 27.7, 24.6, 23.4, 22.6, 21.8, 18.9, 12.2. HRESIMS m/z 505.4003 $[M+H]^+$ (calcd for C30H53N2O4, 505.4000, Δ ppm = 0.6).

Cholyl-cadaverine

Solid methyl cholate (237 μmol, 100 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar, and the solid was dissolved in methanol (3 mL), then neat cadaverine (1.18 mmol, 0.14 mL, 5 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded cholyl-cadaverine. ^{1}H NMR (600 MHz, MeOD) δ 3.95 (s, 1H), 3.80 (d, J = 2.3 Hz, 1H), 3.41 – 3.35 (m, 1H), 3.21 – 3.11 (m, 2H), 2.64 (t, J = 7.2 Hz, 2H), 2.34 – 2.20 (m, 3H), 2.10 (ddd, J = 13.7, 9.3, 7.1 Hz, 1H), 2.04 – 1.93 (m, 2H), 1.93 – 1.84 (m, 2H), 1.83 – 1.71 (m, 3H), 1.65 (d, J = 13.0 Hz, 1H), 1.62-1.53 (m, 4H), 1.53-1.45 (m, 5H), 1.45 – 1.25 (m, 7H), 1.14 – 1.07 (m, 1H), 1.03 (d, J = 6.5 Hz, 3H), 0.98 (td, J = 14.2, 3.0 Hz, 1H), 0.92 (s, 3H), 0.71 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.8, 74.0, 72.9, 69.0, 48.0, 47.5, 43.2, 43.0, 42.3, 41.0, 40.5, 40.2, 36.9, 36.5, 35.9, 35.9, 34.2, 33.4, 33.1, 31.2, 30.3, 29.6, 28.7, 27.9, 25.3, 24.2, 23.2, 17.7, 13.0. HRESIMS m/z 493.4003 $[M+H]^+$ (calcd for C29H53N2O4, 493.4000, Δ ppm = –X.X).

Chenodeoxycholyl-cadaverine

Solid chenodeoxycholic acid methyl ester (271 μmol, 110 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar, and the solid was dissolved in methanol (3 mL), then neat cadaverine (1.35 mmol, 0.16 mL, 5 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded chenodeoxycholyl-cadaverine. ^{1}H NMR (600 MHz, MeOD) δ 3.80 (s, 1H), 3.41 – 3.34 (m, 1H), 3.18 (t, J = 7.0 Hz, 2H), 2.93 (t, J = 7.6 Hz, 2H), 2.30 – 2.20 (m, 2H), 2.16 – 2.06 (m, 1H), 2.04 – 1.90 (m, 3H), 1.90 – 1.80 (m, 2H), 1.80 – 1.58 (m, 6H), 1.58 – 1.46 (m, 6H), 1.42 (m, 3H), 1.34 (m, 6H), 1.23 – 1.14 (m, 2H), 1.13 – 1.04 (m, 1H), 0.97 (d, J = 6.6 Hz, 4H), 0.92 (s, 3H), 0.68 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.9, 72.8, 69.0, 57.2, 51.5, 43.6, 43.0, 41.0, 40.7, 40.6, 40.3, 39.9, 36.9, 36.5, 36.1, 35.8, 34.1, 34.0, 33.3, 31.2, 29.8, 29.2, 28.0, 24.7, 24.6, 23.4, 21.7, 18.9, 12.1. HRESIMS m/z 477.4055 $[M+H]^+$ (calcd for C29H53N2O3, 477.4051, Δ ppm = 0.8).

Cholyl-N-acetyl-cadaverine

Cholyl-cadaverine (45 μmol, 22 mg) was added to a 1-dram scintillation vial equipped with a magnetic stir bar and the solid was dissolved in anhydrous DMF (1 mL). Neat triethylamine (54 μmol, 7.5 μL, 1.2 eq.) and neat acetic anhydride (54 μmol, 5.1 μL, 1.2 eq.) were added and the reaction was stirred at 23 °C for 2 h. The reaction was concentrated in vacuo, triturated with diethyl ether, and purified by silica gel column chromatography. Elution with DCM:MeOH (7:3) afforded cholyl-N-acetyl-cadaverine. Subsequent purification by CombiFlash Nextgen300+ (5.5 g Redisep Gold C18 reverse phase column, evaporative light scanning detector, 13 mL/min, gradient eluent system 0.1% FA H2O (A)/ 0.1% FA ACN (B)), eluted title compound at 45% B. ^{1}H NMR (600 MHz, MeOD) δ 3.95 (s, 1H), 3.80 (d, J = 2.2 Hz, 1H), 3.37 (m, 1H), 3.20 – 3.11 (m, 4H), 2.86 – 2.63 (m, 3H), 2.34 – 2.20 (m, 3H), 2.17 – 2.06 (m, 1H), 2.04 – 1.94 (m, 2H), 1.93 (s, 3H), 1.90 – 1.71 (m, 5H), 1.69 – 1.56 (m, 5H), 1.55-1.48 (m, 5H), 1.45 – 1.26 (m, 6H), 1.17 – 1.07 (m, 1H), 1.02 (d, J = 6.9 Hz, 3H), 0.98 (td, J = 13.8, 2.7 Hz, 1H), 0.92 (s, 3H), 0.70 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.8, 173.2, 74.0, 72.9, 69.0, 48.0, 47.5, 43.2, 43.0, 41.0, 40.5, 40.4, 40.2, 36.9, 36.5, 35.9, 35.9, 34.2, 33.4, 31.2, 30.1, 30.0, 29.6, 28.7, 27.9, 25.3, 24.2, 23.2, 22.5, 17.7, 13.0. HRESIMS m/z 535.4110 $[M+H]^+$ (calcd for C31H55N2O5, 535.4105, Δ ppm = 0.9).

Chenodeoxycholyl-N-acetyl-cadaverine

Chenodeoxycholyl-cadaverine (41.4 μmol, 19.7 mg) was added to a 1-dram scintillation vial equipped with a magnetic stir bar and the solid was dissolved in anhydrous DMF (1 mL). Neat triethylamine (59.6 μmol, 6.9 μL, 1.2 eq.) and neat acetic anhydride (59.6 μmol, 4.68 μL, 1.2 eq.) were added and the reaction was stirred at 23 °C for 2 h. The reaction was concentrated in vacuo, triturated with diethyl ether, and purified by silica gel column chromatography. Elution with DCM:MeOH (7: 3) afforded chenodeoxycholyl-N-acetyl-cadaverine. ^{1}H NMR (600 MHz, MeOD) δ 3.79 (d, J = 1.5 Hz, 1H), 3.41 – 3.33 (m, 1H), 3.20 – 3.12 (m, 4H), 2.31 – 2.20 (m, 2H), 2.09 (ddd, J = 13.8, 9.5, 6.9 Hz, 1H), 2.04 – 1.93 (m, 2H), 1.93 (s, 3H), 1.92 – 1.81 (m, 3H), 1.81 – 1.70 (m, 2H), 1.68 – 1.58 (m, 2H), 1.56 – 1.47 (m, 7H), 1.46 – 1.40 (m, 1H), 1.40 – 1.25 (m, 8H), 1.24 – 1.15 (m, 2H), 1.10 (m,z 1H), 0.99 (dd, J = 19.9, 4.8 Hz, 4H), 0.93 (s, 3H), 0.69 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.8, 173.2, 72.8, 69.0, 57.3, 51.5, 43.7, 43.1, 41.0, 40.7, 40.4, 40.4,

40.2, 36.9, 36.5, 36.2, 35.9, 34.1, 34.0, 33.4, 31.3, 30.1, 30.0, 29.3, 25.2, 24.6, 23.4, 22.6, 21.8, 18.9, 12.2. HRESIMS m/z 519.4161 $[M+H]^+$ (calcd for C31H55N2O4, 519.4156, Δ ppm = 1.0).

*Cholyl-spermidine *mixture of isomers*

Solid methyl cholate (284 μmol, 120 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in methanol (3 mL), then neat spermidine (568 μmol, 82.5 mg, 2 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded cholyl-spermidine. Subsequent purification by CombiFlash Nextgen300+ (5.5 g Redisep Gold C18 reverse phase column, evaporative light scanning detector, 13 mL/min, gradient eluent system 0.1% FA H2O (A)/ 0.1% FA ACN (B)), eluted title compound at 18% B. ^{1}H NMR (600 MHz, MeOD) δ 3.95 (s, 1H), 3.80 (d, J = 2.3 Hz, 1H), 3.40 – 3.34 (m, 1H), 3.26 (t, J = 6.6 Hz, 1H), 3.19 (t, J = 7.1 Hz, 1H), 2.94 – 2.89 (m, 2H), 2.86 – 2.76 (m, 4H), 2.32 – 2.22 (m, 3H), 2.16-2.07 (m, 1H), 2.04 – 1.79 (m, 9H), 1.78-1.70 (m, 5H), 1.67 – 1.51 (m, 9H), 1.47 – 1.24 (m, 6H), 1.11 (qd, J = 11.8, 5.4 Hz, 1H), 1.05 – 1.02 (m, 3H), 0.98 (td, J = 14.2, 3.1 Hz, 1H), 0.92 (s, 3H), 0.71 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 176.0, 72.6, 72.6, 67.7, 46.6, 46.1, 45.6, 41.8, 41.6, 39.6, 39.1, 39.1, 35.6, 35.1, 34.5, 34.5, 32.7, 29.8, 28.2, 27.4, 27.3, 26.5, 24.5, 24.5, 22.8, 21.8, 16.3, 11.6. HRESIMS m/z 520.4478 $[M+H]^+$ (calcd for C31H58N3O3, 520.4473, Δ ppm = 1.0).

*Chenodeoxycholyl-spermidine *mixture of isomers with the same mass*

Solid chenodeoxycholic acid methyl ester (246 μmol, 100 mg) was added to a 20 mL scintillation vial equipped with a magnetic stir bar and the solid was dissolved in methanol (3 mL), then neat spermidine (492 μmol, 71.4 mg, 2 eq.) was added. The reaction was stirred at 65 °C for 72 h. The reaction mixture was concentrated in vacuo and purified by silica gel column chromatography. Elution with MeOH/NH_3 (100:1) afforded chenodeoxycholyl-spermidine. ^{1}H NMR (600 MHz, MeOD) δ 3.80 (s, 1H), 3.41-3.34 (m, 1H), 3.26 (t, J = 6.7 Hz, 1H), 3.19 (t, J = 6.8 Hz, 1H), 2.94-2.86 (m, 2H), 2.86-2.79 (m, 3H), 2.78 – 2.75 (m, 1H), 2.31 – 2.21 (m, 2H), 2.18 – 2.05 (m, 1H), 2.05 – 1.90 (m, 3H), 1.90 – 1.81 (m, 4H), 1.80 – 1.69 (m, 4H), 1.68-1.58 (m, 4H), 1.57-1.47 (m, 5H), 1.47-1.40 (m, 1H), 1.40 – 1.25 (m, 6H), 1.24 – 1.15 (m, 2H), 1.14 – 1.05 (m, 1H), 0.98 (d, J=6.5 Hz, 4H), 0.93 (s, 3H), 0.69 (s, 3H). ^{13}C NMR (151 MHz, MeOD) δ 175.6, 71.5, 67.7, 55.8, 50.2, 42.3, 41.7, 39.7, 39.4, 39.1, 39.0, 38.4, 38.3, 35.6, 35.2, 34.9, 34.5, 32.8, 32.8, 32.7, 32.0, 30.0, 28.0, 26.0, 24.5, 23.4, 23.3, 22.2, 20.5, 17.8, 11.0. HRESIMS m/z 536.4427 $[M+H]^+$ (calcd for C31H58N3O3, 536.4422, Δ ppm = 0.9).

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of polyamines in feline feces

Fecal samples from the Project Survival Cat Haven in California included materials collected from an African lion, an African serval, a cheetah, two jaguars, three jaguarundis, and a tiger. The extracts were prepared in a proportion of 25 mg of fecal material to 1 mL of MeOH:H_2O (1:1), and the samples were homogenized in a Qiagen TissueLyzer II for 5 min at 25 MHz. The samples were centrifuged (14,000 *g*) for 15 min, 400 μL of the supernatants was transferred and evaporated in a Labconco CentriVap, and further stored in –80 °C until analysis. The dried extracts were resuspended in MeOH:H_2O (1:1) to achieve 1 mg/mL. The LC-MS/MS method used for the analyses was the same as previously described in the co-migration analysis section, employing the gradient LC1. The analytical method was performed according to the ICH guidelines[117] based on cholyl-taurine, cholyl-putrescine, cholyl-*N*-acetyl-putrescine, cholyl-cadaverine, and cholyl-glycine. The following parameters were evaluated: specificity, precision (repeatability and intermediate precision), linearity, limit of detection, limit of quantification, and accuracy. More detailed information regarding the methodology used in the validation and quantification is provided below, and the figures of merit are shown in Table S5. Method used to synthesize pure standards synthesized for quantification can be found above in STAR Methods.

LC-MS/MS method validation for polyamines bile amidates quantification

The method validation was performed based on the International Conference on Harmonization (ICH) guidelines.[117] Specificity, linearity, precision, accuracy, and limits of detection and quantification (LOD and LOQ) were determined. For the validation, the African serval fecal sample (AS-1) was selected for containing a significant amount of all the five bile acids: cholyl-glycine, cholyl-putrescine, cholyl-cadaverine, cholyl-taurine, and cholyl-*N*-acetyl-putrescine. To account for carry-over effects observed for the polyamine conjugates, the area relative to the blanks was subtracted from the area of the compounds in the solutions and animal samples, considering only the peaks that were at least in a 2-fold difference.[8] The method employed reached the acceptance criteria specified for each parameter, as detailed in Table S5.

Specificity

The specificity of a method relies on its ability to detect and differentiate the signals of the analytes from the other compounds in the natural matrix or impurities. In this study, the specificity was determined based on the injection of a blank solution (containing sulfadimethoxine as the internal standard, and a solution of the standards cholyl-glycine, cholyl-putrescine, cholyl-cadaverine, cholyl-taurine, and cholyl-*N*-acetyl-putrescine (n = 3). The relative standard deviation (RSD) was calculated based on each peak retention time in the sample AS-1 (African serval). The MS and MS/MS mass spectra of each compound were also used to confirm the specificity and identity of these compounds.

Please cite this article in press as: Mohanty et al., The underappreciated diversity of bile acid modifications, Cell (2024), https://doi.org/10.1016/j.cell.2024.02.019

The peaks of interest presented the following retention times: cholyl-glycine (5.27 min), cholyl-putrescine (4.20 min), cholyl-cadaverine (4.34 min), cholyl-taurine (4.73 min), and cholyl-*N*-acetyl-putrescine (5.43 min). These compounds did not show any interferences when compared to the solution containing only the standards.

Precision (repeatability and intermediate precision)

The precision of the method was determined by analyzing six solutions of AS-1, each one injected one time (n=6), while the repeatability (intraday precision) was estimated as the relative standard deviation (RSD) of the standards concentrations (μg mL–1) measured in two consecutive days. On the first day, the concentrations calculated for the standards were 0.033 (±0.001), 0.036 (±0.001), 0.034 (±0.001), 0.299 (±0.005), and 0.025 (±0.001) for cholyl-glycine, cholyl-putrescine, cholyl-cadaverine, cholyl-taurine, and cholyl-*N*-acetyl-putrescine, respectively. On the second day, the concentration values were 0.032 (±0.002), 0.036 (±0.001), 0.031 (±0.001), 0.305 (±0.012), 0.025 (±0.001), respectively. The RSD values were lower than 5%, and the F-test between the two days revealed that there was no significant difference at F=0.05.

Linearity

The linearity of the method was determined by calibration curves in concentration ranges comprising each compound at the samples of interest. Initially, a stock solution containing cholyl-glycine (1.0 μg mL–1), cholyl-putrescine (4.0 μg mL–1), cholyl-cadaverine (1.5 μg mL–1), cholyl-taurine (50 μg mL–1), and cholyl-*N*-acetyl-putrescine (0.5 μg mL–1) in MeOH–H_2O (1:1) was prepared to acquire analytical curves simultaneously. From this solution, six to seven concentration levels were prepared in MeOH–H_2O (1:1) to obtain calibration curves. The levels ranged from 0.05 to 0.5 μg mL–1, 0.004 to 0.4 μg mL–1, 0.0075 to 0.225 μg mL–1, 0.25 to 15.00 μg mL–1, and 0.0025 to 0.075 μg mL–1 for cholyl-glycine, cholyl-putrescine, cholyl-cadaverine, cholyl-taurine, and cholyl-*N*-acetyl-putrescine, respectively. Each concentration level was injected in triplicate. The linearity was determined by the analytical curves built based on the nominal concentrations and the average of the ratios between the areas of each compound and the internal standard (R=Acompound/AIS). The correlation coefficients (R) were calculated for each compound. All the linear concentration ranges were adequate for all the compounds, with R coefficients above 0.9990.

Limit of detection and limit of quantification

The limits of detection (LODs) and limits of quantification (LOQs) were estimated by the mean of the slopes (a) and the standard deviation of the y-intercept (Sb) on three calibration curves in three low concentrations for each compound (10 to 100 ng mL–1 for cholyl-glycine; 0.4 to 4.0 ng mL–1 for cholyl-putrescine; 0.8 to 7.5 ng mL–1 for cholyl-cadaverine; 25 to 250 ng mL–1 for cholyl-taurine; 0.3 to 2.5 ng mL–1 for cholyl-*N*-acetyl-putrescine). These limits were calculated by the following equations: LOD = 3.3Sb.a–1 and LOQ = 10Sb.a–1.

Accuracy

The accuracy of the method was determined by recovery studies, in which known amounts of the isolated standards were spiked to the AS-1 solution in three different concentrations (low, intermediate, and high) considering the predetermined linearity concentration range. Three replicates for each level were prepared and the samples were analyzed using the validated method. The accuracy was determined by the difference between the nominal and the experimental concentration values. The recoveries obtained were within the acceptance range of 80%–120%.

Statistical analyses

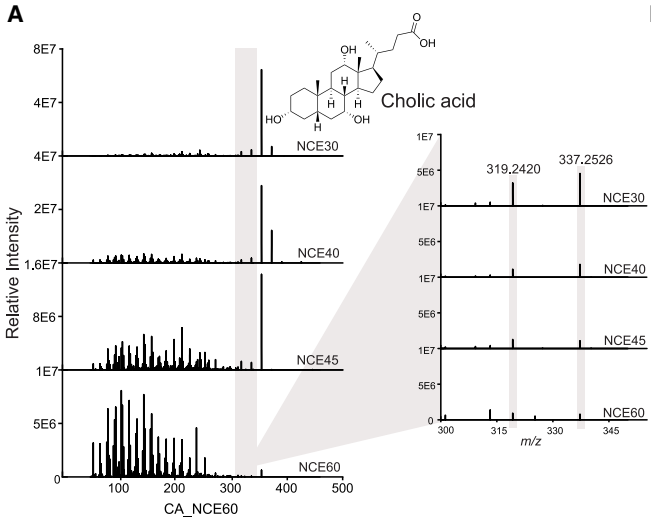
Statistical analyses for comparing different groups in Figure 4B was performed using Kruskal Wallis test followed by a pairwise Wilcoxon test was performed in R using the ‘ggpubr‘ package to calculate p values of significance which were then corrected for multiple comparisons using Benjamini-Hochberg correction. Pairwise Wilcoxon test was performed in R using the ‘rstatix‘ package (version 0.7.2) to calculate p values of significance which were then corrected for multiple comparisons using Bonferroni correction in Figures 6C and 6D.

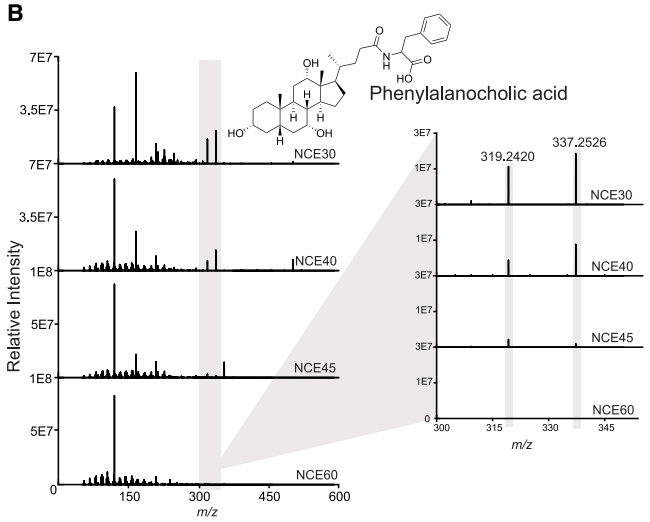
ADDITIONAL RESOURCES

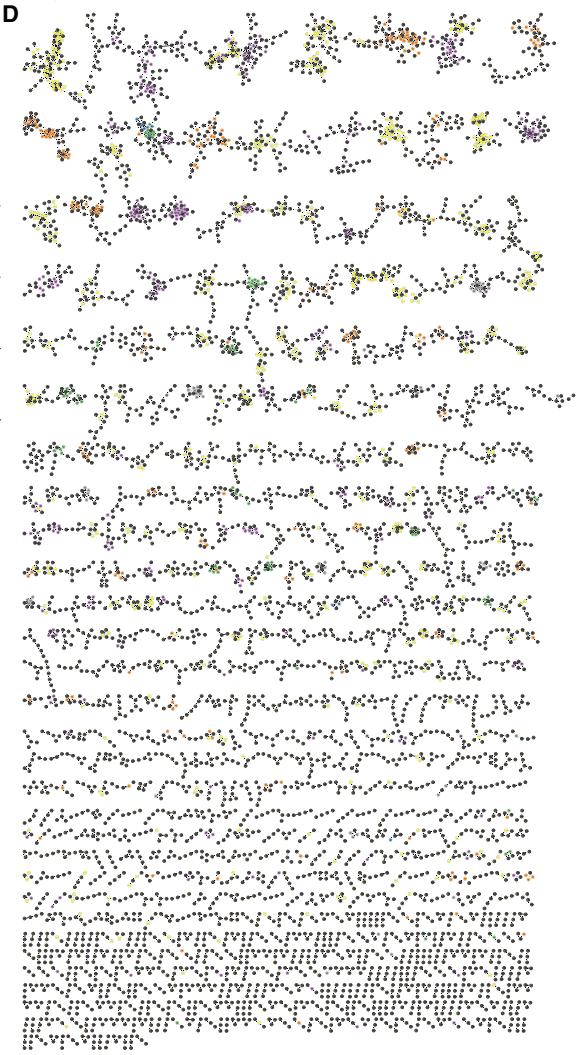
The human Mediterranean diet crossover study from which we analyzed fecal samples for LC-MS/MS analysis has been registered at clinicaltrials.gov as NCT02723617.

CellPress
OPEN ACCESS

Supplemental figures



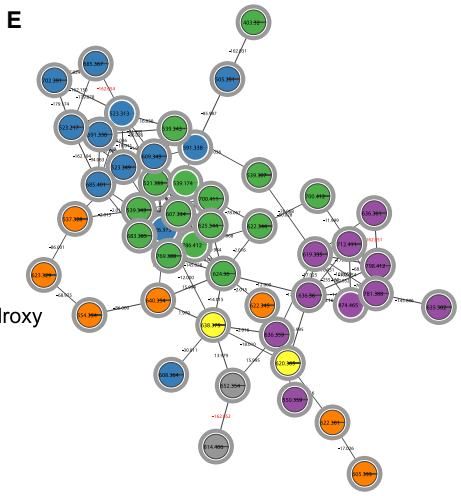




C

QUERY (STAGE 1)
QUERY scaninfo(MS2DATA) WHERE
MS2PROD=337.25:TOLERANCEMZ=0.01:INTENSITYPERCENT=5 AND
MS2PROD=319.24:TOLERANCEMZ=0.01:INTENSITYPERCENT=5

QUERY TRANSLATION
Returning the scan information on MS2
The following conditions are applied to find scans in the mass spec data
Finding MS2 peak at m/z 337.25 with a 0.01 m/z tolerance and a minimum percent intensity relative to base peak of 5.0%
Finding MS2 peak at m/z 319.24 with a 0.01 m/z tolerance and a minimum percent intensity relative to base peak of 5.0%



Nonhydroxy Monohydroxy Dihydroxy Trihydroxy
Tetrahydroxy Pentahydroxy

(legend on next page)

Figure S1. Conserved MS/MS fragment ions and molecular networking of the MassQL query results, related to Figure 1

(A and B) MS/MS fragmentation spectra for (A) cholic acid and (B) phenylalanocholic acid are shown with the zoomed-out region for the two diagnostic MS/MS fragment ions at m/z 319.2429 and 337.2526 for tri-hydroxylated bile acid class. The data were acquired on an Orbitrap with different normalized collision energies (NCEs).

(C) The stage 1 MassQL query for tri-hydroxylated bile acid is shown along with its translation in English as generated on MassQL Sandbox (link for visualization).

(D) Each network represents structurally similar compounds with similar MS/MS fragmentation patterns. The nodes represent each ion feature, while the edges represent their spectral similarity. Only networks with three or more nodes are shown. The Cytoscape file for interactive inspection is provided on the GitHub page (https://github.com/helenamrusso/Bile_acids_modifications).

(E) Highlighted in red are the neutral losses that can be explained by potential glycosylation in the structure. The Cytoscape file for interactive inspection is provided on the GitHub page (https://github.com/helenamrusso/Bile_acids_modifications).

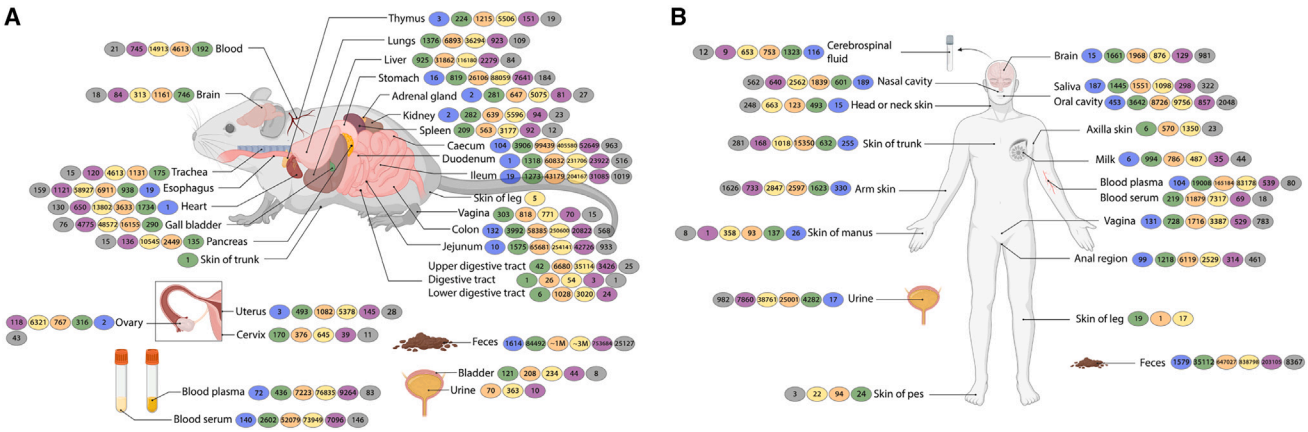


Figure S2. Distributions of all MS/MS spectra of candidate bile acids from fastMASST searches, related to Figure 2

Based on metadata availability in ReDU, the distribution of MS/MS spectra across different body parts is shown for (A) rodents and (B) humans. The numbers in each oval indicate the number of MS/MS spectra observed in each of the sample types. Panels were created using BioRender.com.

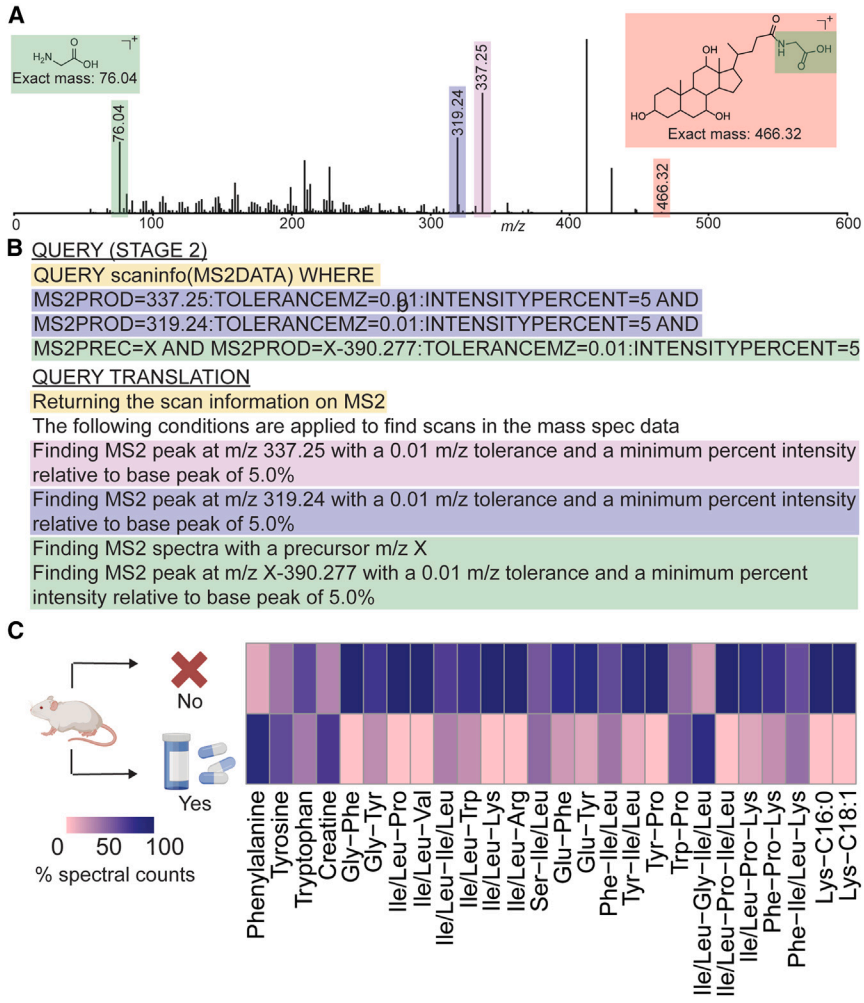


Figure S3. Stage 2 MassQL query and the detection of amines, related to Figure 3

(A) Representative MS/MS spectrum of glycocholic acid (tri-hydroxylated bile acid amidate) with the characteristic fragmentation patterns indicated upon which the stage 2 MassQL query is designed. The loss of glycine is highlighted in green, which is the additional data filtering condition added to the stage 1 queries.

(B) Stage 2 MassQL query and its English language translation for the tri-hydroxylated bile acid.

(C) Relative abundances of amine corresponding delta masses obtained from classical molecular networking concerning antibiotic use in the public dataset MSV000080918 are shown in a heatmap. Of the amines shown here, only Phe, Tyr, and Trp bile amidates were observed, and only in the no antibiotic-treated samples. These three amino acids were observed in the non-antibiotic-treated samples, but not conjugated to bile acids, where only taurine conjugates were observed. This observation disfavors the conjugation of amines to bile acids as a non-enzymatic addition.

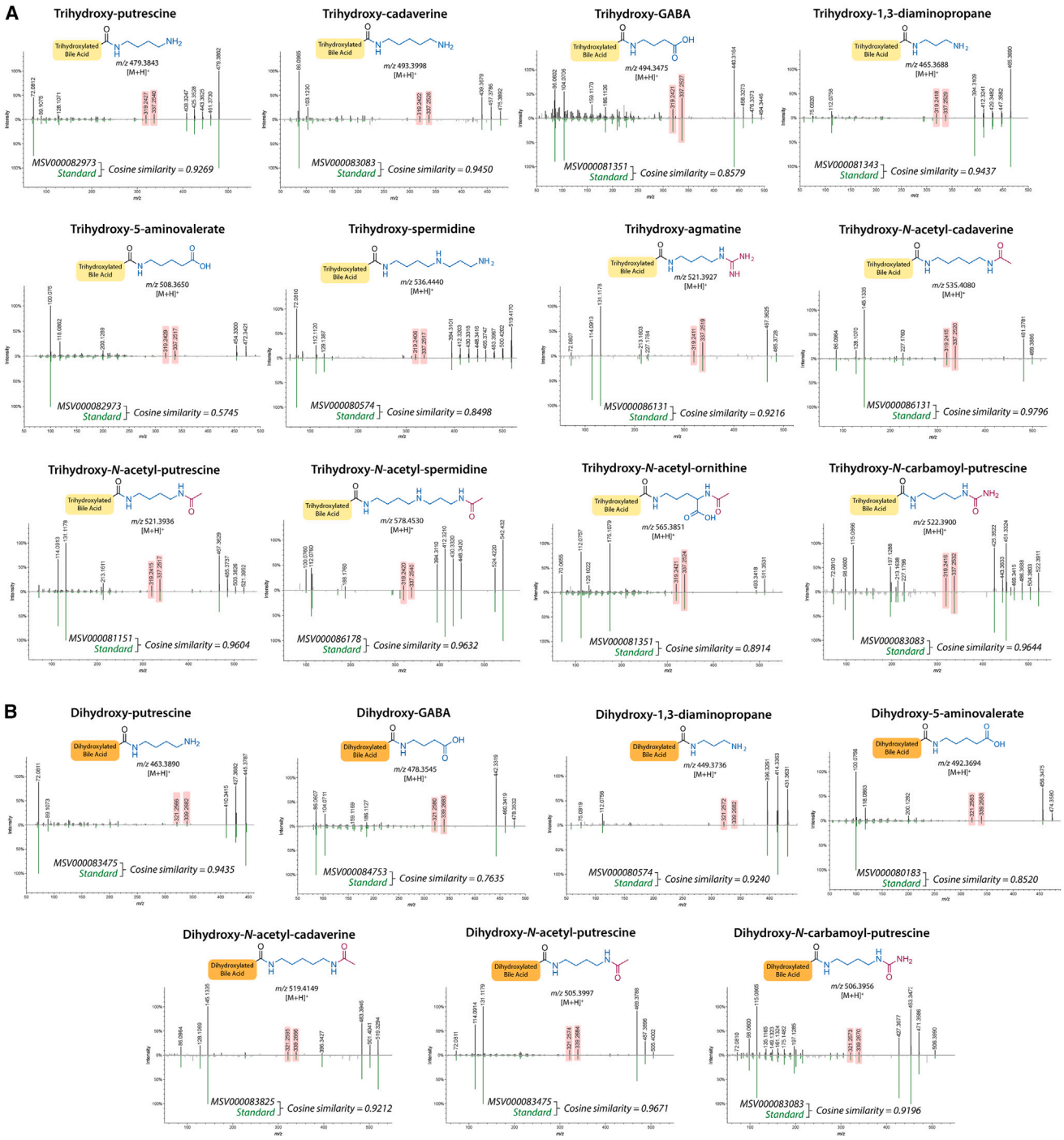
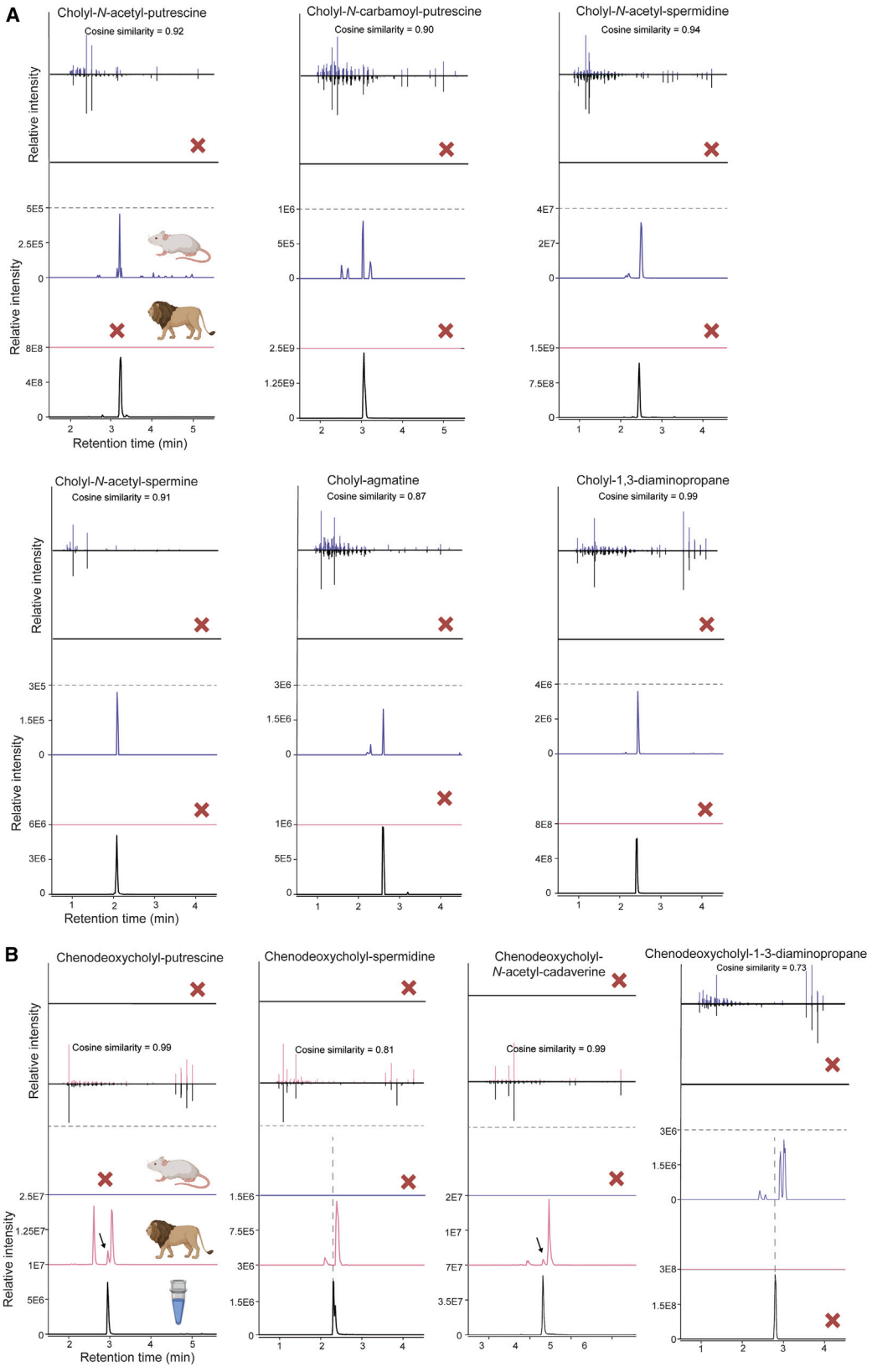


Figure S4. Polyamine bile amidates from public data (MS level 2 annotation), related to Figure 4

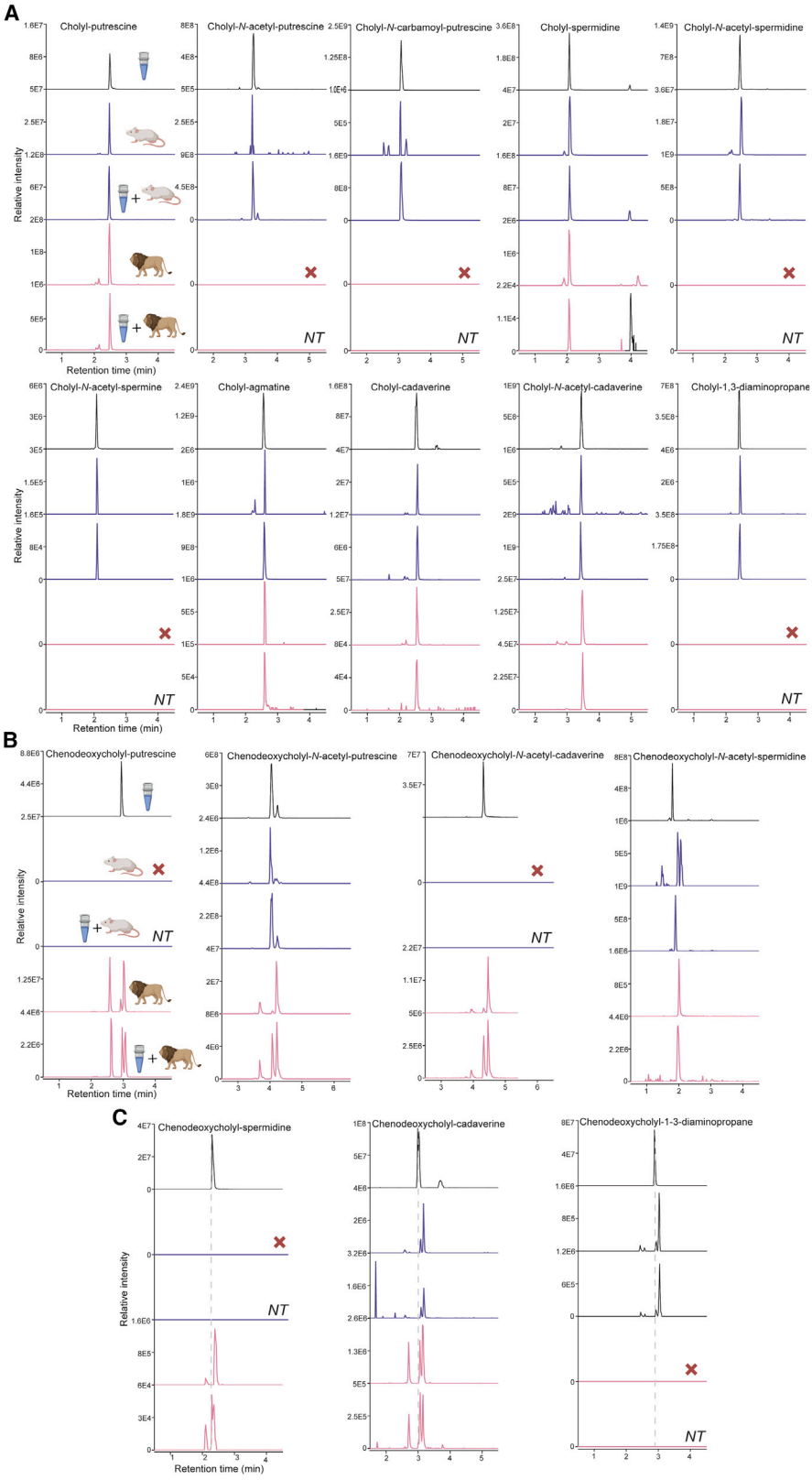
MS/MS spectral matches for candidate polyamine bile amidates with tri-hydroxylated and di-hydroxylated bile acids from public data (MSV000080183, MSV000080574, MSV000081151, MSV000081343, MSV000081351, MSV000082973, MSV000083083, MSV000083475, MSV000083825, MSV000084753, MSV000086131, MSV000086178) (https://massive.ucsd.edu/) with synthetic standards. The dataset related to each MS/MS spectra in the top (black) panels is specified in each mirror plot. MS/MS mirror plots can be visualized directly from the raw data in the GNPS dashboard with the following links for the plots in the order they appear in the figure: trihydroxy-putrescine, trihydroxy-cadaverine, trihydroxy-GABA, trihydroxy-1,3-diaminopropane, trihydroxy-5-aminovalerate, trihydroxy-spermidine, trihydroxy-agmatine, trihydroxy-*N*-acetyl-cadaverine, trihydroxy-*N*-acetyl-putrescine, trihydroxy-*N*-acetyl-spermidine, trihydroxy-*N*-acetyl-ornithine, trihydroxy-*N*-carbamoyl-putrescine, dihydroxy-putrescine, dihydroxy-GABA, dihydroxy-1,3-diaminopropane, dihydroxy-5-aminovalerate, dihydroxy-*N*-acetyl-cadaverine, dihydroxy-*N*-acetyl-putrescine, dihydroxy-*N*-carbamoyl-putrescine.



(legend on next page)

Figure S5. Validation of polyamine bile amidates (MS level 1 annotation), related to Figure 5

Retention time matches of the (A) polyamine cholic acid amidates and (B) polyamine chenodeoxycholic acid amidates detected in two studies, MSV000080574 (blue) and MSV000086131 (pink), with synthetic standards and fecal sample extracts spiked with standards to confirm co-migration. The red cross in the figure represents no detection of those polyamine bile acid amidates in lion or mice fecal samples. MS/MS mirror plots can be visualized with the following links for the plots in the order they appear in the figure: cholyl-*N*-acetyl-putrescine (mice), cholyl-*N*-carbamoyl-putrescine (mice), cholyl-*N*-acetyl-spermidine (mice), cholyl-*N*-acetyl-spermine (mice), cholyl-agmatine (mice), cholyl-1,3-diaminopropane (mice), chenodeoxycholyl-putrescine (lion), chenodeoxycholyl-spermidine (lion), chenodeoxycholyl-*N*-acetyl-cadaverine (lion), chenodeoxycholyl-1,3-diaminopropane (mice).

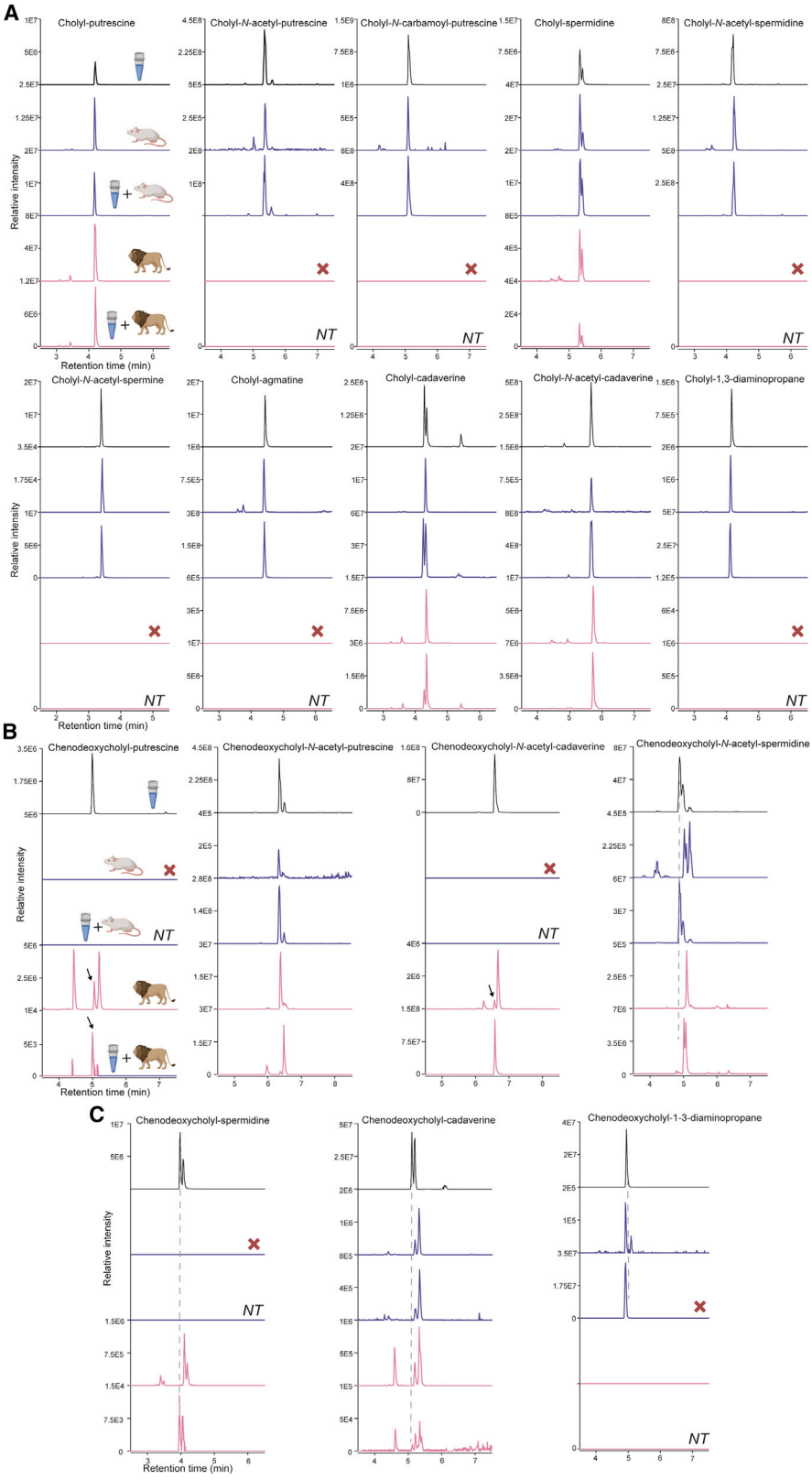


(legend on next page)

Figure S6. Validation of polyamine bile amidates (MS level 1 annotation), related to Figure 5

(A and B) Retention time matches of the (A) polyamine cholic acid amidates and (B) polyamine chenodeoxycholic acid amidates detected in two studies, MSV000080574 (blue) and MSV000086131 (pink), with synthetic standards and fecal sample extracts spiked with standards. The red cross in the figure represents no detection of those polyamine bile acid amidates in lion or mice fecal samples.

(C) Chenodeoxycholic acid amidates that do not match in retention time with synthetic standards. The red cross in the figure represents no detection of those polyamine bile acid amidates in lion fecal samples. NT: not tested.



(legend on next page)

Figure S7. Validation of polyamine bile amidates with second chromatography method, related to Figure 5

(A and B) Retention time matches with a second chromatography method for the (A) polyamine cholic acid amidates and (B) polyamine chenodeoxycholic acid amidates synthesized and detected in two studies, MSV000080574 (blue) and MSV000086131 (pink), with synthetic standards.

(C) Chenodeoxycholic acid amidates that do not match in retention time with synthetic standards. The red cross in the figure represents no detection of those polyamine bile acid amidates in lion fecal samples. NT: not tested.