Supporting Information

On-site Diagnosis of Poultry Coccidiosis by Miniature Mass Spectrometer and Machine Learning

Wei-Chieh Wang,[†,§] Hsin-Hsiang Chung,[†,§] Ewelina P. Dutkiewicz[†,§], Jhih-Ying Wong[‡], Wen-Chin Yang[‖], Cicero Lee-Tian Chang[*,‡], and Cheng-Chih Hsu[*,†]

[†]Department of Chemistry, National Taiwan University, Taipei 10617, Taiwan

[‡]Department of Veterinary Medicine, National Chung Hsing University, Taichung 40227, Taiwan

[‖]Agricultural Biotechnology Research Center, Academia Sinica, Taipei 11529, Taiwan

* Corresponding author

Email: ccrhsu@ntu.edu.tw (C. -C. Hsu)

Email: ltchang@nchu.edu.tw (C. L. -T. Chang)

Table of Contets

1. Experimental Procedures

Chemicals and materials. Acetonitrile (ACN) and methanol (MeOH) were purchased from J.T. Baker (LCMS-grade, Phillipsburg, NJ, USA). Formic acid (FA) was purchased from Honeywell (98% puriss grade, NJ, USA). Ultrapure water is obtained by PureLab Classic (ELGA, UK). Chemical standards including choline chloride (>99%) , L-methionine (>98%), and genistein (>98%) were purchased from Sigma-Aldrich (Merk, Munich, Germany) for identification verification.

***Eimeria tenella* disease model construction.** The *Eimeria tenella* disease model was constructed at the chicken coop in Department of Veterinary Medicine, National Chung Hsing University, Taiwan. Thirty 4-day-old male chickens were purchased from the Fortune Taiwan company (Taiwan) with the ISO 22000 and IACUC 105-037 certificates. The chickens were randomly divided into infected group (N=15) and healthy group (N=15). The chickens in the infected group were orally infected with 3,000 *Eimeria tenella* oocysts. Chickens in both groups were housed in the similar environment but separated from the other group. Chicken feces from cecal were collected from second to seventh day post-infection (**Table S1**). The weight of each feces sample was ~1.0 g. All the samples were stored under -80 ^{o}C before analysis. To confirm that the orally infection of *E. tenella* was successfully performed, we checked all *E. tenella*-infected chickens by tracking oocyst shedding on day 5 post-infection. All of the *E. tenella*-infected chickens also showed clinical signs, including from none to bloody droppings, paleness, decreased diet intake, and sick bird syndrome with inactivity gradually. In addition, in the end of the experiment, we checked the cecum sections of infected and control chickens using microscopic examination to observe if there is coccidial pathogenic lesions (**Figure S8**).

Chicken feces sample pretreatment. We weighted 100.0 ± 2.0 mg from each chicken fecal sample. The sample was then mixed with 1,000 uL 50% MeOH and homogenized using homogenizer (Precellys® Evolution, Bertin Technologies SAS, Montigny-le-Bretonneux, France) for 90 seconds. Next, samples were centrifuged at 12,000 rpm for 10 minutes. The supernatant (200 μL) was then added with 600 μL cold ACN for protein precipitation. Samples were briefly vortexed, and left on ice for 10 minutes. The precipitate was removed by centrifuging at 12,000 rpm for 10 minutes. Though the high-speed centrifugation (12,000 rpm) was used in the lab experiment, we have demonstrated that mini centrifuge can be used as an alternative to remove precipitate for on-site analysis (**Figure S9**).

Miniature mass spectrometry analysis. We obtained the metabolic profile of chicken fecal using the miniature mass spectrometer (Expression CMS, Advion, NY, USA). The miniature mass spectrometer weighs 32 kg, and has dimensions of 66 × 28 × 56 cm (length × width × height). A single quadrupole with 1.0 Da resolution is adopted as mass analyzer for the miniature mass spectrometer. Samples were introduced into the mass spectrometer using flow injection analysis (FIA). The mobile phase was 50% MeOH with 0.1% FA at the flow rate of 40 μL/min. The injection volume of each sample was 5 μL. Data was acquired under positive-ion mode with the mass range of *m/z* 100-900. The resolution of the miniature mass spectrometer is 1.0 Da. The electrospray ionization (ESI) parameters were set as: capillary temperature: 250 °C; ESI voltage: +4.0 kV. Nitrogen was used as sheath gas for ESI with a flow rate of 3.5 liter per minute. The analysis time for each sample was 4 minutes. The whole diagnosis platform, including miniature mass spectrometer, rotary pump,

gas cylinders and laptop, was carried by a mobile cart with dimensions of 101 × 102 × 50 cm (length × width × height).

Data processing and machine learning model construction. Raw files were converted to CSV format using Advion Data Express (version 3.2.7.1, Advion, NY, USA), and imported into Matlab software for further processing. For each sample, we averaged the MS spectra within the time period of 16 to 60 sec after injection, which represented around 25 spectra. The data was binned to 1.0 Da resolution. To reduce random noise, peaks with intensity lower than 3 times blank samples signal intensity were excluded. All mass spectra were normalized to total ion count.

After preprocessing, data was transferred to the commercially available data mining software RapidMiner (version 9.2.001) for machine learning model construction. The data was randomly separated into training set and validation set with the ratio of 7 to 3 each day (**Table S2**). The training set was used to train the least absolute shrinkage and selection operator (LASSO) model. The LASSO model was constructed by the build-in generalized linear model algorithm with the value of alpha set to 1. The performance of the model was first evaluated using 10-fold cross validation. After training, the model was further evaluated using the validation set.

High-weighted features identification. To identify high-weighted features selected by LASSO model, we analyzed the chicken feces extracts using liquid chromatography high-resolution tandem mass spectrometry (LC-HR-MS/MS). The LC-MS/MS experiments were performed using a hybrid LTQ-Orbitrap Elite mass spectrometer (Thermo Scientific, Waltham, Massachusetts, US) coupled with Vanquish UHPLC Systems (Thermo Scientific, San Jose, USA). A binary gradient with mobile phase A of 0.1% FA and with mobile phase B of ACN and 0.1% FA was used. Liquid chromatography was performed using a 1.7 μm ACQUITY UPLC® CSH C18 column (100 × 2.1 mm, Waters, USA). The flow rate was set to 0.2 mL/min. The gradient started with 2 minutes 5% of phase B, and gradually increased to 99% of phase B during 2-13 minutes. The gradient was then kept at 99% of phase B for 3 minutes. The ESI parameters were set as: heater temperature: 180 °C; capillary temperature: 280 °C; sheath gas flow rate: 30 arb; auxiliary gas flow rate: 5 arb; ESI voltage: +3.5 kV. Data was acquired using FT positive-ion mode with the mass range of *m/z* 100-1,000. The resolution was set at 60,000. The tandem mass spectra of the high-weighted compounds were acquired using collision-induced dissociation (CID) fragmentation. The normalized dissociation energy was set at 25. If multiple peaks were observed in the *m/z* binning range, the peak with the highest intensity was chosen. For the identification of choline, normal phase LC-MS/MS was used. Liquid chromatography was performed using a 1.7 μm ACQUITY UPLC® BEH Amide column (150 × 2.1 mm, Waters, Ireland). The flow rate was set to 0.25 mL/min. The gradient started with 1 minutes 95% of phase B, and gradually decreased to 50% of phase B during 1-5 minutes. The gradient was then kept at 50% of phase B for 3 minutes. The MS parameters were the same as previously described, with the normalized dissociation energy set at 35.

The identification was determined by the exact mass and MS/MS spectrum of the selected feature. The identification results and the confidence level for each feature was listed in **Table S3**. The confidence level of metabolite assignment is according to the guidelines reported by the Chemical Analysis Working Group (CAWG) of the Metabolomics Standards Initiative (MSI).[1] There are four confidence levels for metabolite

identification, where level one represents identified compounds and level four represents unknown compounds. In this paper, we annotate features with level one if the retention time, exact mass and MS/MS spectrum are identical to commercial standards (**Figure S7**). The identification is labeled as level two if the error of exact mass is less than five ppm, and the MS/MS spectrum is interpretable (**Figure S6**) or similar to databases (**Figure S2**). If the identification is only based on exact mass, the confidence level is labeled as three.

Reference

1. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, *et al.* "Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)" *Metabolomics.* **2007**, 3, 211–221.

2. Supporting Tables

Table S1. Number of feces collected at each day for healthy group and infected group.

	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Total
Healthy group	15	15	15	15	15	15	90
Infected group	15	15	15	15	15	15	90

Table S2. Number of samples used in training set and testing set for LASSO model construction.

	Healthy	Infected	Total
Training set	66	66	132
Testing set	24	24	48

Table S3. The weighting and tentatively identification result of each selected *m/z* by the LASSO model. LPC – lysophosphatidylcholine; Leu – leucine; Val – valine; Asp – aspartic acid; Gly – glycine.

Feature	Weighting	Compound	Chemical formula	Species	*m/z* exp.	Mass error (ppm)	Level[a]
395	1.50	Glaucarubolone	$C_{20}H_{26}O_8$	$[M+H]^+$	395.170	0.5	Level 3
150	0.44	Methionine	$C_5H_{11}NO_2S$	$[M+H]^+$	150.058	-3.3	Level 1
436	0.22	Amino-cholestane-3,5,6-triol	$C_{27}H_{49}NO_3$	$[M+H]^+$	436.377	-3.7	Level 3
274	0.18	Deoxyadenosine	$C_{10}H_{13}N_5O_3$	$[M+Na]^+$	274.092	1.6	Level 3
123	0.11	-	$C_2H_6N_2O_4$	$[M+H]^+$	123.040	-0.3	Level 4
583	0.08	Tyr-Asn-Thr-Trp	$C_{28}H_{34}N_6O_8$	$[M+H]^+$	583.253	3.1	Level 3
503	0.08	Malonyldaidzin	$C_{24}H_{22}O_{12}$	$[M+H]^+$	503.117	-3.4	Level 2
519	0.07	Malonylgenistin	$C_{24}H_{22}O_{13}$	$[M+H]^+$	519.112	-3.3	Level 2
520	0.06	LPC(18:2)	$C_{26}H_{50}NO_7P$	$[M+H]^+$	520.338	-3.3	Level 2
510	0.03	LPC(17:0)	$C_{25}H_{52}NO_7P$	$[M+H]^+$	510.354	-3.4	Level 2
420	0.01	-	$C_{24}H_{29}N_5O_2$	$[M+H]^+$	420.240	2.4	Level 4
121	-0.81	Phenylacetaldehyde	C_8H_8O	$[M+H]^+$	121.065	-3.3	Level 2
271	-0.49	Genistein	$C_{15}H_{10}O_5$	$[M+H]^+$	271.059	-3.3	Level 1
104	-0.38	Choline	$C_5H_{14}NO$	M^+	104.107	-3.8	Level 1
245	-0.24	Leu-Leu	$C_{12}H_{24}N_2O_3$	$[M+H]^+$	245.185	-2.8	Level 2
527	-0.20	-	$C_{28}H_{34}N_2O_8$	$[M+H]^+$	527.239	0.6	Level 4
457	-0.18	Leu-Leu-Leu-Val	$C_{23}H_{44}N_4O_5$	$[M+H]^+$	457.337	-2.9	Level 2
586	-0.18	-	$C_{27}H_{47}N_5O_7S$	$[M+H]^+$	586.327	-0.7	Level 4
134	-0.11	Aspartic acid	$C_4H_7NO_4$	$[M+H]^+$	134.044	-7.7	Level 2
191	-0.09	Asp-Gly	$C_6H_{10}N_2O_5$	$[M+H]^+$	191.066	-0.6	Level 2
272	-0.04	N-(3-Sulfopropyl)-methionine	$C_8H_{17}NO_5S_2$	$[M+H]^+$	272.063	2.6	Level 3
317	-0.01	2-hydroxy-octadecatrienoic acid	$C_{18}H_{30}O_3$	$[M+Na]^+$	317.208	-3.5	Level 3
285	-0.01	Glycetein	$C_{16}H_{12}O_5$	$[M+H]^+$	285.076	-2.3	Level 2

[a] The confidence level of metabolite assignment is according to the guidelines reported by the Chemical Analysis Working Group (CAWG) of the Metabolomics Standards Initiative (MSI).[1]

1. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, *et al.* "Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI)" *Metabolomics.* **2007**, 3, 211–221.

3. Supporting Figures

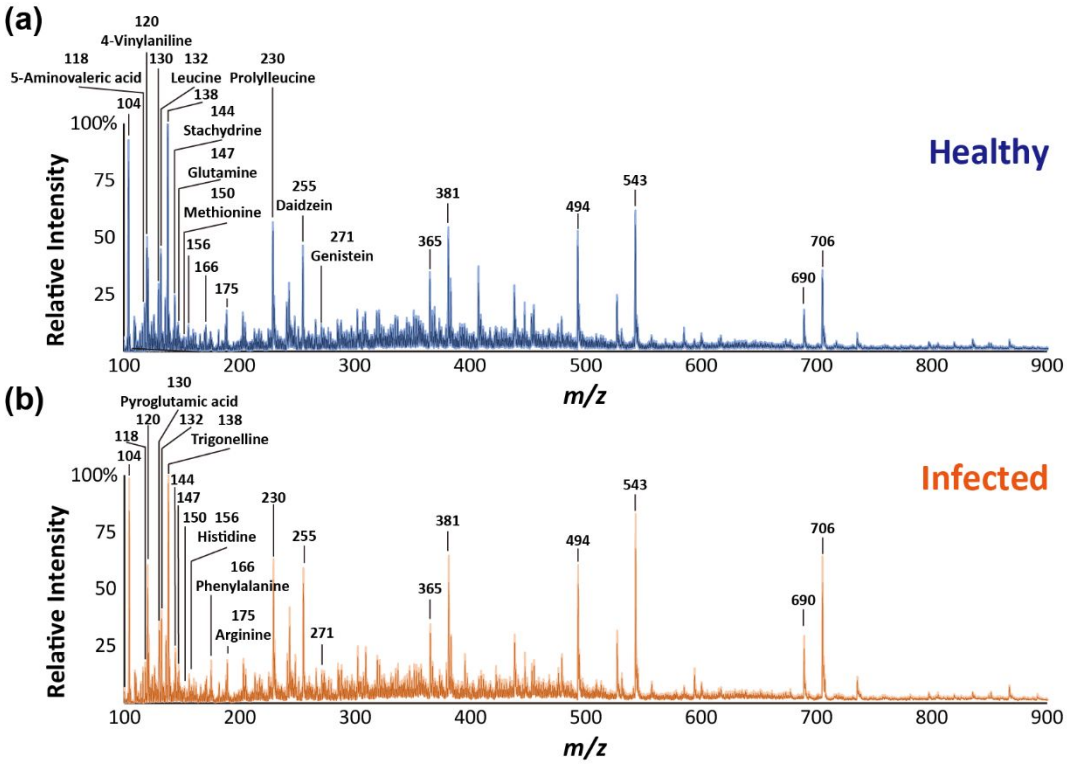


Figure S1. Averaged mass spectra of healthy **(a)** and infected **(b)** chicken fecal extracts acquired using miniature mass spectrometer (Expression CMS, Advion, NY, USA). The data is presented as mean values (line) ± interquartile range (shadow). The metabolites labeled were identified by subsequent high-resolution tandem mass spectrometry analysis (**Figure S2**).

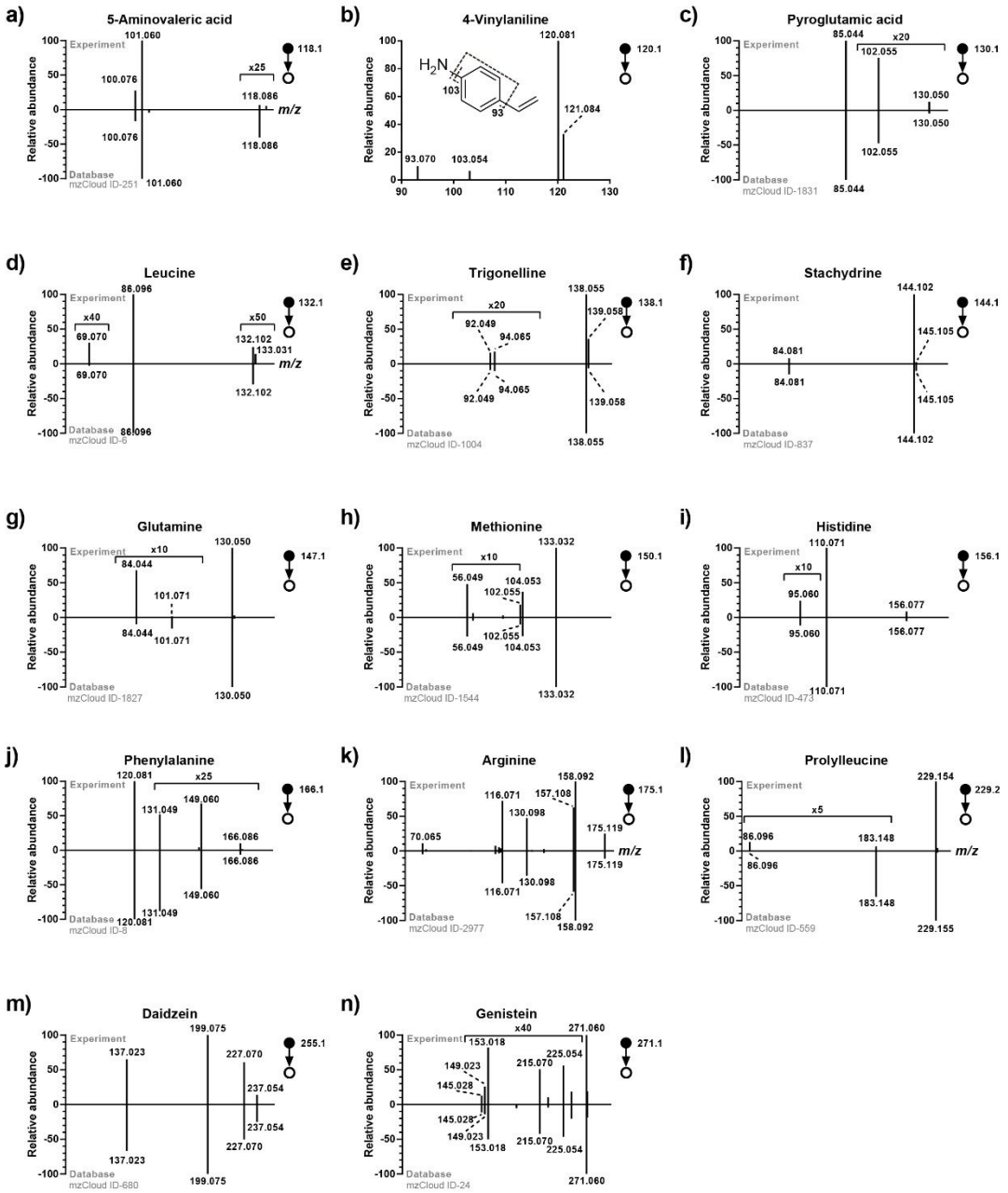


Figure S2. The MS/MS spectra of 14 metabolites observed in chicken feces extracts: (**a**) 5-Aminovaleric acid, (**b**) 4-Vinylaniline, (**c**) Pyroglutamic acid, (**d**) Leucine, (**e**) Trigonelline, (**f**) Stachydrine, (**g**) Glutamine, (**h**) Methionine, (**i**) Histidine, (**j**) Phenylalanine, (**k**) Arginine, (**l**) Prolylleucine, (**m**) Daidzein, and (**n**) Genistein. The identification was determined by the exact mass (mass error < 5 ppm) and MS/MS spectrum. The face-up shows the experimental MS/MS spectrum collected by high-resolution mass spectrometer (LTQ-Orbitrap Elite mass spectrometer, Thermo Scientific, US), and the face-down shows the spectrum from mzCloud database (the mzCloud ID for each compound is provided).

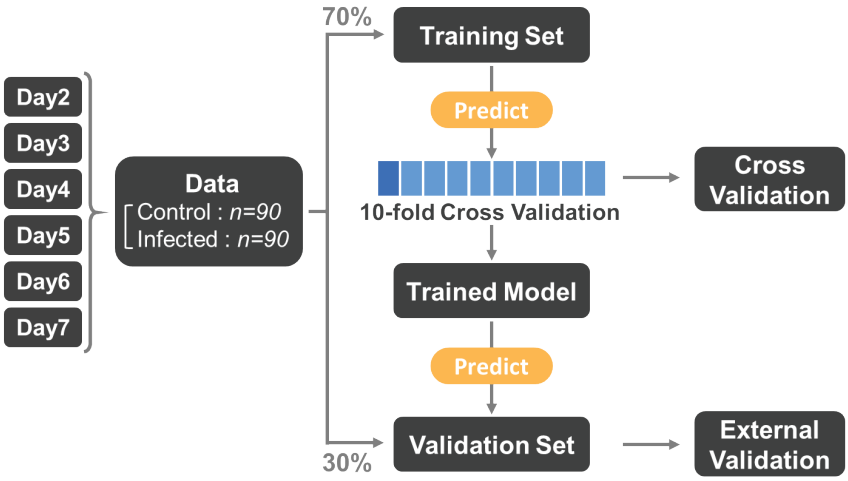


Figure S3. Workflow for constructing Lasso model for coccidiosis diagnosis.

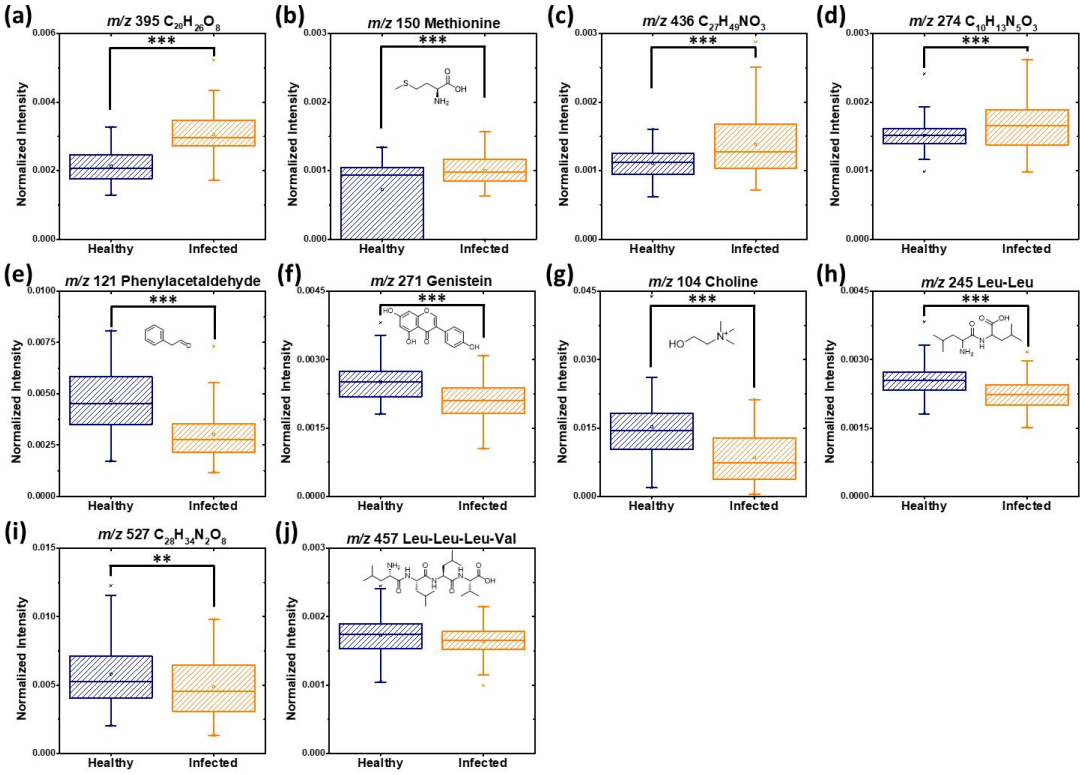


Figure S4. The relative abundance of top-10 high weighted *m/z* in healthy group and infected group. Statistical analysis (*t* test) results are shown as $* \ p < 0.05$; $** \ p < 0.01$; $*** \ p < 0.001$.

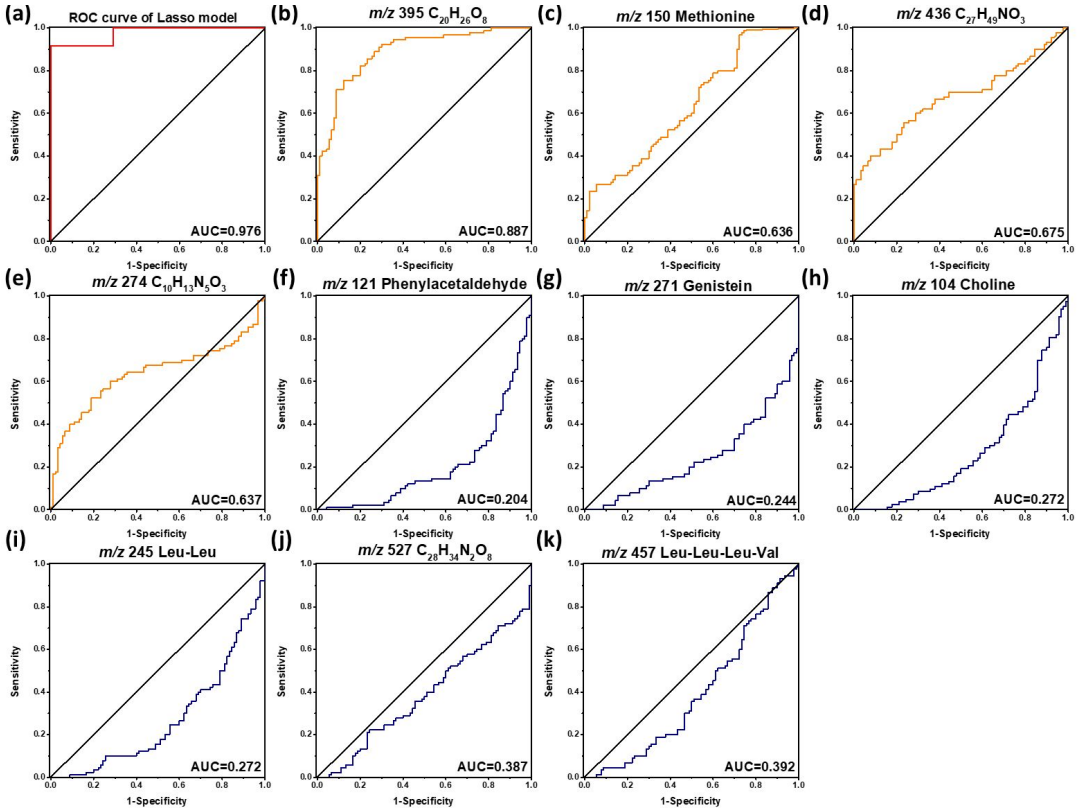
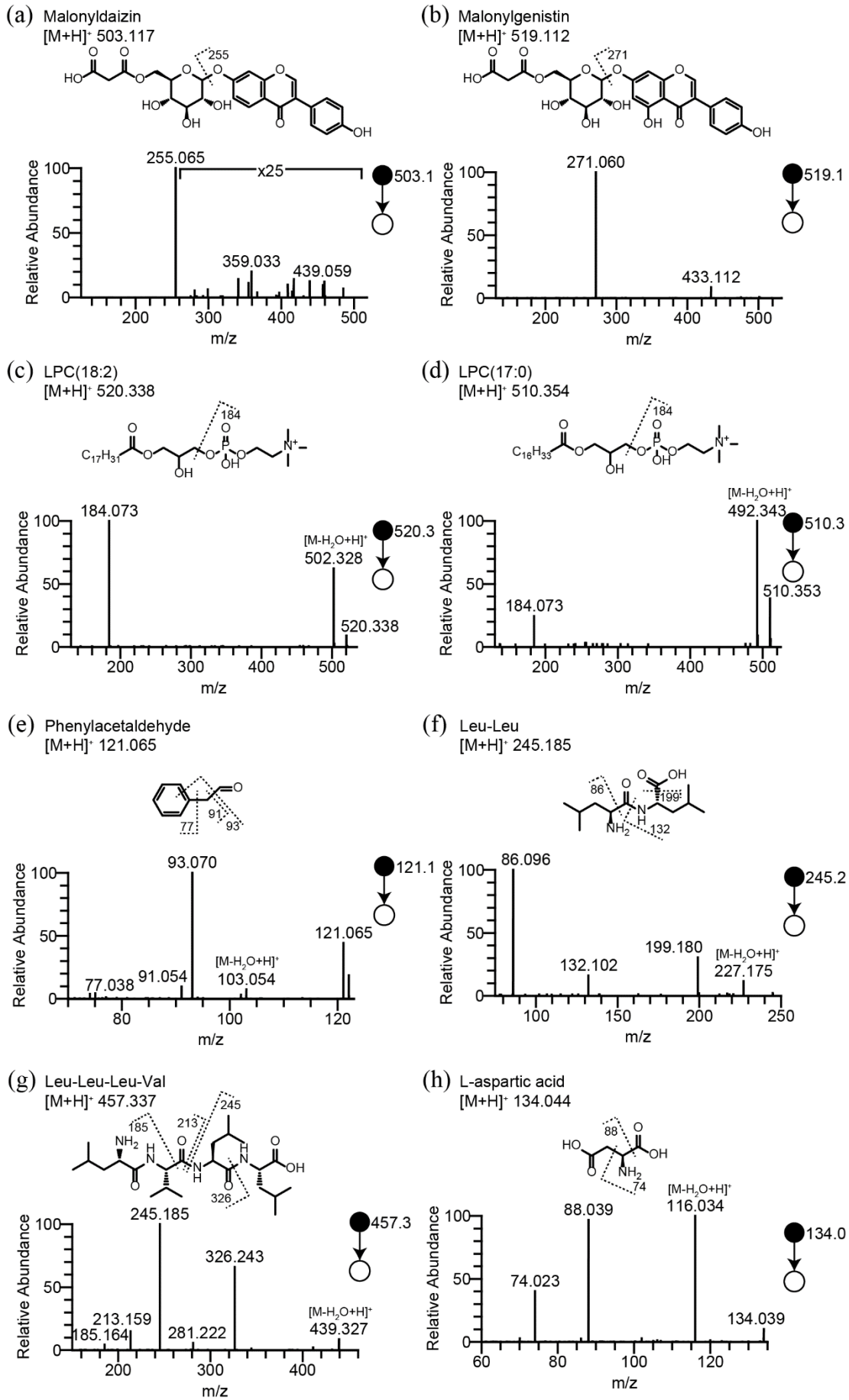


Figure S5. Receiver operating characteristic (ROC) curve of Lasso model (**a**) and top-10 high weighted *m/z* (**b-k**).


(a) Malonyldaizin
[M+H]+ 503.117
255
255.065
x25
503.1
359.033
439.059
Relative Abundance
m/z
(b) Malonylgenistin
[M+H]+ 519.112
271
271.060
519.1
433.112
(c) LPC(18:2)
[M+H]+ 520.338
184
184.073
[M-H2O+H]+
502.328
520.3
520.338
(d) LPC(17:0)
[M+H]+ 510.354
184
[M-H2O+H]+
492.343
510.3
510.353
184.073
(e) Phenylacetaldehyde
[M+H]+ 121.065
91
77
93
93.070
121.1
121.065
[M-H2O+H]+
103.054
91.054
77.038
(f) Leu-Leu
[M+H]+ 245.185
86
199
132
86.096
245.2
199.180
[M-H2O+H]+
227.175
132.102
(g) Leu-Leu-Leu-Val
[M+H]+ 457.337
185
213
245
326
245.185
457.3
326.243
213.159
185.164
281.222
[M-H2O+H]+
439.327
(h) L-aspartic acid
[M+H]+ 134.044
88
74
[M-H2O+H]+
116.034
88.039
134.0
74.023
134.039

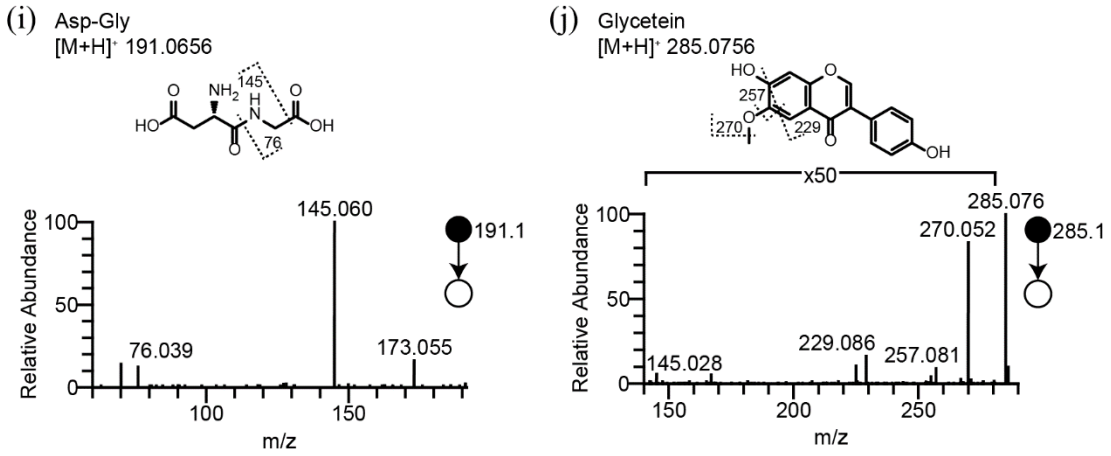


Figure S6. MS/MS spectra of diagnostic features selected by the Lasso model. Proposed molecular structures and fragmentation sites are shown for each species: (**a**) malonyldaidzin, (**b**) malonylgenistin, (**c**) LPC(18:2), (**d**) LPC(17:0), (**e**) phenylacetaldehyde, (**f**) Leu-Leu, (**g**) Leu-Leu-Leu-Val, (**h**) aspartic acid, (**i**) Asp-Gly, (**j**) glycetein.

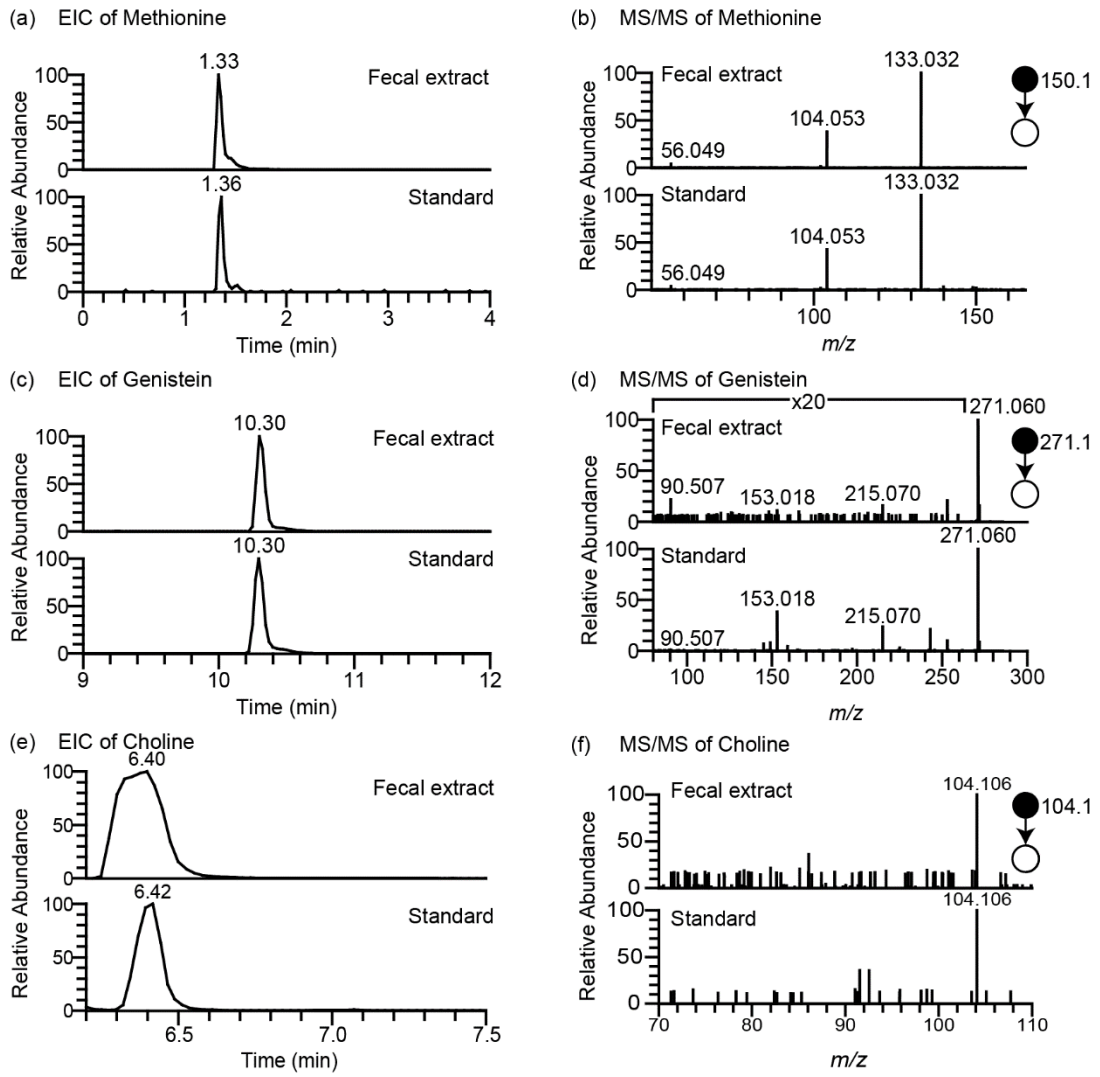


Figure S7. The extracted ion chromatogram (EIC) and MS/MS spectra of methionine (**a**, **b**), genistein (**c**, **d**), and choline (**e**, **f**) acquired from standards and fecal extract.

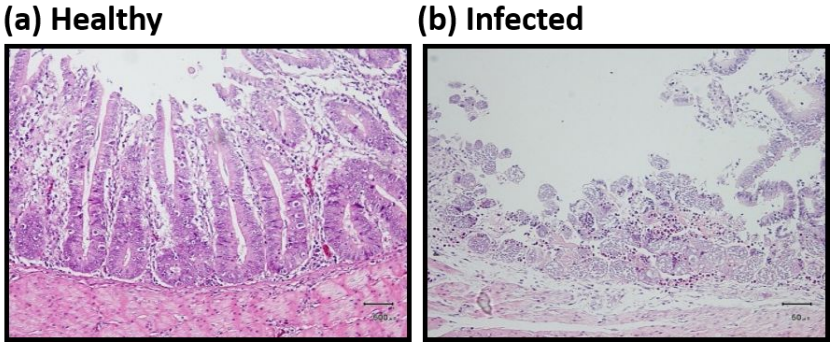


Figure S8. Histopathological lesions in cecum of chicken in healthy chicken (**a**) and *E. tenella*-infected chicken (**b**). Cecum of chickens in unchallenged control group did not show obvious pathological changes. For cecum tissues in *E. tenella* challenged group, the gut villi were broken, and the structure was vague. Red blood cells and inflamed cells were infiltrated in cecum tissues.

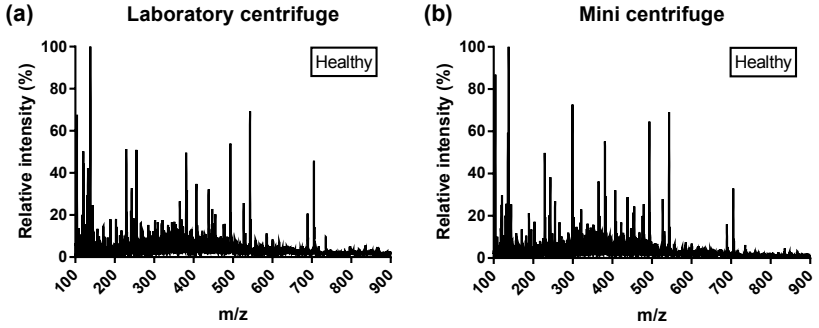


Figure S9. MS spectra of chicken fecal extracts prepared by laboratory centrifuge (12,000 rpm) (**a**), and mini centrifuge (**b**). The two spectra show high similarity (cosine similarity: 0.91), and are both predicted correctly as healthy.