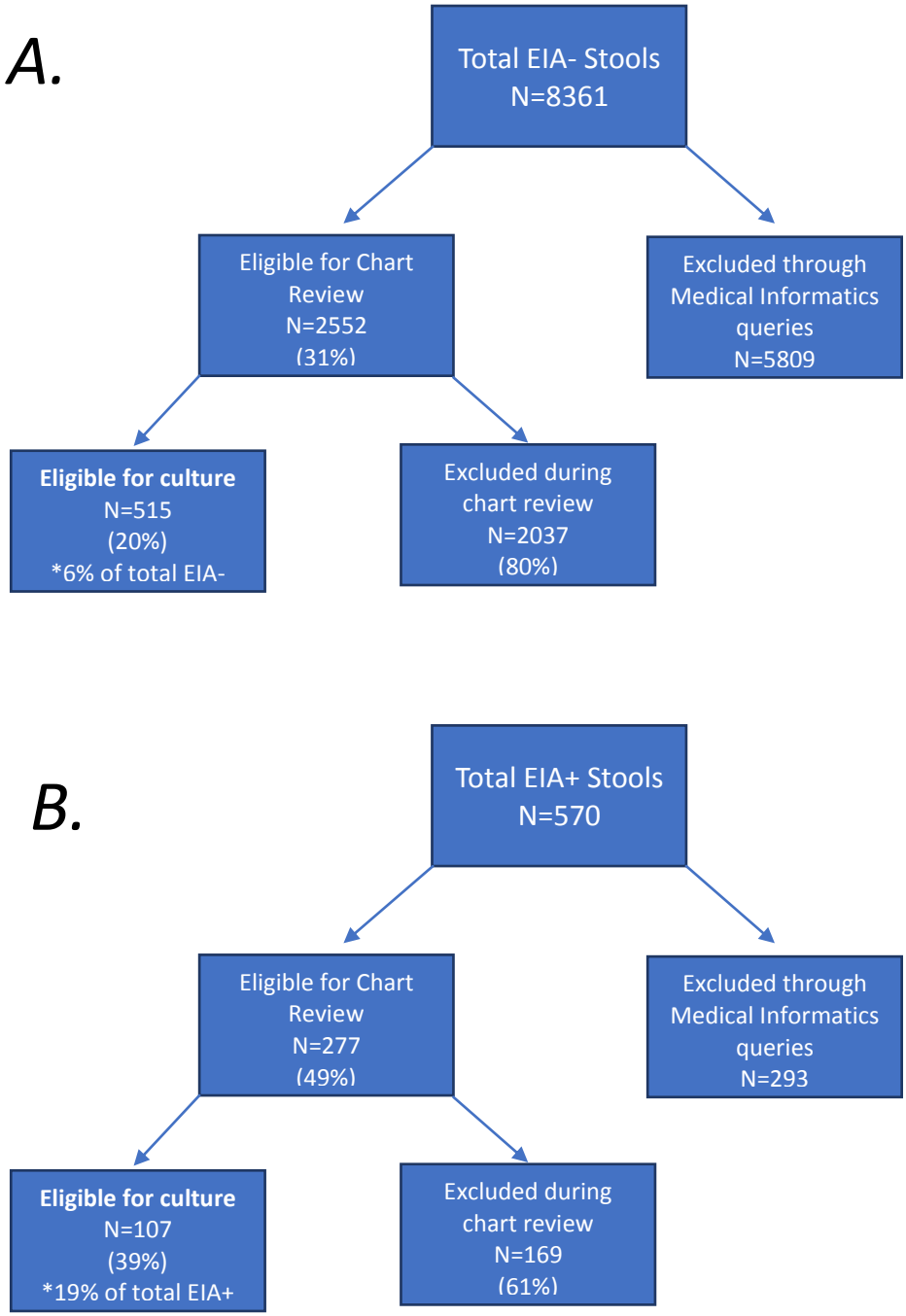


SFig. 1. Cohort assembly from clinical stool specimens. (**A**) Flow chart for selection of Cx-/EIA- and Cx+/EIA- specimens. (**B**) Flow chart for selection of Cx+/EIA+ specimens.

Supplementary Table 1. CDI-associated metabolites from cross-validation of logistic regression model.

Mass@RT	Frequency [A]	Median Odds Ratio (95% CI)	Compound
173.0@4.71	100%	1.54 (1.40 – 1.68)	4-methylpentanoic acid, TMS derivative [B]
117.0@4.70	99%	1.15 (1.04 – 1.30)	4-methylpentanoic acid, TMS derivative [B]
159.0@7.55	99%	1.13 (1.02 – 1.25)	2-hydroxy-4-methylpentanoic acid
86.0@6.75	98%	0.91 (0.84 – 0.98)	isoleucine
215.0@24.80	96%	0.87 (0.77 – 0.98)	cholenoic acid (**bile acid BA1**)
73.0@13.79	95%	0.92 (0.83 – 0.99)	ribitol
67.0@20.17	94%	1.14 (1.02 – 1.27)	eicosatrienoic acid
179.0@12.06	94%	1.11 (1.01 – 1.23)	tyrosol, 2 TMS derivative
204.0@19.45	91%	0.87 (0.76 – 0.99)	glyceryl glycoside
217.0@14.72	89%	0.92 (0.84 – 0.99)	fructose
73.0@11.41	83%	1.05 (1.00 – 1.13)	L-5-oxoproline, 2 TMS derivative
73.0@11.32	81%	1.05 (1.00 – 1.13)	*trans*-4-hydroxycyclohexanecarboxylic acid
255.0@27.45	80%	0.95 (0.88 – 1.00)	hydroxycholenoic acid (**bile acid BA2**)
73.0@14.37	69%	0.96 (0.88 – 1.00)	*no match*
73.0@11.12	68%	0.96 (0.89 – 1.00)	*no match*
205.0@8.12	63%	0.94 (0.86 – 1.00)	glycerol, 3 TMS derivative
194.0@10.01	63%	0.96 (0.90 – 1.00)	194.0@10.011011
79.0@19.99	59%	1.04 (1.00 – 1.14)	arachidonic acid, TMS derivative
142.0@12.16	57%	0.96 (0.88 – 1.00)	*no match*
73.0@15.99	56%	0.97 (0.90 – 1.00)	D-sorbitol, 6 TMS derivative
217.0@12.74	47%	0.97 (0.90 – 1.00)	D-(-)-ribofuranose, tetrakis(trimethylsilyl) ether (isomer 2)
75.0@6.77	43%	1.03 (1.00 – 1.13)	indole-3-acetic acid
202.0@16.29	43%	1.03 (1.00 – 1.11)	*no match*
73.0@16.17	33%	0.97 (0.90 – 1.00)	*no match*
145.0@6.51	30%	1.03 (1.00 – 1.12)	2-hydroxy-3-methylbutyric acid, 2 TMS derivative
255.0@25.45	25%	0.97 (0.90 – 1.00)	deoxycholic acid (**bile acid BA3**)
158.0@8.35	24%	1.03 (1.00 – 1.13)	*allo*-isoleucine, 2 TMS derivative
84.0@11.56	24%	0.98 (0.92 – 1.00)	*no match*
73.0@8.90	23%	0.98 (0.91 – 1.00)	glyceric acid, 3 TMS derivative
217.0@15.71	22%	0.98 (0.92 – 1.00)	meglumine, 5 TMS derivative
204.0@15.65	22%	0.98 (0.92 – 1.00)	*no match*
147.0@9.95	21%	0.98 (0.91 – 1.00)	D-(-)- erythrofuranose, tris(trimethylsilyl) ether (isomer 2)
204.0@15.13	20%	0.98 (0.92 – 1.00)	*no match*
73.0@12.53	18%	1.03 (1.00 – 1.12)	*alpha*-arabinopyranose, 4 TMS derivative
179.0@14.22	18%	0.98 (0.92 – 1.00)	3-(4-hydroxyphenyl)propanoic acid (phloretic acid) [C]
73.0@12.86	18%	0.98 (0.92 – 1.00)	*no match*
230.0@11.47	16%	0.98 (0.90 – 1.00)	*no match*
73.0@12.39	16%	0.98 (0.91 – 1.00)	*no match*
73.0@11.88	16%	0.98 (0.93 – 1.00)	*no match*
174.0@15.81	14%	1.02 (1.00 – 1.08)	tyramine
151.0@4.93	13%	0.98 (0.90 – 1.00)	phenol, TMS derivative

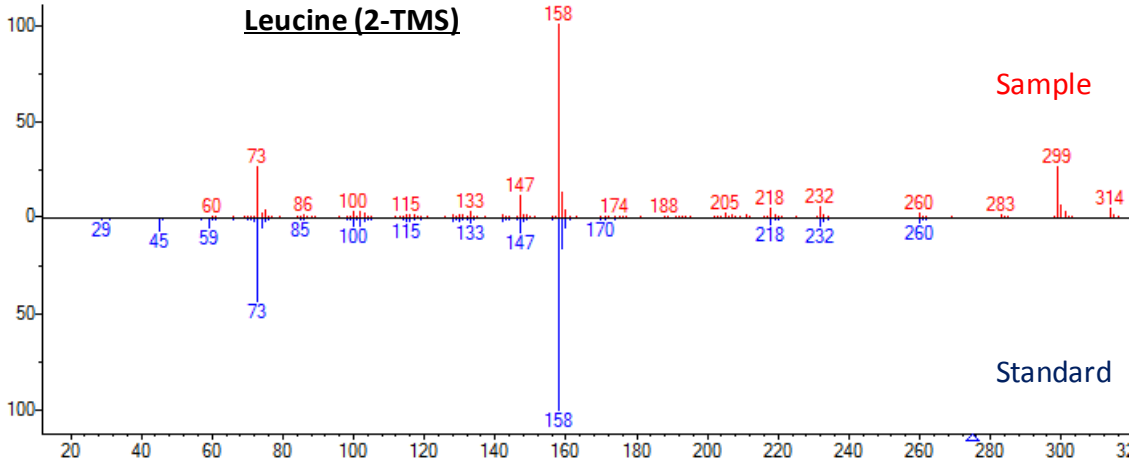
Mass@RT	Frequency [A]	Median Odds Ratio (95% CI)	Compound
86.0@6.44	12%	0.98 (0.92 – 1.00)	L-leucine, TMS derivative
73.0@14.90	11%	0.98 (0.92 – 1.00)	*no match*

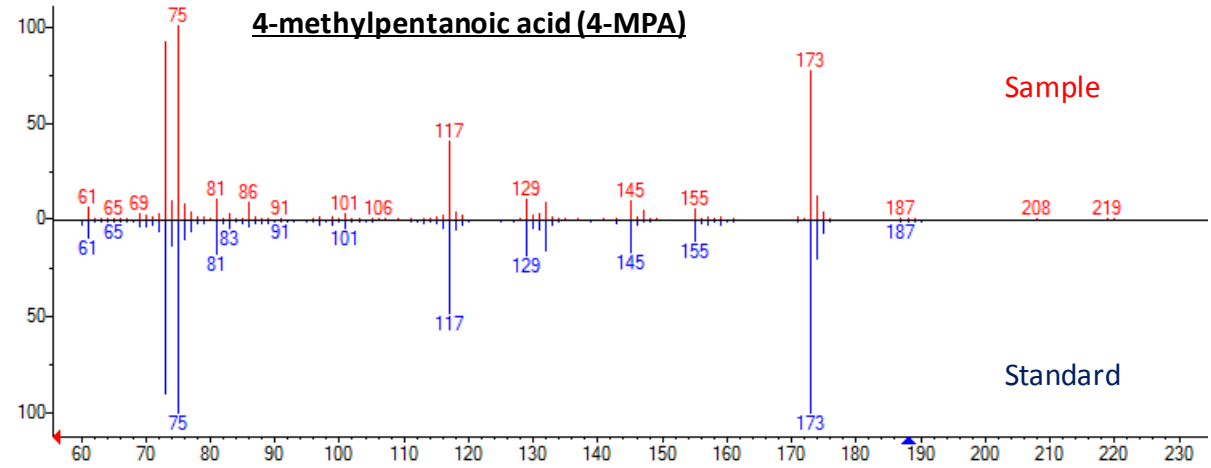
[A] Proportion of cross-validation models including the metabolite
[B] Two ions from same metabolite were independently resolved
[C] This is also referred to as desamino tyrosine

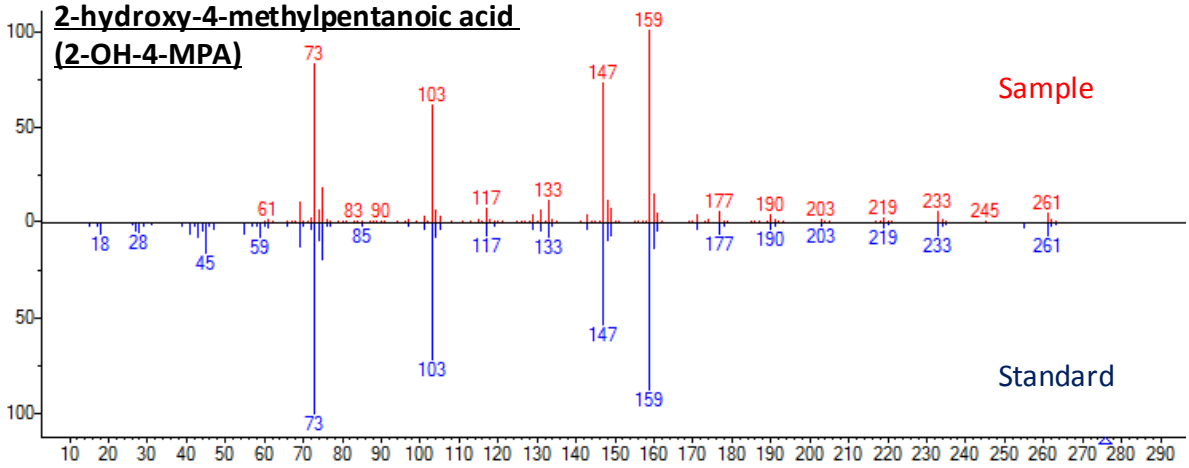
Supplementary Table 2. Metabolites quantified by selective ion monitoring (SIM).

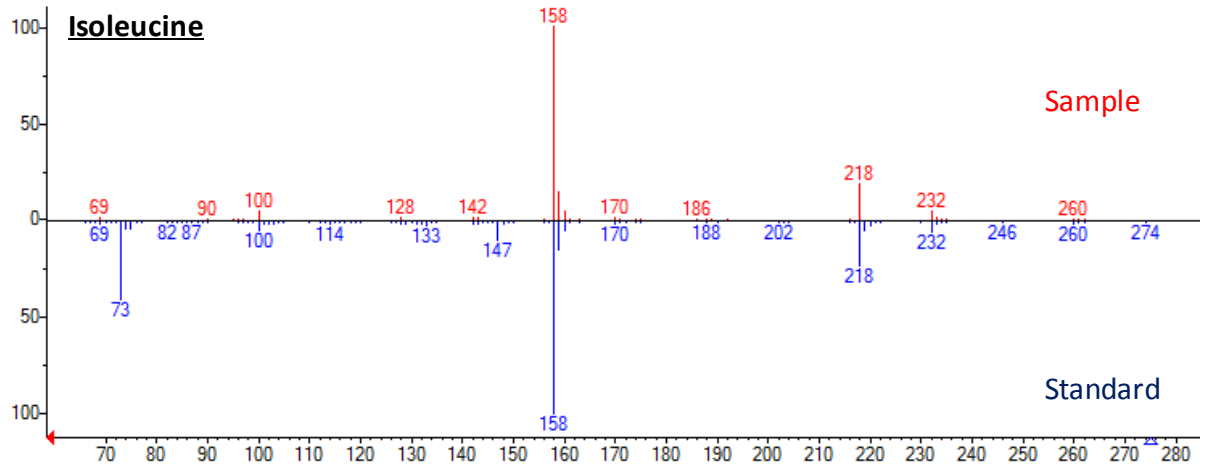
Compound	***m/z***	**GC retention time (min)**
4-methylpentanoic acid (4MP, isocaproic acid)	173	4.73
2-hydroxy, 4-methylpentanoic acid (2OH-4MP)	159	7.54
2-hydroxy, 3-methylpentanoic acid (2OH-3MP)	159	7.59
2-(4-hydroxyphenyl)acetic acid	179	12.8
3-(4-hydroxyphenyl)propionic acid (phoretic acid)	179	14.1
allo-isoleucine	158	8.37
isoleucine	158	8.4
leucine	158	8.1
tyrosine	218	16.1
cholic acid (CA)	253	27.5
cholenoic acid (CE)	215, 430	24.7
trehalose	361	22.8
^{13}C-trehalose	367	22.8

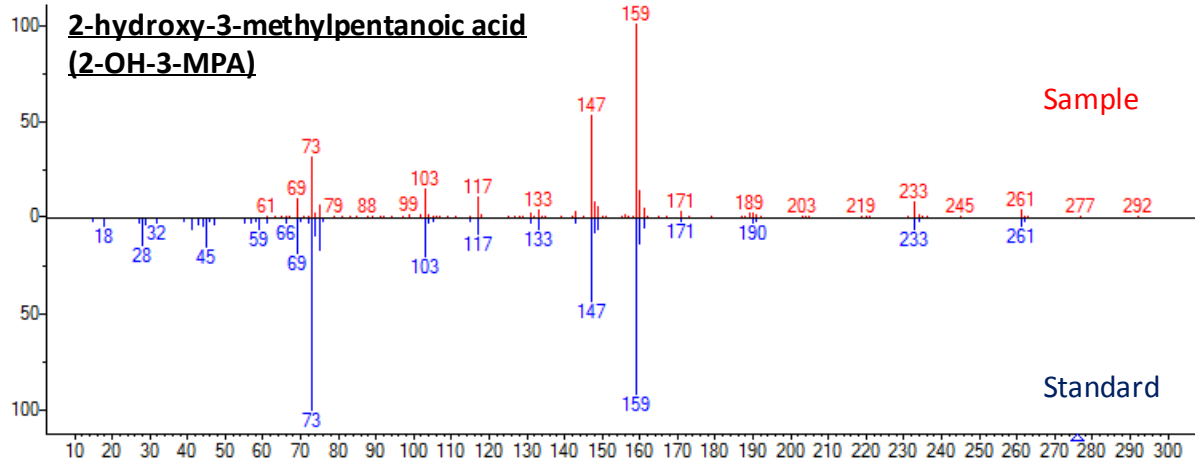
Identification of Stickland metabolites in the fecal metabolome

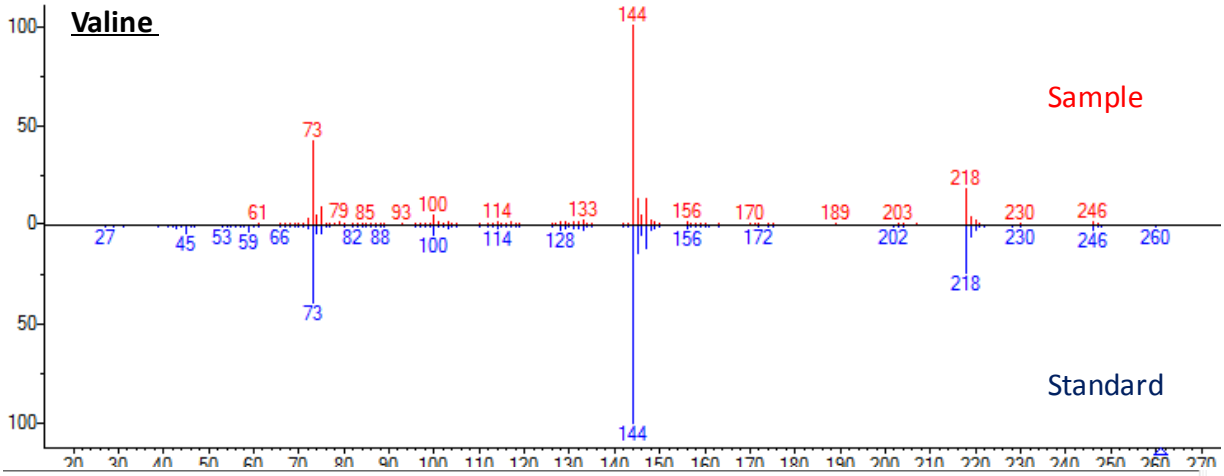


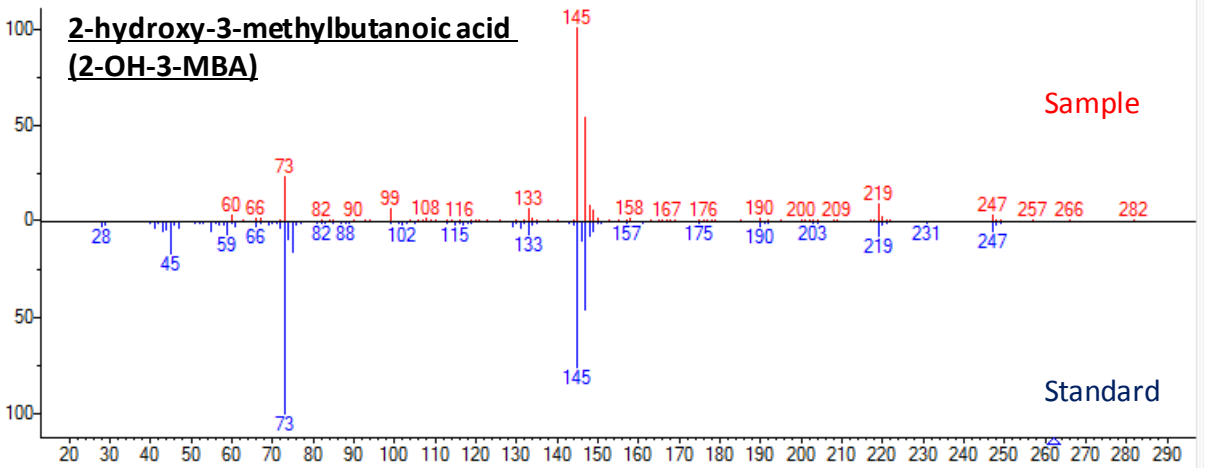


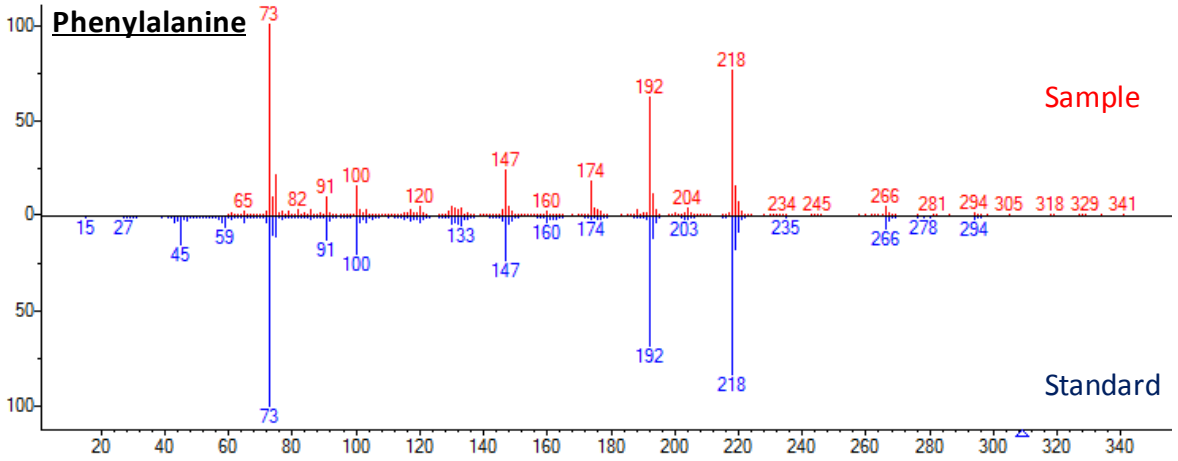


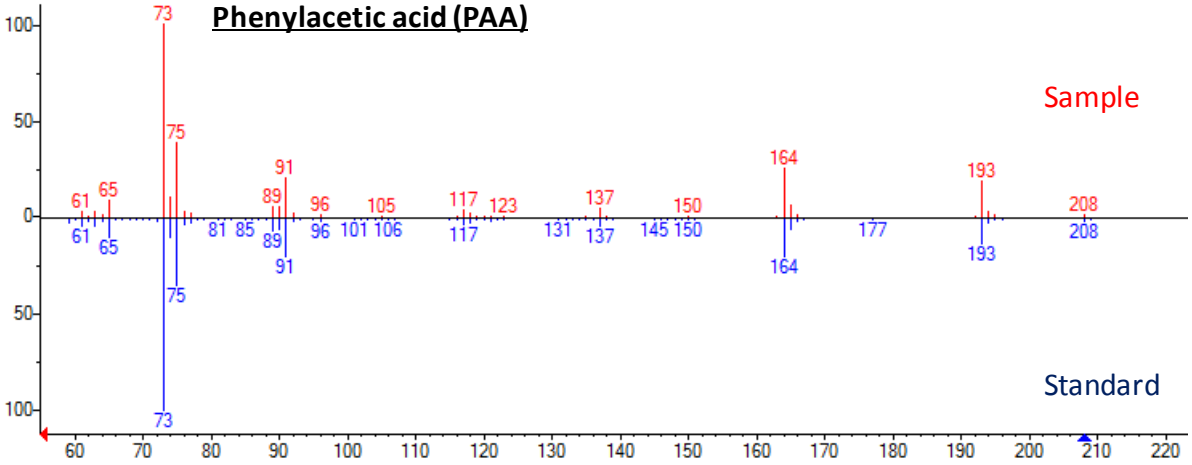

Isoleucine
Sample
Standard

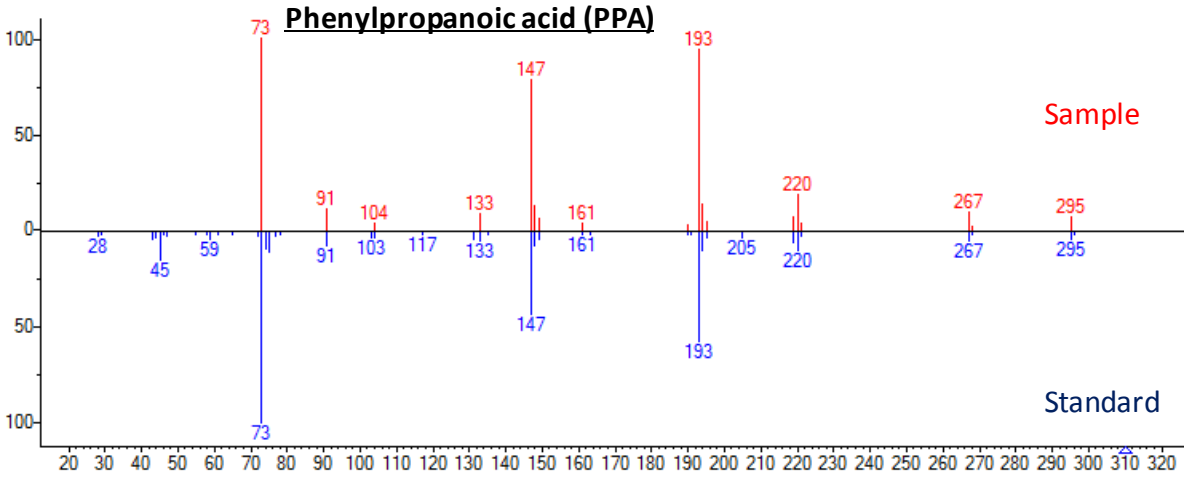

2-hydroxy-3-methylpentanoic acid
(2-OH-3-MPA)
Sample
Standard

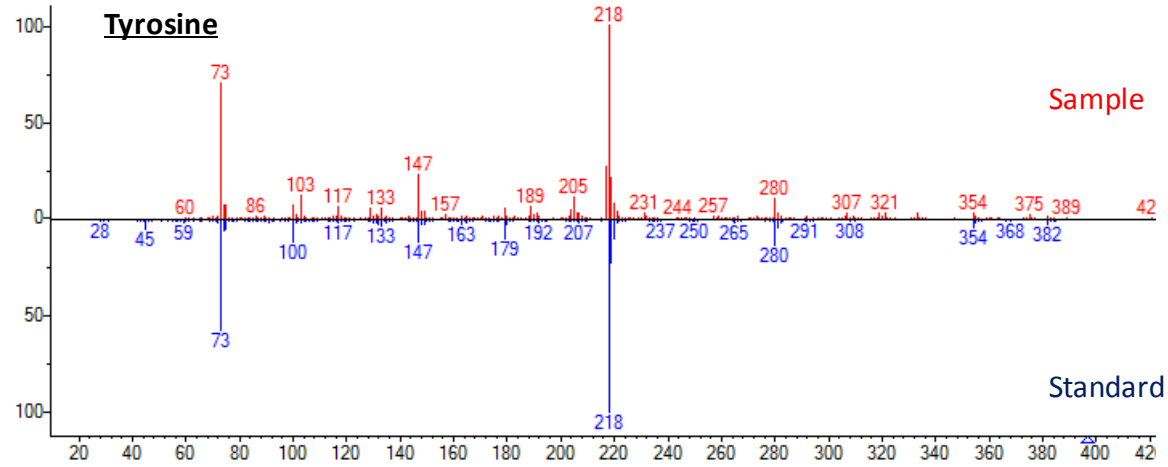

Valine
Sample
Standard

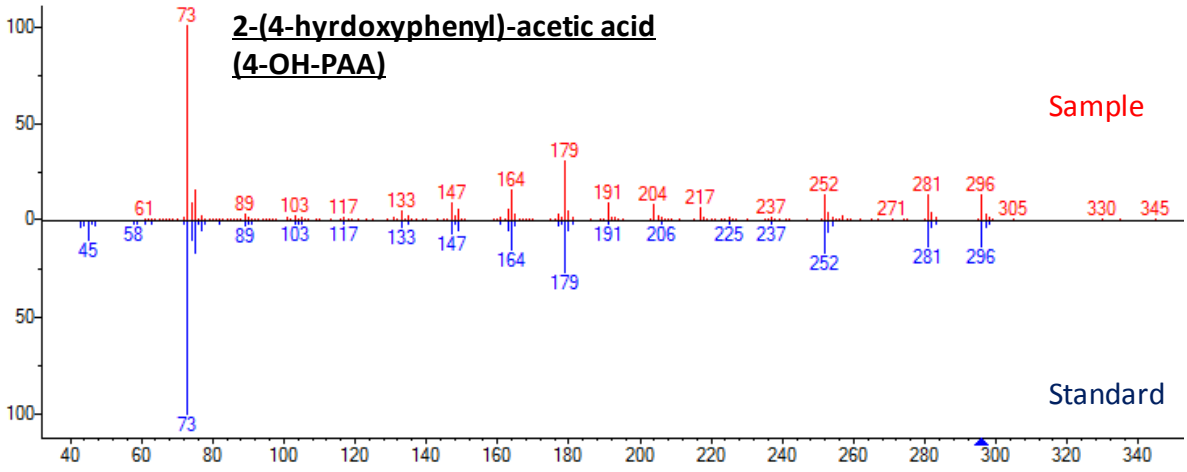

2-hydroxy-3-methylbutanoic acid
(2-OH-3-MBA)
Sample
Standard

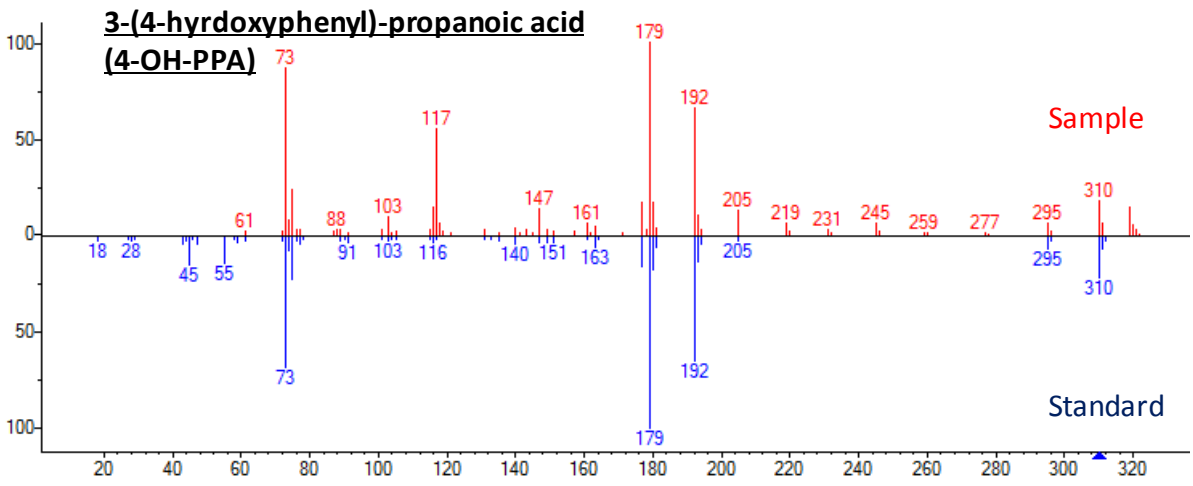

Phenylalanine
Sample
Standard

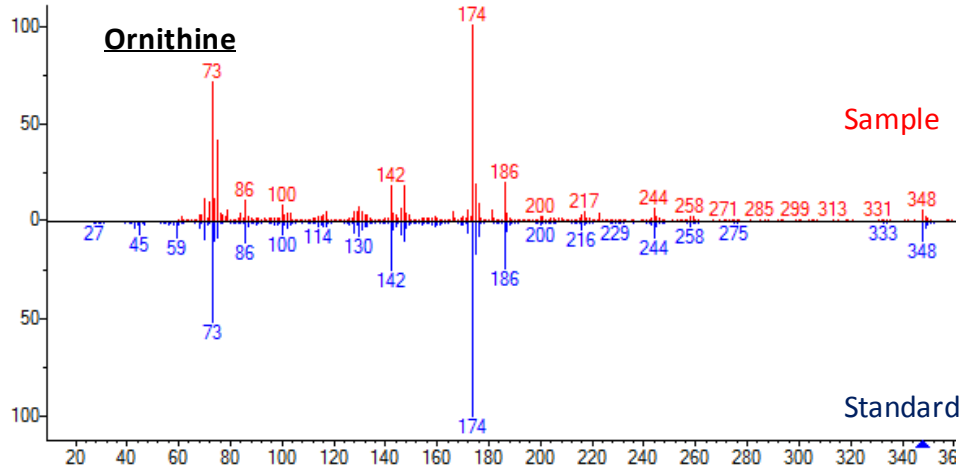

Phenylacetic acid (PAA)
Sample
Standard

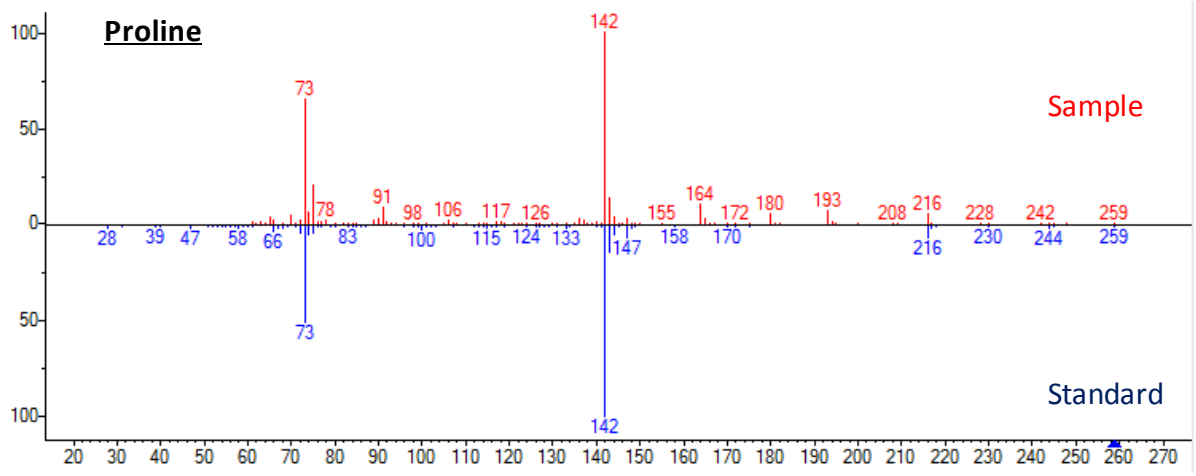

Phenylpropanoic acid (PPA)
Sample
Standard

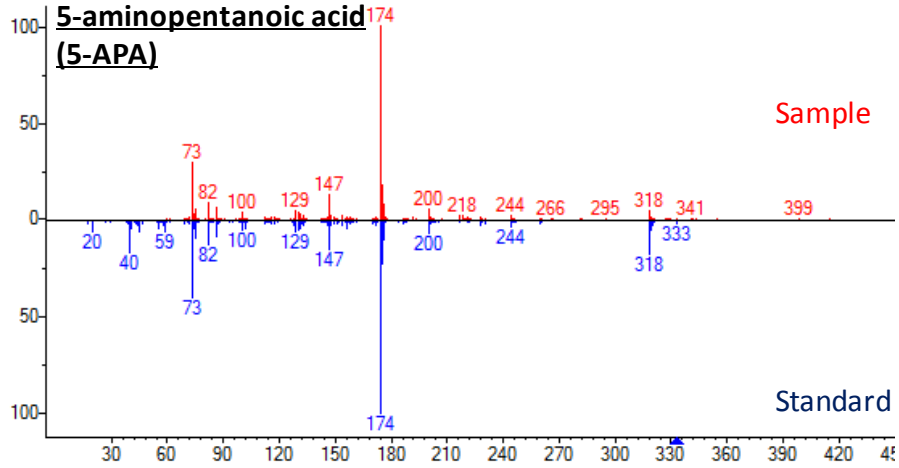

Tyrosine
Sample
Standard

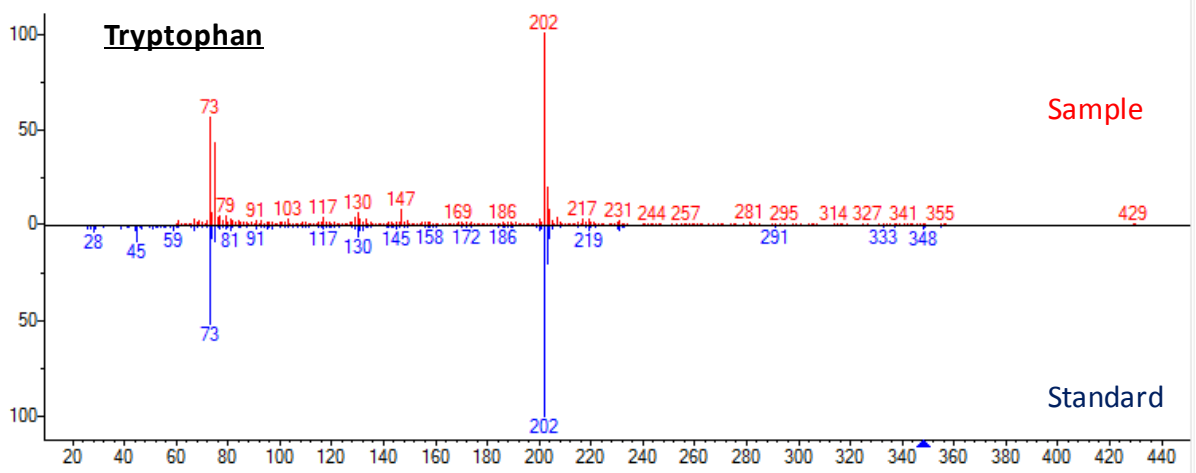

2-(4-hyrdoxyphenyl)-acetic acid
(4-OH-PAA)
Sample
Standard

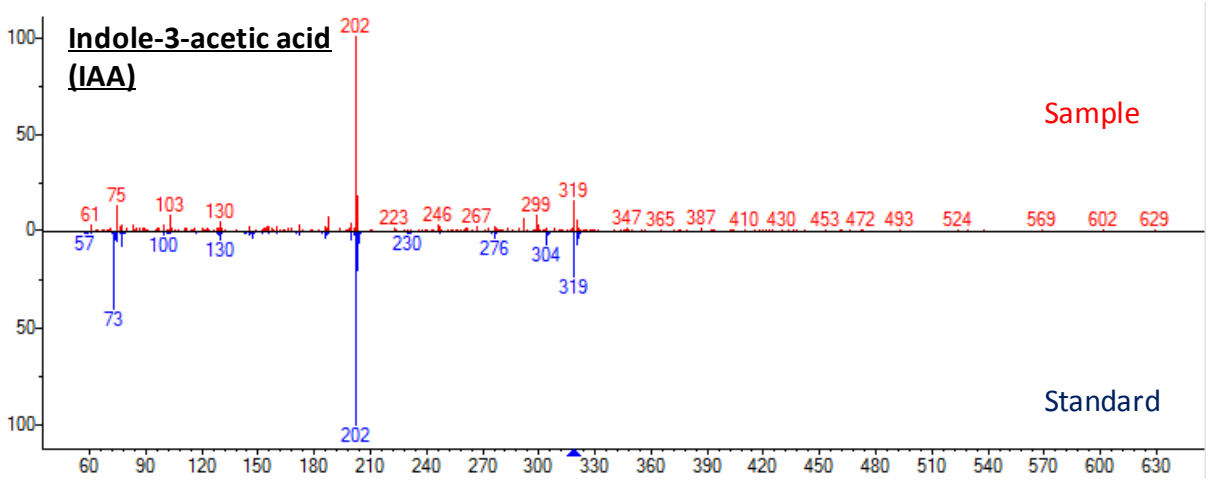

3-(4-hyrdoxyphenyl)-propanoic acid
(4-OH-PPA)
Sample
Standard


Ornithine
Sample
Standard


Proline
Sample
Standard


5-aminopentanoic acid
(5-APA)
Sample
Standard

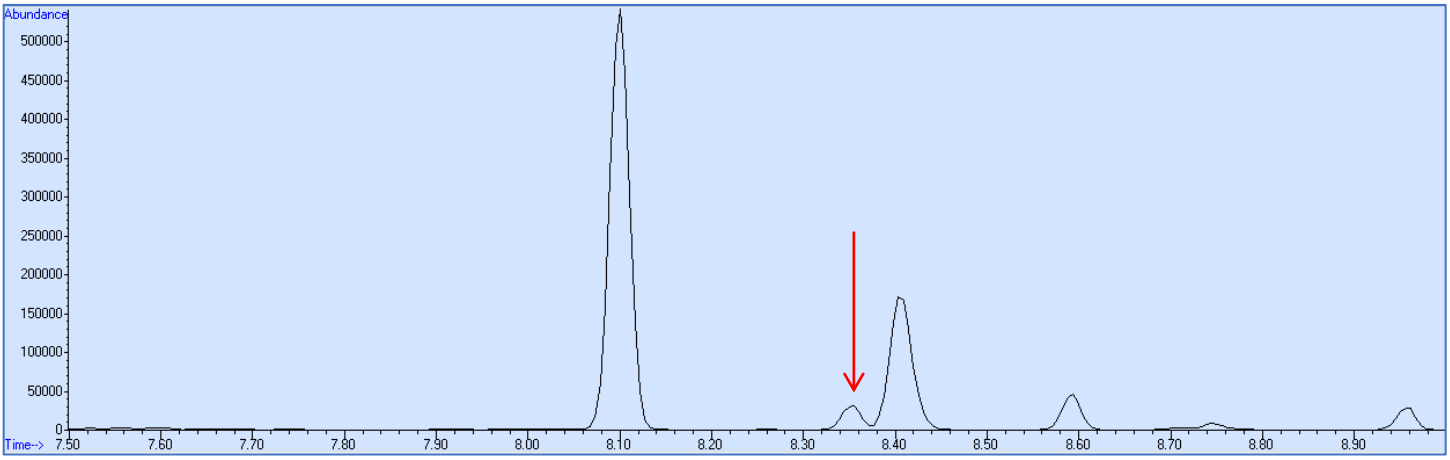

Tryptophan
Sample
Standard



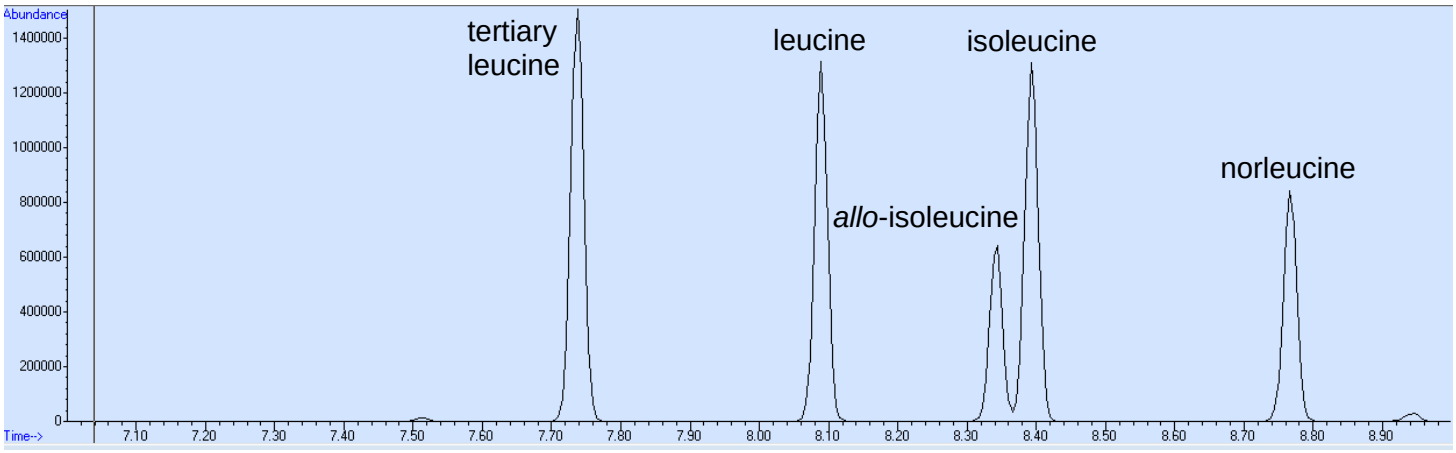
SFig 2. Shown here are the GC-MS spectra for each detected Stickland metabolite (red) compared to the standard spectra in the NIST 14 library. Sample spectra were obtained from a pooled quality control sample using the AMDIS GC-MS software package.

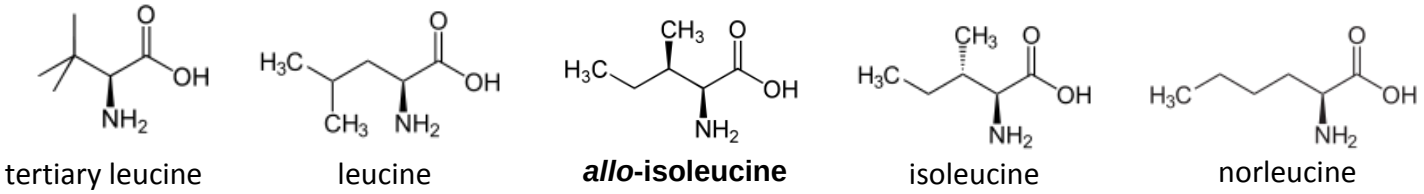
Identification of *allo*-isoleucine in the fecal metabolome

Among the metabolites associated with Cx+/EIA+ were two closely eluting peaks with mass spectra matching that of isoleucine. were two chromatographically-distinct molecules with mass spectra matching isoleucine (**SFig 3**). The first of these metabolites was positively-associated with CDI (**STable 1**). With this GC-MS method, differing peaks with identical mass spectra can result from the presence of isomers (but not enantiomers). When we compared these peaks with authentic leucine, tertiary leucine, and *allo*-isoleucine, the first peak matched L-leucine and the second peak matched *allo*-isoleucine.

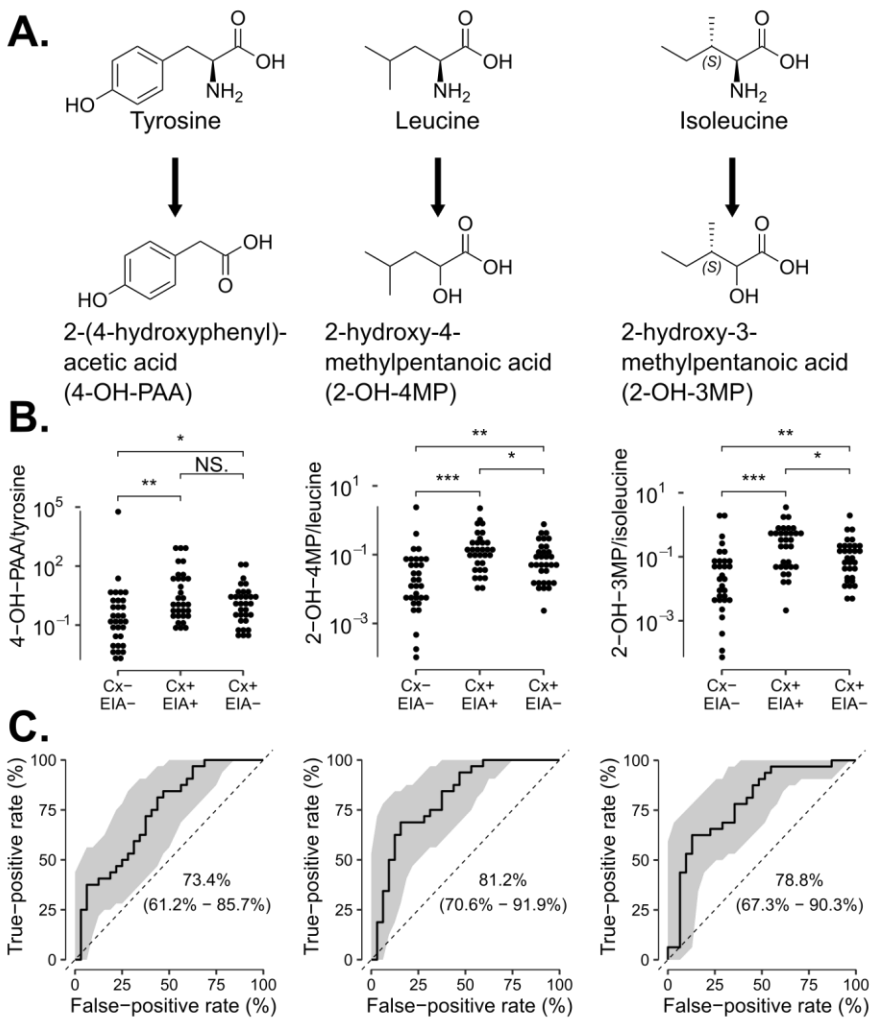


SFig. 3. Shown here is a selected ion (*m/z* 158) GC-MS chromatogram at from a Cx+/EIA+ specimen showing a peak (arrow) associated with Cx+/EIA+ specimens in multivariate analyses. Both this and the immediately subsequent peak at 8.4 min (identified as isoleucine) possess similar EI mass spectra and match the NIST14 Library for isoleucine. The peak at 8.1 min corresponds to isoleucine based on comparison to an authentic standard.





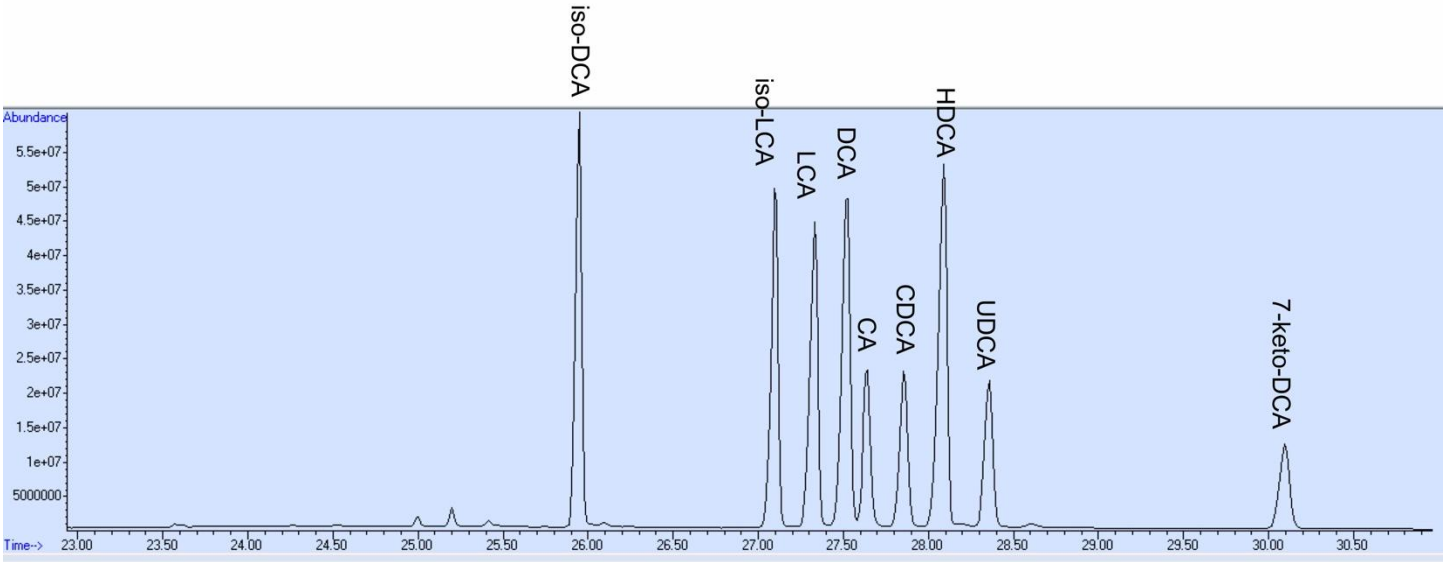
SFig. 4. Shown here is a selected ion (*m/z* 158) GC-MS chromatogram of isoleucine isomers. Of these, *allo*-isoleucine yields a peak that elutes immediately prior to isoleucine, possesses a similar mass spectrum, and corresponds to the Cx+/EIA+-associated peak noted above in **SFig. 3**.



SFig. 5. This figure corresponds to Fig. 3 in the main text. Shown here are the product/precursor ratios of several Stickland metabolites monitored by SIM. (A) Structures of Stickland precursors and products. (B) Dot plots of product/precursor ratios for all three patient groups. (C) ROC plots showing true- and false-positive rates at varying threshold product/precursor ratios. The area under the curves and 95% confidence intervals are displayed on the plots.

Identification of fecal bile acids detected by GC-MS

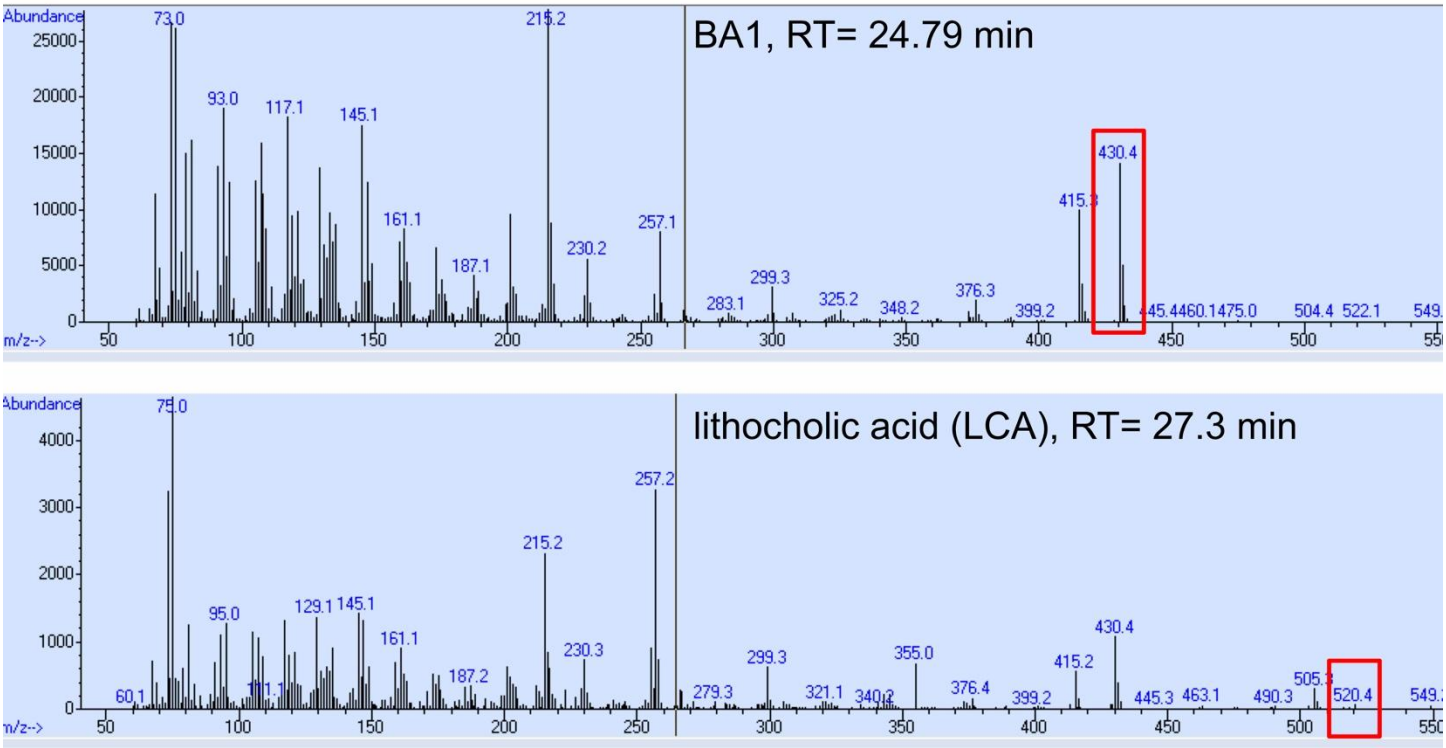
Three bile acid features (BA1, 2, 3; **supplementary Table 3**) were identified among the top 26 cross-validated CDI correlates (**Table 2, supplementary Table 1**). Using the NIST 14 database, the closest spectral match to the first feature was lithocholic acid (LCA), with deoxycholic acid (DCA) as the closest match to the next two features. Upon comparison of these three features with authentic standards, only the third feature (BA3, m/z 255, RT 27.45) exhibited a mass spectrum and retention time matching its preliminary identification (DCA). BA1 and BA2 both exhibited shorter (by 2 and 2.5 minutes) retention times than their closest NIST14 matches (LCA and DCA, respectively) indicating incorrect, but structurally related, metabolite identification. Neither BA1 nor BA2 co-eluted with authentic standards of isodeoxycholic, isolithocholic, lithocholic, deoxycholic, cholic, chenodeoxycholic, hyodeoxycholic, ursodeoxycholic, or 7-ketodeoxycholic acids.



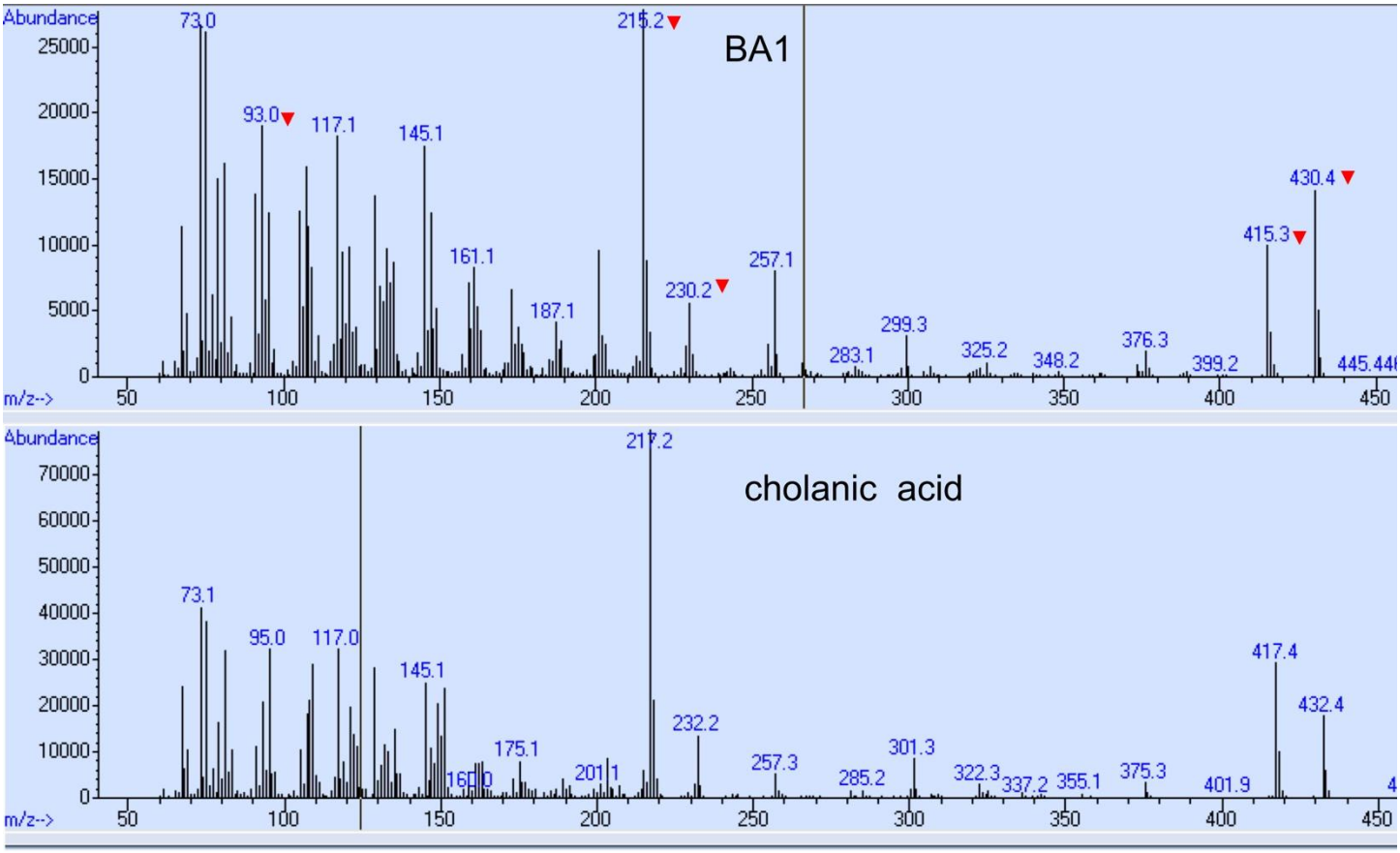
SFig. 6. GC-MS chromatogram of authentic isodeoxycholic acid (iso-DCA), isolithocholic acid (iso-LCA), lithocholic acid (LCA), deoxycholic acid (DCA), cholic acid (CA), chenodeoxycholic acid (CDCA), hyodeoxycholic acid (HDCA), ursodeoxycholic acid (UDCA), or 7-ketodeoxycholic acid (7-keto-DCA).

The BA1 EI mass spectrum shares multiple high intensity ions with LCA, suggesting shared chemical features (**SFig. 7**). Notably, however, the LCA molecular ion (520.4 m/z) and its $[M-15]^{\bullet+}$ fragment ion were absent. Instead, the EI spectrum exhibited putative molecular and $[M-15]^{\bullet+}$ fragment ions with lower molecular weights of *m/z* 430.4 and 415.2. This corresponds to a difference of 90 m/z units between LCA and BA1, consistent with an LCA derivative lacking one TMS-derivatized oxygen and one hydrogen. This corresponds to a non-hydroxylated bile acid. Cholanic acid, a non-hydroxylated bile acid (**SFig. 8**), exhibited a retention time very

similar (0.16 min longer, **SFig. 6**) to that of BA1 with a similar mass spectrum pattern that differed in that several major ions were 2 *m/z* units greater than those of BA1. Together these results are consistent with very similar chemical structures that differ in the presence of one double bond. We assign the structure of BA1 as cholenoic acid, a monounsaturated cholanic acid derivative.

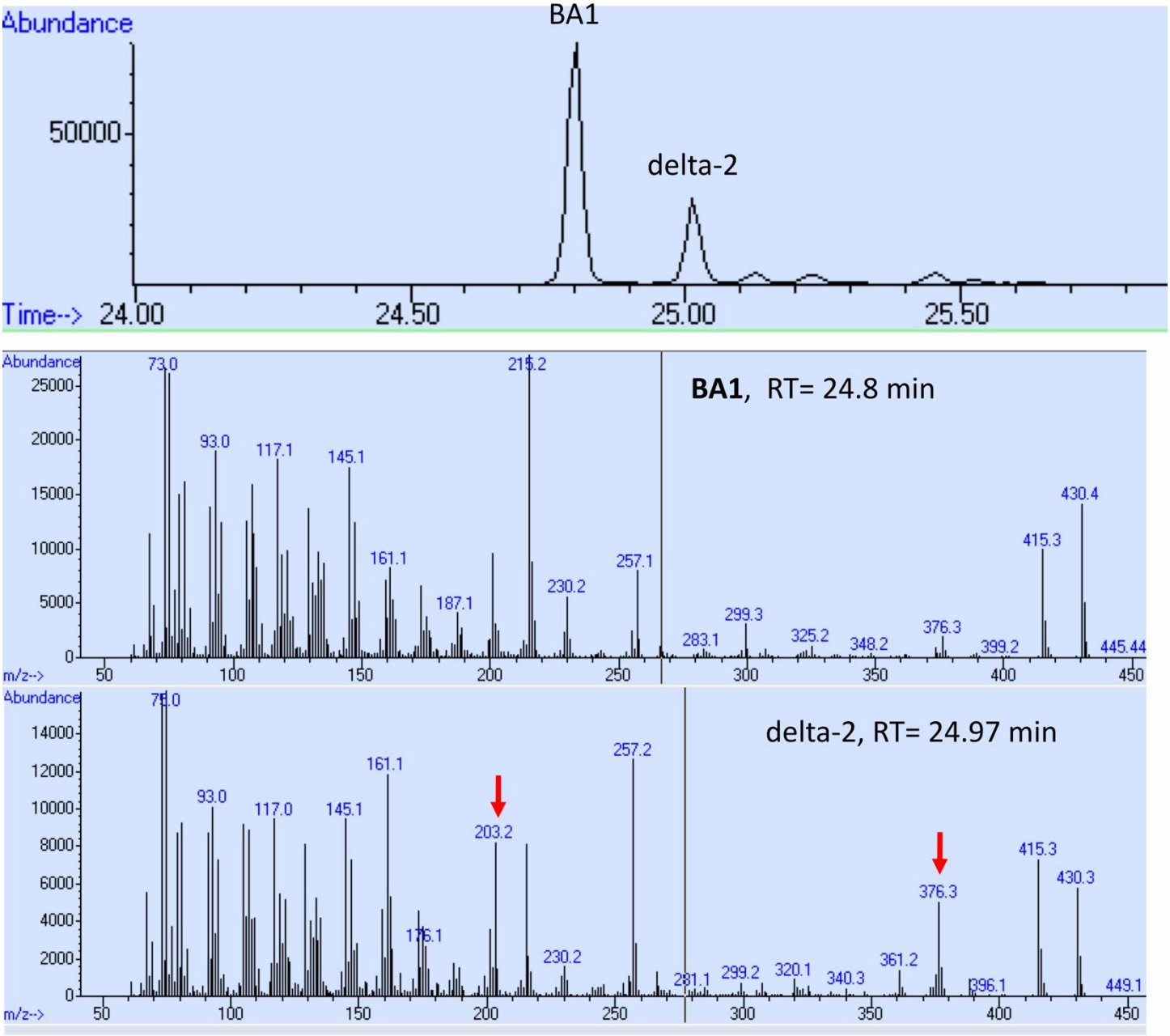


SFig. 7. EI-MS spectra of the fecal metabolite BA1 (top) and authentic lithocholic acid (bottom). Molecular ions are indicated in the red boxes. RT: gas chromatographic retention time.

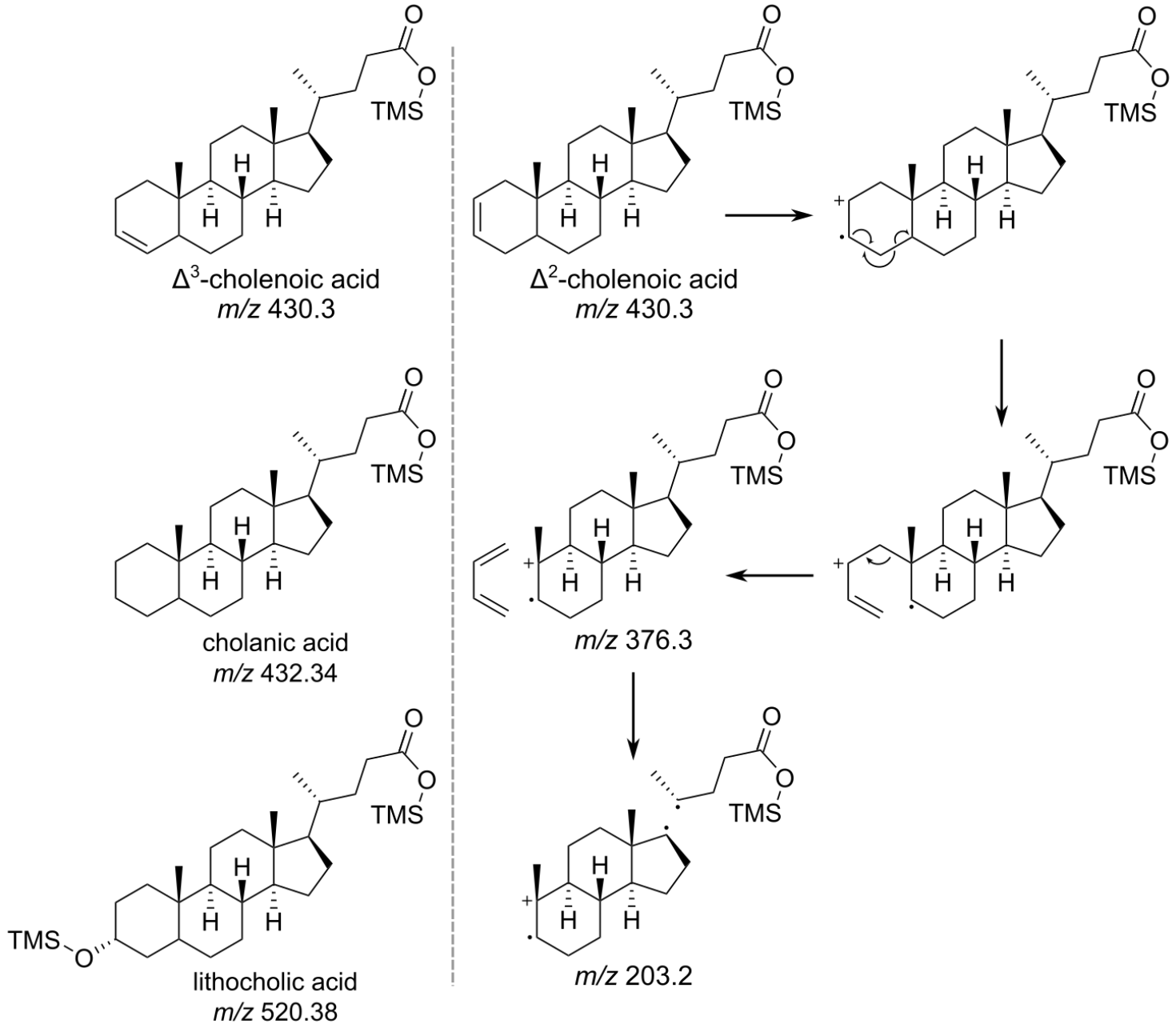


SFig. 8. EI-MS spectra of the fecal metabolite BA1 (top) and authentic cholanic acid (bottom).

Selected-ion monitoring of *m/z* 430 (the BA1 molecular ion) resolved a second, closely eluting peak (0.2 minutes later) in some samples that exhibits a highly similar EI mass spectrum (**SFig. 9**). These two peaks resemble the two cholenoic acid isomers differing by double bond location previously reported as *Fusobacterium* products (1). Careful comparison of the EI spectra between these two metabolites revealed prominent quantitative intensity differences in ions at *m/z* 203.2 and 376.3. The more intense ion at *m/z* 376.3 in the second isomer's spectrum can be rationalized as the product of the more facile retro-Diels-Alder fragmentation of delta-2-cholenoic acid, suggesting that BA1 is the delta-3 isomer instead (**SFig. 10**). Together, these studies support diminished delta-3-cholenoic acid as a CDI correlate.



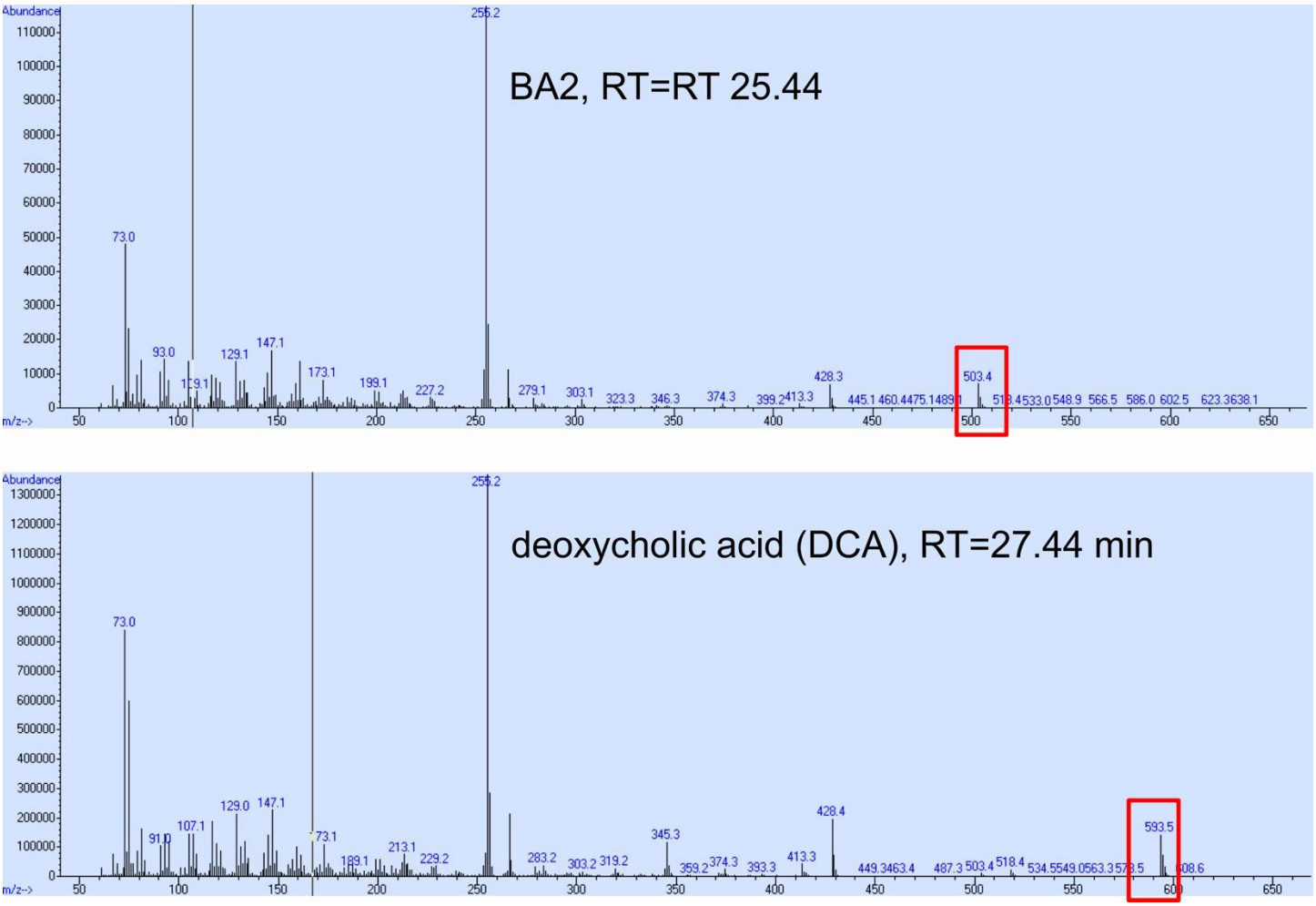
SFig. 9. GC-MS selected ion chromatogram at *m/z* 430 of fecal specimen showing BA1 with closely-eluting peak (delta-2) (*top panel*). EI mass spectra of BA1 (*middle panel*) and delta-2 (*bottom panel*). Red arrows denote peaks with quantitative differences between the two features.



SFig. 10. Proposed, favored retro-Diels-Alder fragmentation mechanism for delta-2-cholenoic acid.

The BA2 EI mass spectrum largely resembled that of DCA, also suggesting shared chemical features (**SFig. 11**). Notably, however, the DCA molecular ion (*m/z* 593.5) was absent. Instead, the EI spectrum of BA2 exhibited a putative molecular ion at *m/z* 503.4, a difference of 90 m/z units between DCA and BA2. As for BA1, this is consistent with a DCA derivative lacking one TMS-derivatized oxygen and one hydrogen. The relationship between BA2 and DCA spectra is very similar to the difference between the spectra of BA3 (delta-3-cholenoic acid) and LCA. This supports the identity of BA3 as a monounsaturated, monohydroxylated bile

acid. There is no commercially available reference standard with this structure. In the absence of supportive data on double bond or hydroxyl location, we refer to this bile acid as hydroxycholenoic acid (**STable 3**).



SFig. 11. EI-MS spectra of the fecal metabolite BA2 (*top*) and authentic deoxycholic acid (*bottom*). Molecular ions are indicated in the red boxes. RT, gas chromatographic retention time.

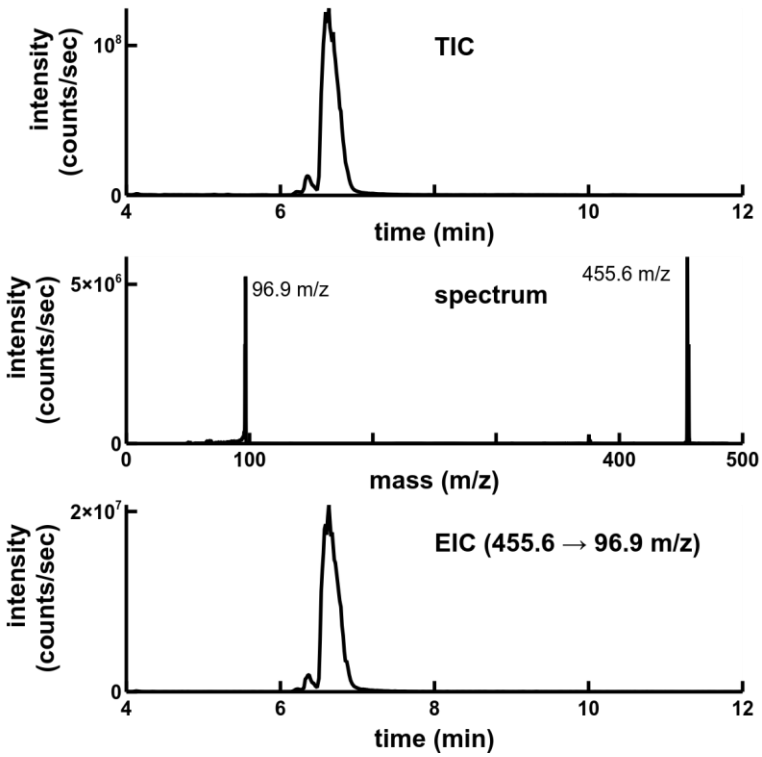
Supplemental Table 3. Identification bile acids among cross-validated CDI features.

feature	retention time	base peak (*m/z*)	feature identification
BA1	24.8 min	215	delta-2-cholenoic acid
BA2	25.44 min	255	hydroxycholenoic acid
BA3	27.44 min	255	deoxycholic acid (DCA)

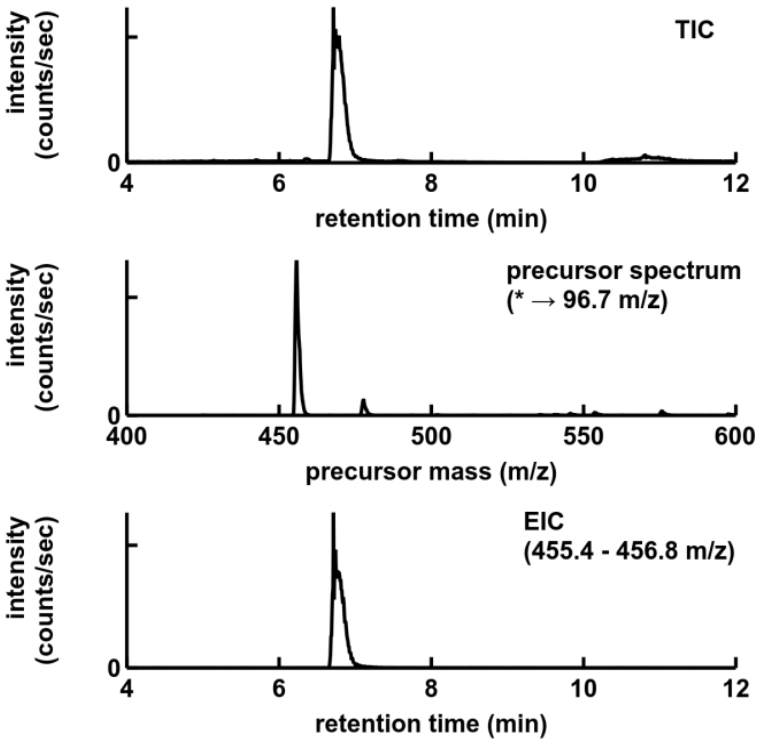
Measurement of bile acids by liquid chromatography-mass spectrometry

Motivated by the association of LCA-S degradation products with CDI, we screened patient samples for the presence of LCA-S and other sulfated bile acids. Lithocholic acid-3-sulfate (LCA-S) was initially detected in

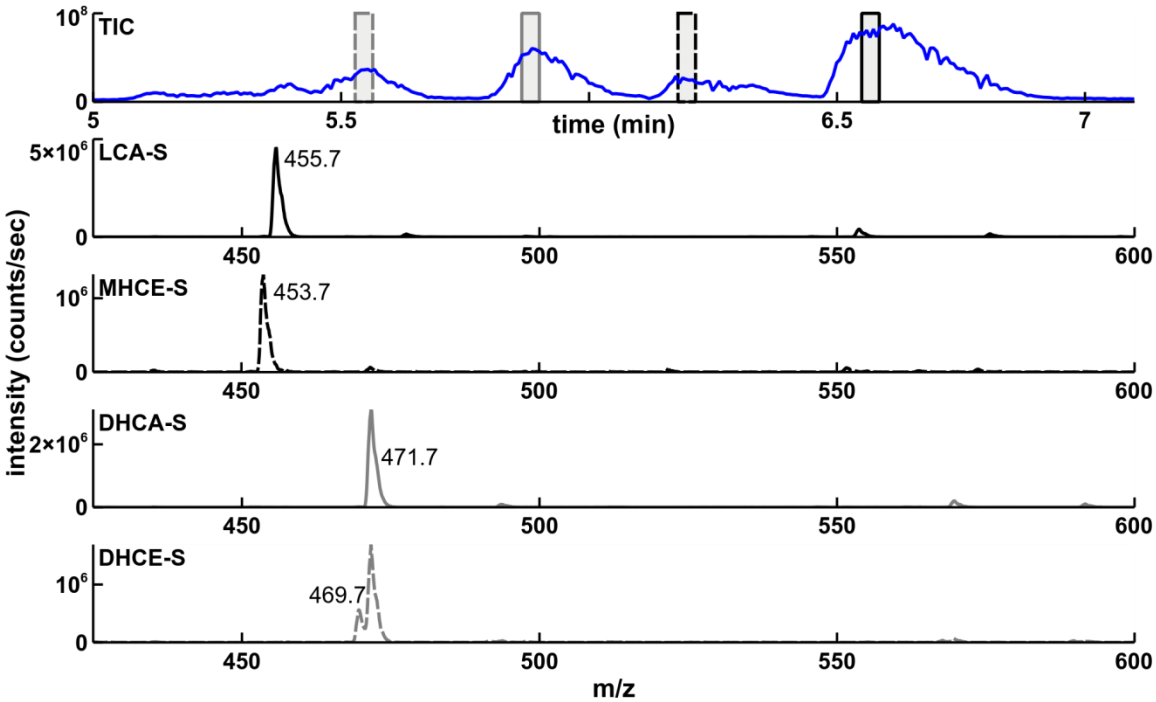
patient samples based on its characteristic 96.9 m/z product ion, which represents the sulfate group (**SFig 12**). An authentic standard of LCA-S confirmed the 96.9 m/z product ion and the retention time of LCA-S (**SFig 13**). A precursor ion scan of patient samples using this sulfate ion revealed multiple sulfated compounds (**SFig 14**) with masses corresponding to isomers of lithocholic acid sulfate (LCA-S; 455.8 m/z), deoxycholic acid sulfate (DHCA-S; 471.7 m/z), unsaturated lithocholic acid sulfate (MHCE-S; 453.8 m/z), and unsaturated deoxycholic acid sulfate (DHCE-S; 469.7 m/z). These presumed bile acid sulfates were measured by LC-ESI-MS/MS using parameters reported in **Supplemental Table 4**. Multiple peaks with the same precursor mass were presumed to represent structural isomers and were measured separately.



SFig 12. Product ion spectrum of LCA-S in a representative patient sample. The LCA-S spectrum exhibits a 96.9 m/z fragment characteristic of sulfated bile acids.



SFig 13. Precursor ion spectrum of LCA-S standard from its aliphatic sulfate product ion (96.7 m/z)



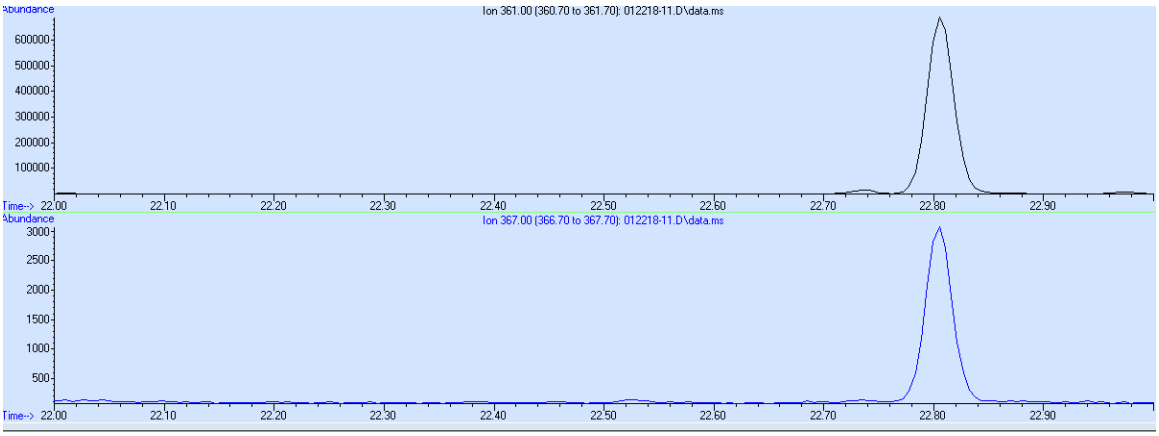
SFig 14: Precursor ion spectra of bile acid sulfates in a representative patient sample. Shaded boxes in top panel are time periods from which the corresponding spectra were extracted in subsequent panels.

Supplementary Table 4. Instrument settings for detecting bile acids via LC-MS/MS.

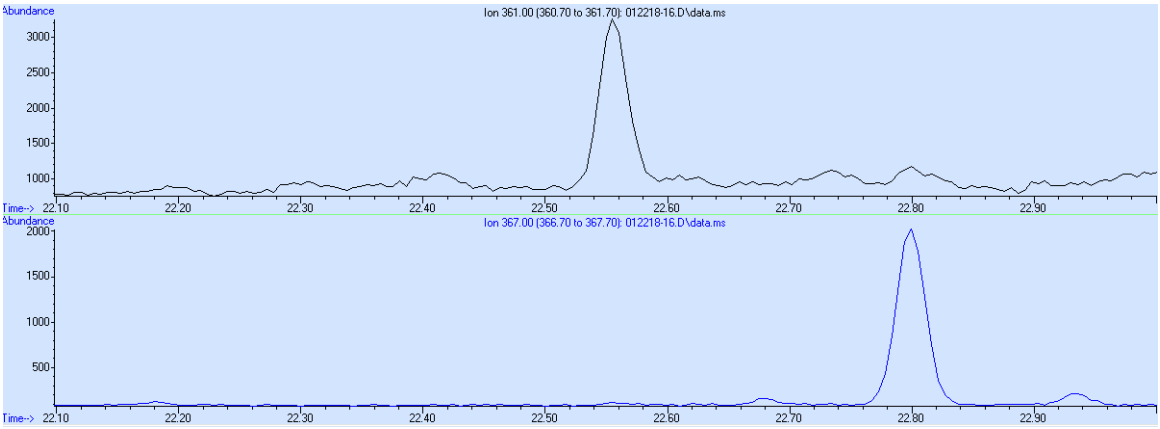
Instrument Settings			
Ion Spray Voltage		-4.5 kV	
Heater Temperature		500 °C	
Nebulizer Gas		40	
Auxiliary Gas		40	
Collision Cell Exit Potential		-12 V	
Declustering Potential		-10 V	
Abbreviation	**Precursor (m/z)**	**Product (m/z)**	**Collision Energy (V)**
MCA	407	387	-45
CA	407	343	-50
CDCA	391	391	-50
DCA	391	345	-50
MHCE-S	453.8	96.9	-80
DHCE-S	469.7	96.9	-80
G-CA	464	402	-50
G-CDCA	448	386	-50
G-DCA	448	402	-50
G-LCA	432	388	-45
G-UDCA	448	386	-50
HDCA	391	373	-45
LCA	375	375	-50
LCA-S	455.8	96.9	-80
T-CA	514	514	-50
T-CDCA	498	107	-80
T-DCA	498	107	-80
T-LCA	482	107	-80
T-UDCA	498	107	-80
UDCA	391	391	-50
DHCA-S	471.7	96.9	-80

Identification of fecal trehalose

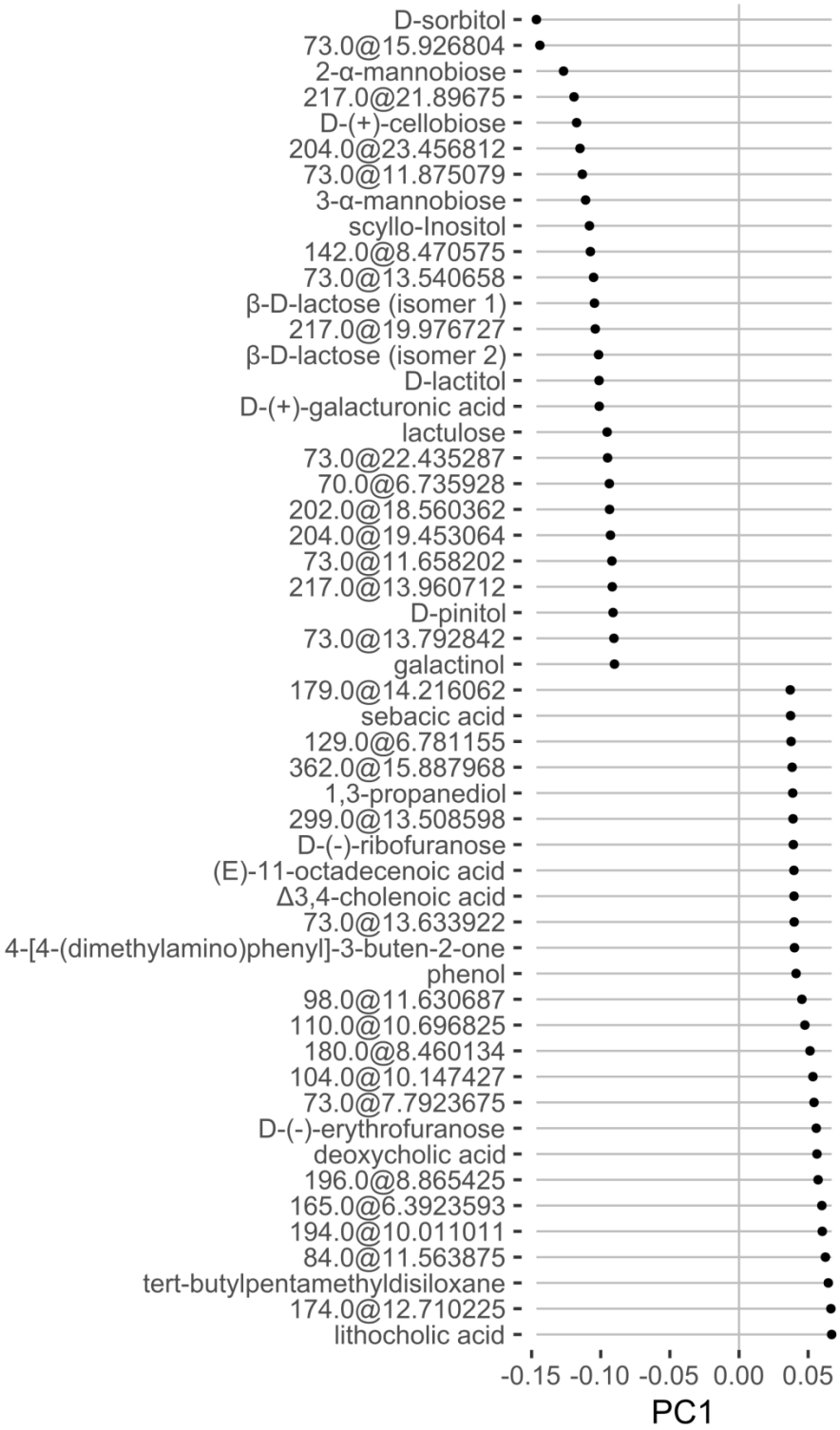
A



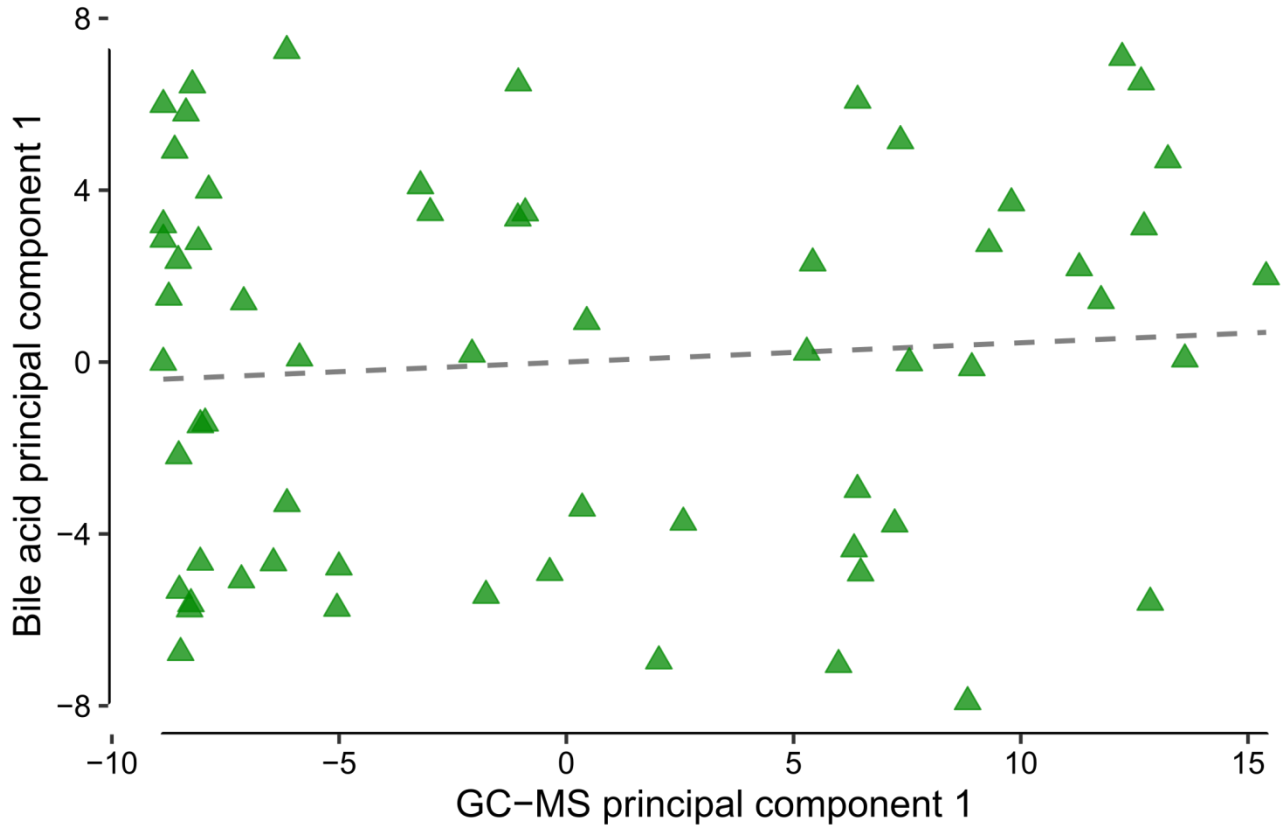
B



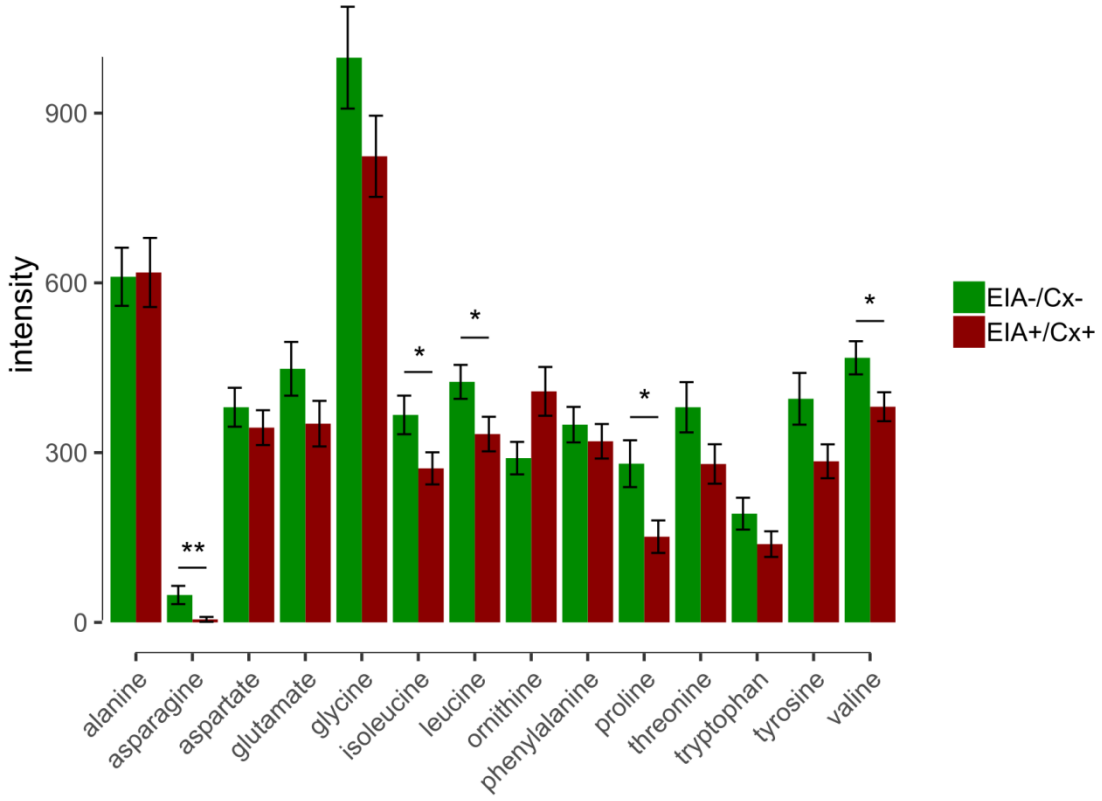
SFig.15. Selected ion monitoring (SIM) GC-MS chromatograms of fecal specimens containing $^{13}C_6$-trehalose internal standard and monitoring for peaks corresponding to trehalose. Within each panel, the top chromatogram is the monitored ion for trehalose at *m/z* 361 and the bottom chromatogram is the ion for $^{13}C_6$-trehalose at *m/z* 367. (A) GC-MS chromatograms from patient fecal specimen in which trehalose was detected. This is indicated by the prominent, co-eluting peaks at 22.8 minutes. This specimen was determined to possess 230 pmol of trehalose. A comparable *m/z* 361 peak at this retention time was observed in the full scan metabolomic profiling data for this specimen. (B) GC-MS chromatograms from patient fecal specimen in which trehalose was undetectable. The *m/z* 361 chromatogram does not possess a co-eluting peak at 22.8 minutes that exceeds the background level.



SFig 16. Putative identities of the upper and lower 1% of features in GC-MS PC1 from Figure 8D. Peak spectra were matched against the NIST 14 spectral library. Features without a database match are labeled as "most intense fragment @ retention time in minutes".



SFig 17. Bile acid PC1 values (Figure 7) plotted against GC-MS PC1 values (Figure 8) for all Cx-/EIA- fecal specimens (n=62). Grey line represents least-squares regression line fitted to all data points (r^2=0.007).



SFig 18. Amino acids detected by GC-MS in patient samples. Bars represent mean peak intensity and errors represent SEM. Significant differences between groups were identified using the Mann-Whitney U-test (** $p<0.01$, * $p<0.05$).

References

1. Robben J, Janssen G, Merckx R, and Eyssen H. Formation of delta 2- and delta 3-cholenoic acids from bile acid 3-sulfates by a human intestinal *Fusobacterium* strain. *Appl Environ Microbiol.* 1989;55(11):2954-9.