Metaproteomics Analysis Reveals the Adaptation Process for the Chicken Gut Microbiota

Yue Tang,^a Anthony Underwood,^b Adriana Gielbert,^a Martin J. Woodward,^c Liljana Petrovska^a

Veterinary Laboratories Agency, Addlestone, Surrey, United Kingdom^a; Public Health England, London, United Kingdom^b; Food and Nutritional Sciences, University of Reading, Whiteknights, Reading, United Kingdom^c

The animal gastrointestinal tract houses a large microbial community, the gut microbiota, that confers many benefits to its host, such as protection from pathogens and provision of essential metabolites. Metagenomic approaches have defined the chicken fecal microbiota in other studies, but here, we wished to assess the correlation between the metagenome and the bacterial proteome in order to better understand the healthy chicken gut microbiota. Here, we performed high-throughput sequencing of 16S rRNA gene amplicons and metaproteomics analysis of fecal samples to determine microbial gut composition and protein expression. 16 rRNA gene sequencing analysis identified *Clostridiales*, *Bacteroidaceae*, and *Lactobacillaceae* species as the most abundant species in the gut. For metaproteomics analysis, peptides were generated by using the Fasp method and subsequently fractionated by strong anion exchanges. Metaproteomics analysis identified 3,673 proteins. Among the most frequently identified proteins, 380 proteins belonged to *Lactobacillus* spp., 155 belonged to *Clostridium* spp., and 66 belonged to *Streptococcus* spp. The most frequently identified proteins were heat shock chaperones, including 349 GroEL proteins, from many bacterial species, whereas the most abundant enzymes were pyruvate kinases, as judged by the number of peptides identified per protein (spectral counting). Gene ontology and KEGG pathway analyses revealed the functions and locations of the identified proteins. The findings of both metaproteomics and 16S rRNA sequencing analyses are discussed.

The domestic poultry industry is an important livestock sector producing meat and eggs for human consumption, the vast majority of which are produced by large commercial enterprises in intensive industrial systems (1).

The gut microbiota carries out vital processes for normal digestive functions of the host, and it can be considered an acquired organ comprising very large numbers of diverse bacterial cells that can perform differing functions (2, 3), perhaps the most commercially important being its contribution to feed conversion. The mature microbiota is highly diverse, with over 1,000 bacterial species in chickens (4). Colonization of the chicken gastrointestinal (GI) tract by commensal bacteria is an ongoing process which begins immediately after hatching, and the microbiota of the small intestine is established by week 2 (5–8). Microbes in the GI tract may be grouped into either commensal organisms or transient and potential pathogens. The commensals are adapted to the host environment and are often considered beneficial by providing vitamins, amino acids, and short-chain fatty acids to the host: acetate, butyrate, and succinate are commonly produced, with butyrate being the preferred energy source for host epithelial cells (9). The normal microbiota also mitigates against pathogens by mechanisms that are not yet fully understood.

The gut microbiota is composed of a large number of bacteria comprising very diverse species that are presumed to both compete and cooperate synergistically for the use and catabolism of nutrient resources (10). We hypothesized that while the high-throughput genomic approach is an appropriate tool for assessment of the state of the gut microbiota, such as bacterial diversity, proteomics analyses would provide a fingerprint of metabolic and stress responses of the bacteria to the specific environment of the chicken gut, where nutrient limitation and higher body temperatures, 40°C to 43°C, provide stress. As a preliminary step in making this assessment, we wished to provide proof of principle by performing protein expression proteomics and 16S rRNA se-

quencing analyses and using bioinformatics analysis to identify major protein functions and pathways. These analyses provided starting points to evaluate the chicken gut microbiota and potentially to develop biomarkers for infections.

MATERIALS AND METHODS

Bacterial extraction. Two pooled fecal samples from 30 18-week-old white leghorn chickens fed Attlee's nonmedicated poultry feed (Lillico Attlee, Dorking, Surrey, United Kingdom) and housed in biosecure rooms in groups of 15 were collected for this study. Bacterial cells were extracted from freshly collected samples by using the differential centrifugation method (11). Briefly, fresh chicken fecal samples were collected from the two rooms in the early afternoon (2 to 3 p.m.). To obtain sufficient cells for protein extraction, the samples were pooled and dispensed into three tubes, each containing 5 g. The 5-g pooled fecal samples were resuspended by vortexing in phosphate-buffered saline (PBS) (138 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 2.8 mM KH₂PO₄ [pH 7.2]) containing 0.1% Tween 80 in 25-ml Universal tubes. The samples were spun at 300 × g for 10 min at 8°C, and the supernatants were collected. Fresh PBS containing 0.1% Tween 80 was added to the sample tubes, and the tubes were vortexed to resuspend the pellets. This cycle was repeated 4 times to gather bacterial cells. The cells in the supernatants were pelleted by centrifugation at 14,000 × g for 20 min at 4°C. The recovered cells were washed three times in 50 ml PBS plus 0.1% Tween 80 through resuspension and centrifugation and stored at –20°C.

Received 4 September 2013 Accepted 16 October 2013

Published ahead of print 8 November 2013

Address correspondence to Yue Tang, yue.tang@ahvla.gsi.gov.uk, or Liljana Petrovska, Liljana.petrovska@ahvla.gsi.gov.uk.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.02472-13>.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.
doi:10.1128/AEM.02472-13

16S rRNA gene sequencing analysis. Portions of bacterial cells purified from the chicken fecal samples used for proteomics were gathered for DNA isolation. Approximately 200 mg of each fecal sample was placed directly into a Maxwell 16 Tissue DNA Purification kit sample cartridge (Promega, USA). Genomic DNA was purified by using the Maxwell 16 Instrument (Promega, USA), as recommended by manufacturer, which includes mechanical disruption. Eluted DNA was stored at -20°C until future use. DNA yield and purity were determined by using a NanoDrop spectrophotometer.

Aliquots of extracted DNA were amplified with universal primers for the V4 and V5 regions of the 16S rRNA gene. Primers U515F (5'-GTGYC AGCMGCCGCGGTA) and U927R (5'-CCCGYCAATTCMTTTRAGT) were designed to permit amplification of both bacterial and archaeal ribosomal gene regions while providing the best possible taxonomic resolution based on previously reported information (12, 13). Forward fusion primers consisted of GS FLX Titanium primer A and the library key 5'-C CATCTCATCCCTGCGTGTCTCCGACTCAG, together with one of a suite of 16 10-bp multiplex identifiers (MIDs) (Roche Diagnostics, United Kingdom). Reverse fusion primers included GS FLX Titanium primer B and the library key 5'-CCTATCCCCTGTGTGCCTTGGCAGTCTCAG. Amplification was performed with FastStart HiFi polymerase (Roche Diagnostics Ltd., United Kingdom) by using the following cycling conditions: 94°C for 3 min and 30 cycles of 94°C for 30 s, 55°C for 45 s, and 72°C for 1 min, followed by 72°C for 8 min. Amplicons were purified by using Ampure XP magnetic beads (Beckman Coulter, United Kingdom), and the concentration of each sample was measured by using the fluorescence-based Picogreen assay (Invitrogen, USA). Concentrations were normalized before pooling samples in batches of 8, each of which would be subsequently identified by its unique MID. Pooled samples were then subjected to unidirectional sequencing from the forward primer on the 454 GS FLX Titanium platform according to the manufacturer's instructions (Roche Diagnostics).

The bar-coded pyrosequencing data were analyzed by using the QIIME database (<http://www.microbio.me/qiime/>) and QIIME version 1.5.0-dev (14). The QIIME database processing pipeline allows correction for possible biases arising from different depths of sequence across samples by splitting the raw sequence data into samples by bar codes and filtering the low-quality reads using QIIME database default parameters. Each library was subsampled to an even sequencing depth of a minimum of 500 sequences per sample and clustered into operational taxonomic units (OTUs) by using a closed-reference OTU-picking protocol at 97% sequencing identity (similar to the species level) using UCLUST (15).

A representative sequence for each OTU was chosen for downstream analysis based on the most abundant sequence from each OTU. PyNAST was used to align sequences with a minimum length of 150 bp and a minimum identity of 75.0% (16). OTUs were assigned to a taxonomy by using Ribosomal Database Project (RDP) Naive Bayes classifier v 2.2 with the confidence level set at 0.8 (17). β diversity (diversity between groups of samples) was used to generate principal coordinate plots for each sample using unweighted and weighted UniFrac distances (18).

Protein extraction. Bacterial cells were disrupted by sonication (3 times for 10 s) (19), and cell extracts were recovered by centrifugation at $14,000 \times g$ for 10 min at 4°C . Protein contents were determined by using Coomassie Plus protein assay reagent (Thermo Scientific) with bovine serum albumin (BSA) as the standard, according to the manufacturer's instructions.

Proteolytic digestion with trypsin. Digestion was carried out according to the instructions provided with the Fasp protein digestion kit, which is based on filtering out undesirable substances during sample preparation (Protein Discovery, USA). In brief, 0.4 mg of total protein resuspended in 30 μl UPX Universal protein extraction buffer was heated at 99°C for 10 min. Two hundred microliters of urea sample solution was mixed with protein samples in spin filters, and the filters were spun at $14,000 \times g$ for 15 min. The flowthrough was discarded. Ten microliters of $10\times$ iodoacetamide solution and 90 μl urea sample solution were added

to the filters and mixed by vortexing for 1 min. The filters were kept in the dark for 20 min at room temperature (RT) before being centrifuged at $14,000 \times g$ for 10 min. The spin filters were centrifuged again for 15 min after the addition of 100 μl urea sample solution, and this step was repeated twice. The flowthrough was discarded. Ammonium bicarbonate solution (100 μl) was added to the filters, and the filters were centrifuged at $14,000 \times g$ for 10 min; this step was also repeated twice. Trypsin (5 μg ; Promega, United Kingdom) in 100 μl ammonium bicarbonate solution was added to the filter. The mixture was vortexed and incubated at 37°C for 18 h. The filter was transferred into a fresh tube before peptides were collected at $14,000 \times g$ for 10 min. The collection of peptides was repeated after the addition of 100 μl ammonium bicarbonate solution to the filter. Finally, 100 μl 0.5 M NaCl was added, and peptides were collected again. The peptide concentration was measured by using a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA) at 280 nm. The peptides were desalted by using C_{18} columns before being analyzed by quadrupole time of flight (QTOF) mass spectrometry or being fractionated by strong anion exchange (SAX).

Peptide fractionation by strong anion exchanges. SAX was performed as described previously (20). Briefly, 30 to 50 μg tryptic peptide was loaded at pH 11 onto a tip-based anion exchanger constructed by using six layers of an Empore anion exchange disk (3M, Bracknell, United Kingdom). The column was equilibrated, and fractions were eluted by using Britton-Robinson buffer (20 mM acetic acid, 20 mM phosphoric acid, and 20 mM boric acid titrated with NaOH to the desired pH). Fractions were eluted subsequently with buffer solutions of pH 11, 8, 6, 5, 4, and 3. The fractionated peptides were desalted by using C_{18} columns.

C_{18} desalting. Two types of cartridges were used for desalting: Sep Vac 3cc (200 mg) C_{18} cartridges (Waters, United Kingdom) for peptides after the Fasp method was performed and Empore C_{18} 3-ml solid-phase-extraction cartridges (3M, Bracknell, United Kingdom) for peptides after SAX fractionation was performed. Two buffers were used for both cartridges. Buffer A contained 98% H_2O , 2% acetonitrile, and 0.1% formic acid, and buffer B contained 80% acetonitrile, 20% H_2O , and 0.1% formic acid. The wash and elution steps were carried out in accordance with the manufacturer's instructions. The eluted peptides were dried with a Speed Vac. The peptides were then resuspended in buffer A, and the concentration was measured with a NanoDrop 1000 spectrophotometer at an A_{280} before use.

LC-MS/MS analysis. The peptide samples were analyzed by using a 1200 series nanoscale liquid chromatography (NanoLC) instrument connected online to a 6520 QTOF mass spectrometer (mass accuracies better than 2 ppm for mass spectrometry [MS] and 5 ppm for tandem MS [MS/MS] and mass resolution of up to 20,000) via an orthogonal electrospray Chipcube interface (Agilent Technologies, United Kingdom). Peptide samples (1 μg in 0.1% formic acid) were injected for LC-MS/MS analysis with a 120-min gradient from 3% (vol/vol) acetonitrile–0.1% (vol/vol) formic acid to 40% (vol/vol) acetonitrile–0.1% formic acid and from 3% (vol/vol) acetonitrile–0.1% (vol/vol) formic acid to 75% (vol/vol) acetonitrile–0.1% formic acid by using a high-capacity (160-nl) 150-mm C_{18} chip (Agilent Technologies, United Kingdom). The mass spectrometer was operated with mass ranges of 250 to 2,000 m/z for MS and 50 to 2,000 m/z for MS/MS at a constant flow rate of 500 nl/min. Acquisition rates were 3 spectra/s for MS and 5.01 spectra/s for MS/MS, and the mass reference was 922.009798 m/z .

Spectrum identification and gene ontology analysis. Mascot (version 2.3; Matrix, United Kingdom) analysis was carried out to identify peptides and to search for proteins in the NCBI nonredundant (nr) database (release date, 28 March 2011). The following settings for a database search from the QTOF data were employed: mass tolerances of 0.3 Da for MS spectra and 0.3 Da for MS/MS spectra. The target decoy search was selected to obtain false discovery rates (FDRs). The significance threshold, P , was adjusted so that the false discovery rate was $<5\%$ when a peptide matched above the homology or identity threshold. Gene ontology (GO)

TABLE 1 Bacterial families and order identified by 16S rRNA sequencing

Bacterial family or order ^a	No. of sequences	% of total sequences
<i>Clostridiales</i>	2,422	25
<i>Bacteroidaceae</i>	2,206	21
<i>Lactobacillaceae</i>	2,011	19
<i>Enterococcaceae</i>	551	5
<i>Porphyromonadaceae</i>	1,383	13
<i>Eubacteriaceae</i>	487	5
<i>Ruminococcaceae</i>	314	3
<i>Lachnospiraceae</i>	205	2
<i>Veillonellaceae</i>	195	2
<i>Rikenellaceae</i>	140	1

^a Due to the resolution of 16S rRNA gene sequencing, *Clostridiales* is the order name.

and pathway analyses of the identified proteins were performed with Protein Information Resource (iProXpress) bioinformatics tools.

RESULTS

16S rRNA gene sequencing analysis. To gain information on the bacterial population in the healthy chicken gut using freshly voided fecal samples as a surrogate, total DNA extracted from the pooled sample was analyzed. Bacterial taxon detection, with the most abundant phyla, class, and genera, is summarized in Fig. S1 in the supplemental material. The bacterial community in the healthy chicken gut was dominated by members of the phyla *Firmicutes* and *Bacteroidetes*, at 55.99 and 35.5%, respectively, with lower percentages of *Proteobacteria* at 1.7%, *Actinobacteria* at 0.1%, and all others at 6.7%. The most frequently identified bacterial families are listed in Table 1. The most frequently identified bacterial taxa were *Clostridiales* (2,422 sequences), *Bacteroidaceae* (2,206), and *Lactobacillaceae* (2,011).

Metaproteomics analysis of chicken fecal samples. Chicken gut microbial metaproteomes were obtained from the same pooled fresh fecal samples used for 16S rRNA sequencing. Peptides identified by both the Fasp method and fractionation by strong anion exchange were analyzed by LC-MS/MS. A flow diagram of the peptide fractionation and analysis is presented in Fig. S2 in the supplemental material. To maximize protein identification, the peptides identified by SAX were analyzed with two gradient elution profiles: 75% buffer B and 40% buffer B in 120 min. Spectrum identification and subsequent protein identification were performed by using the Mascot search engine (Matrix, United Kingdom) and the NCBI nr database. To ensure high confidence, the false discovery rate (FDR) was set at <5% (21–24). The details of the Fasp runs, the percent FDR, and the number of identified proteins in each step are presented in Table S1 in the supplemental material. The list of proteins identified by using Mascot is presented in Table S2 in the supplemental material, and the list of proteins mapped to UniProt identifications is presented in Table S3 in the supplemental material.

A total of 3,673 proteins were identified in this study (see Table S2 in the supplemental material). Although the peptides were further fractionated by SAX, there were still proteins identified by the Fasp method alone. Of 812 proteins identified by the Fasp method, 436 proteins were identified by both methods (Fig. 1). To remove redundant proteins from the NCBI nr database, the identified proteins were mapped to UniProtKB accession number identifications from NCBI gi numbers (25).

FIG 1 Venn diagram of the number of shared proteins determined by both the Fasp method and the Fasp SAX method.

A total of 3,487 proteins were mapped to UniProtKB identifications, representing 799 genera (see Table S3 in the supplemental material), of which 11% (380) belonged to *Lactobacillus* spp. and 4.4% (155) belonged to *Clostridium* spp. (Table 2).

The abundance of a protein is usually correlated with peptides identified in a given sample. Spectral counting is a popular method for quantitative proteomics, as the more peptides identified, the more abundant a protein in the sample (26). It is interesting to note that although pyruvate kinases were the most abundant proteins as determined by peptide counting, including 27 peptides from *Lactobacillus crispatus* JV-V01 pyruvate kinase (see Table S2 in the supplemental material), they were not the most frequently identified proteins in this study (Table 3). Other abundant bacterial proteins were phosphoglycerate kinase and elongation factors (see Table S2 in the supplemental material).

When MS/MS spectra are searched against a sequence database, peptides, rather than proteins, are matched. Each protein is identified with at least one unique peptide. In many cases, the matched peptides are not unique to a single protein. Given a set of peptides, sometimes several proteins can be identified, although only one protein is present on the protein list. These peptides are also called semiunique peptides. This is particularly problematic for metaproteomics, as many bacterial species in the gut are closely related. For example, there were 11 peptides matching glyceraldehyde-3-phosphate dehydrogenase of *Lactobacillus johnsonii* NCC 533; the same set of peptides matched glyceraldehyde-3-phosphate dehydrogenase type I of *Lactobacillus gasseri* 202-4. This matching information is detailed for identified proteins in Mascot searches. To inspect our data set for the extents of this occurrence and as a quality control, we performed two checks with

TABLE 2 Numbers of proteins from bacteria and eukaryotes identified by proteomics

Bacterial genus	No. of proteins	% of total proteins identified
<i>Lactobacillus</i>	380	11
<i>Clostridium</i>	155	4
<i>Streptococcus</i>	66	2
<i>Bacteroides</i>	65	2
<i>Bacillus</i>	54	2
<i>Plasmodium</i>	52	1
<i>Prevotella</i>	44	1
<i>Eubacterium</i>	41	1
<i>Ruminococcus</i>	39	1
<i>Escherichia</i>	34	1
<i>Enterococcus</i>	34	1
<i>Tetrahymena</i>	33	1
<i>Turicibacter</i>	24	1
<i>Vibrio</i>	23	1
<i>Staphylococcus</i>	20	1

TABLE 3 Most frequently identified proteins

Protein	No. of identifications	Reference(s) ^a
60-kDa chaperone (GroEL)	349	16, 27–30
Glyceraldehyde-3-phosphate dehydrogenase	279	31–33
Glutamate dehydrogenase	200	34
Elongation factor G	155	30, 35, 36
DNA-directed RNA polymerase subunit beta'	118	
Phosphoglycerate kinase	87	36
Elongation factor Tu	64	30, 37, 38
Enolase (2-phosphoglycerate dehydratase)	62	31
50S ribosomal protein L7/L12	57	30
Pyruvate kinase	44	36
Glucose-6-phosphate isomerase	43	
DNA-directed RNA polymerase	42	
Formate-tetrahydrofolate ligase	36	
L-Lactate dehydrogenase	23	33
Chaperone protein DnaK	18	16, 27–30
Fructose-bisphosphate aldolase	18	36
Triosephosphate isomerase	17	36

^a References for proteins with adaptation-related functions.

100 randomly identified proteins each time; we identified 32 and 39 occasions where the same set of peptides matched more than one protein. Furthermore, of 100 proteins identified with more than one peptide, 61 belonged to the same bacterial species, 27 belonged to the same genus, and 12 belonged to the same family/order. Therefore, the majority of matching proteins belonged to the same species but were identified with different protein identifications.

Bacterial adaptation in the chicken gut. The most frequently identified proteins in this study were chaperone proteins, including 349 GroEL and 18 DnaK proteins (Table 3). Proteases involved in the heat shock response, such as 1 Clp protease and 2 FtsH proteases (39), were also identified (see Table S3 in the supplemental material). Other stress proteins included 5 cold shock proteins, 7 cytochromes, 4 superoxide dismutases, 2 thioredoxins, and 4 peroxidases (see Table S3 in the supplemental material), which may suggest that to live in the chicken gut, bacteria need to adapt to the environment. Also, many of the frequently identified proteins (Table 3) may be involved in coping with stress conditions, although, by definition, their normal functions are not in the stress response. For example, elongation factor G (EF-G) and EF-Tu are part of the protein translation machinery in *Escherichia coli*, but they have also been shown to possess chaperone properties (35, 37). Other identified stress-related proteins included 30S ribosomal protein S1, 30S ribosomal protein S3, 50S ribosomal protein L7/L12, EF-G, polynucleotide phosphorylase, DnaK and GroEL enolases, polynucleotide phosphorylase, and pyruvate-flavodoxin oxidoreductase (Table 3), many of which have been shown to be upregulated in different pure bacterial cultures under acid stress, metabolic stress, oxidative stress, and other stress conditions.

Other proteins. There were 146 host proteins identified in this study, of which 47 were mapped to UniProt identifications, including two mucins (see Table S2 in the supplemental material). Secretory IgA (SIgA) produced in mucosal linings serves as the

first line of defense against microorganisms through a mechanism called immune exclusion (40). There were 69 immunoglobulin proteins which were not mapped to UniProt identifications. Close inspection revealed that they were unique proteins which differed in the variable regions, as judged by sequencing alignment. These proteins can be divided into three groups: the light chains, the heavy chains, and the joint chains. The difference in protein identifications is likely due to the variable regions from both light and heavy chains.

There were 19 glycine (soybean) proteins, 11 rice (*Oryza sativa*) proteins, 4 maize (*Zea*) proteins, as well as 1 onion protein identified in this study. According to the ingredients reported by Lillico Attlee, the feed contains soybeans, wheat, and maize.

Gene ontology and pathway analysis. Although many of the identified proteins have been implicated in adaptation, it is difficult to quantify how many of the proteins are actually used by bacteria for these functions. Therefore, gene ontology analysis is still important, as it describes the major functions and locations of proteins. GO SLIMs are cut-down versions of the GOs containing a subset of the terms in the whole GO.

Functional annotation of the proteins identified was performed by using iProXpress software. The following search criteria were set: GO SLIM cellular component, GO SLIM molecular function, and GO SLIM biological process.

The cellular component search identified 1,022 of the 3,487 proteins as cytoplasm proteins (GO:0005737) (Fig. 1A). Other components were cell part (GO:0044464; 383 proteins), membrane (GO:0016020; 191 proteins), protein complex (GO:0043234; 153 proteins), and ribosome (GO:0005840; 120 proteins).

The largest group in the molecular function group was nucleotide binding proteins (GO:0000166; 1,648 proteins) (Fig. 1B). For example, GroEL binds to ATP, while EF-G binds to GTP for their functions. Other functions frequently identified were oxidoreductase activity (GO:0016491; 695 proteins), nucleic acid binding (GO:0003676; 631 proteins), and transferase activity (GO:0016740; 550 proteins). As the total number exceeded 3,487, some proteins must have been defined in several functional groups.

Proteins required for carbohydrate metabolism were the most frequently identified proteins in the biological process group (GO:0005975; 752 proteins) (Fig. 1C). Other groups included oxidation reduction proteins (GO:0055114) (709 proteins), alcohol metabolic proteins (GO:0006066; 690 proteins), and metabolite and energy proteins (GO:0006091; 401 proteins).

The iProXpress package also provided the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis. The major pathways were metabolic pathways (KEGG 01100; 639 proteins) (Fig. 1D), biosynthesis of secondary metabolites (KEGG 01110; 360 proteins), microbial metabolism in diverse environments (KEGG 01120; 346 proteins), glycolysis (KEGG 00010; 325 proteins), and RNA degradation (KEGG 03018; 233 proteins).

DISCUSSION

In this study, we took great care to process freshly voided fecal samples as rapidly as possible, and it is possible that some of the proteins identified reflect the state of the bacteria that have responded to the fecal *ex vivo* environment (aerobic and ambient temperature), so some caution is required if translating the findings directly to the *in vivo* environment.

Large-scale sequencing reduces errors in estimations of bacte-

rial populations in the gut. In this study, >10,000 sequences were identified by using the conserved V4 and V5 regions of the 16S rRNA gene (Table 1). Other 16S rRNA gene sequencing studies have also shown a wide range of bacteria present in the gut for each bacterial genus (8, 41–43). However, as demonstrated in this study, it is not always possible to correlate the bacterial population in the gut with the protein population, as different bacterial species express different proteins under stress and other environmental and variable conditions.

Species identification is different between proteomics and 16S rRNA gene sequencing. For proteomics, more abundant proteins are identified with more peptides. However, each protein is listed only once. For 16S rRNA gene sequencing, the abundance of a species is represented by the number of times that the gene is sequenced. Therefore, 16S rRNA gene sequencing reflects the true proportion of a bacterial species within a population, assuming that each cell contains only one copy of the 16S rRNA gene. However, previous studies have shown that 16S rRNA-based analyses of metagenomic samples may be biased because of unequal PCR amplification of 16S rRNA genes from different species and could also be complicated by amplification artifacts such as chimeric sequences and sequencing errors (44). A previous study by Ashelford et al. (45) showed that at least 1 in 20 16S rRNA sequences currently in public repositories contains substantial anomalies.

Biased amplification of 16S rRNA genes can be overcome by using whole-genome shotgun (WGS) sequencing. This method of metagenomic sequencing also has its disadvantages. The relative organism abundances inferred from WGS metagenomic sequences can vary significantly depending on the DNA extraction and sequencing protocols utilized (46). In addition, shotgun metagenomic sequencing is generally considered not deep enough to detect rare species in complex communities, and low-abundance species are best identified through 16S rRNA gene sequencing (47). A recent comparative study (48) showed significant differences in the bacterially diverse populations derived from 16S rRNA gene sequencing and whole-genome shotgun metagenome sequencing of the same sample. The differences were not due to the different depths of sampling of the two methods and indicated that 16S rRNA gene sequencing can profile the bacterial communities in greater detail than can metagenomics. The results of that study indicate that even when corrected for depth, conclusions derived from 16S rRNA gene sequencing and shotgun metagenome sequencing cannot be directly compared. Our findings also seem to support this conclusion.

In this study, GroEL and DnaK were among the most frequently identified proteins, indicating stress responses. The metaproteomics analyses suggested that these stress-related proteins were much less frequently identified in human (49) and pig (Y. Tang, A. Underwood, A. Gielbert, M. J. Woodward, and L. Petrovska, unpublished data) fecal samples. Chicken body temperatures vary between 39.8°C and 43.6°C at different times of the day, while the human body temperature is 37°C and the pig body temperature is 38.8°C. Several studies have shown that, at least in pure culture, *E. coli*, *Clostridium acetobutylicum*, *Lactobacillus plantarum*, and *Campylobacter jejuni* heat shock responses were triggered when the growth temperature was increased to 42°C (16, 27–29). Thus, the main stress factor might be the higher body temperatures of the chicken gut. Other factors, such as acid stress, metabolic stress, or uric acid, might also play a role. Birds excrete excess nitrogen in the form of solid uric acid crystals as composite

excreta with fecal contents, as there is no separation of urinary and fecal excretion as in mammals. However, uric acid does not dissolve readily and is relatively nontoxic (50), but nevertheless, it may have an impact on protein degradation and/or induction of bacterial responses, although we reasoned that such an impact should be limited given that sample handling was swift, reducing exposure to the presence of the relatively insoluble uric acid. Having demonstrated the potential of the methodologies developed in this study, it is appropriate that a series of proteomics analyses of different parts of the chicken digestive tract could be performed to produce a comprehensive picture of the gut microbiota, as bacterial diversity is considerably increased along the tract from the stomach to the colon (51), and protein expression may also vary to adapt to different environments.

Although GroEL proteins were the most frequently identified proteins, GO SLIM biological process analysis indicated that most identified proteins were involved in metabolic processes of carbohydrate, alcohol, and protein (Fig. 2) instead of protein folding, suggesting that once the adaptation was induced, other processes functioned normally. This notion was supported by the KEGG pathway analysis, as many identified proteins were involved in metabolism, biosynthesis of secondary metabolites, and other metabolically related pathways. The peptides identified suggested that metabolically related proteins were more abundant, indicating that the gut bacteria were thriving.

On average, 2 to 6 peptides were associated with the GroEL proteins, indicating that these proteins were not very abundant in our data set. This raised the question of why so many of these proteins were identified in chicken fecal samples instead of some more abundant proteins, such as pyruvate kinases. The answer might lie in the sequence diversity of GroEL proteins. To be identified as a new protein, at least one unique peptide has to be identified. Proteins such as pyruvate kinases may be more conserved so that it is difficult to find many of them with unique peptides. Also, the peptides identified most frequently by using the specific methods developed in this study were assigned to GroEL and pyruvate kinase proteins. We had anticipated the detection of presumed high-abundance ribosomal proteins, as these are found in multiple copies per cell, perhaps in the range of 100 to 1,000 copies each. While these were detected, but not in the abundance of GroEL and pyruvate kinases, we suggest that methods to quantify peptides by comparison with internal standards, such as ribosomal proteins, need to be explored.

The protein extraction method has an effect on the identified proteins. In this study, although the buffer for protein extraction contained the detergent 0.1% Tween 80, only 8% of proteins were membrane proteins (Fig. 2). Genomic sequencing analysis predicts that 20 to 30% of proteins produced by most organisms are integral membrane proteins (52). This comparison suggests that 0.1% Tween 80 did not release all membrane proteins into the solution. In a previous human metaproteomics study, the cell pellet was resuspended in 6 M guanidine, which should solubilize more membrane proteins (49). To fractionate membrane proteins, 1 to 1.5% Triton X-100 was used for a human T cell line study (53). Also, the Fasp method used to generate peptides was centered on a filter column. The filter can clean samples efficiently for large-scale protein identification (54). However, some small proteins might have been lost through the filter during sample preparation, as the cutoff point was 30 kDa. Although many small proteins with molecular masses of <10 kDa were identified in this

A Cellular component

- GO:0005737 cytoplasm 44%
- GO:0044464 cell part 16%
- GO:0016020 membrane 8%
- GO:0043234 protein complex 7%
- GO:0005840 ribosome 5%
- GO:0030529 ribonucleoprotein complex 4%
- GO:0005576 extracellular region 4%
- GO:0009986 cell surface 3%
- GO:0005856 cytoskeleton 2%
- GO:0005634 nucleus 2%
- GO:0009536 plastid 1%
- GO:0030313 cell envelope 1%
- GO:0030312 external encapsulating structure 1%
- GO:0043190 ATP-binding cassette 1%

B Molecular function

- GO:0000166 nucleotide binding 27%
- GO:0016491 oxidoreductase activity 11%
- GO:0003676 nucleic acid binding 10%
- GO:0016740 transferase activity 9%
- GO:0016787 hydrolase activity 8%
- GO:0048037 cofactor binding 6%
- GO:0043167 ion binding 6%
- GO:0003824 catalytic activity 6%
- GO:0005515 protein binding 4%
- GO:0016829 lyase activity 3%
- GO:0005215 transporter activity 2%
- GO:0005198 structural molecule activity 2%
- GO:0016874 ligase activity 2%
- GO:0016853 isomerase activity 2%
- GO:0008233 peptidase activity 2%

C Biological process

- GO:0005975 carbohydrate metabolic process 13%
- GO:0055114 oxidation-reduction process 13%
- GO:0006066 alcohol metabolic process 12%
- GO:0006091 generation of precursor metabolites and energy 7%
- GO:0046483 heterocycle metabolic process 7%
- GO:0006412 translation 7%
- GO:0019538 protein metabolic process 6%
- GO:0009117 nucleotide metabolic process 6%
- GO:0006457 protein folding 6%
- GO:0006520 cellular amino acid metabolic process 5%
- GO:0008152 metabolic process 4%
- GO:0006793 phosphorus metabolic process 4%
- GO:0006351 transcription, DNA-dependent 4%
- GO:0050896 response to stimulus 2%

D Pathway statistics

- (KEGG) 01100 Metabolic pathways 23%
- (KEGG) 01110 Biosynthesis of secondary metabolites 13%
- (KEGG) 01120 Microbial metabolism in diverse environments 12%
- (KEGG) 00010 Glycolysis / Gluconeogenesis 12%
- (KEGG) 03018 RNA degradation 8%
- (KEGG) 00230 Purine metabolism 6%
- (KEGG) 00240 Pyrimidine metabolism 4%
- (KEGG) 00250 Alanine, aspartate and glutamate metabolism 4%
- (KEGG) 00330 Arginine and proline metabolism 3%
- (KEGG) 00910 Nitrogen metabolism 3%
- (KEGG) 03020 RNA polymerase 3%
- (KEGG) 00620 Pyruvate metabolism 3%
- (KEGG) 00680 Methane metabolism 3%
- (KEGG) 00290 Valine, leucine and isoleucine biosynthesis 1%

FIG 2 Gene ontology analysis and pathway statistics of the identified proteins. The analysis was carried out with the Protein Information Resource bioinformatics tool, based on the 3,487 proteins mapped to UniProt identifications.

study (see Table S3 in the supplemental material), it was a factor to consider.

Another issue in proteomic analysis is undersampling. It is often observed that many low-abundance proteins remain unde-

tected in large-scale proteomic analyses. This is caused by the limited rate of MS/MS spectra acquired from a mass spectrometer. To overcome this, peptide fractionation to reduce sample complexity and resampling are common techniques (17). In this study, a sim-

ple SAX fractionation step proved to be effective. It was also found that the different combinations of Mascot searches resulted in 30% more proteins being identified. In this study, 812 proteins were identified with Fasp only, and 3,297 proteins were identified when both Fasp and SAX methods were used in a sequential manner. In other metaproteomics studies, the numbers of identified proteins reported were between 1,000 and 2,000 (49, 55). In a single run, >2,000 proteins were identified from HeLa cells by using Fasp only (54). When both Fasp and SAX were combined, 4,000 to 5,000 proteins were identified from mouse hippocampus tissues (20). Those methods were developed by using an LTQ-Obitrap mass spectrometer, which is a more sensitive instrument than the Agilent 6520 QTOF instrument used in this study. The sample preparations for the Fasp and SAX methods were simple and fast and therefore were suitable for large-scale protein identifications.

In a previous human metaproteomics study, the most abundant microbial proteins were glyceraldehyde-3-phosphate dehydrogenase, ribosomal proteins, and electron transfer flavoproteins (49), whereas in the chicken gut, the most abundant microbial proteins were the chaperone GroEL, glyceraldehyde-3-phosphate dehydrogenase, and glutamate dehydrogenase. In that same human study, the most frequently identified proteins were from *Bacteroides*, *Bifidobacterium*, and *Clostridium* (49), which showed a good correlation with the fingerprint data (56), while in this study, the most frequently identified proteins were from *Lactobacillus*, followed by *Clostridium* and *Streptococcus*. Among the differences in diets, environments, and host immunity, higher body temperature might be accounted for the most.

In conclusion, a large number of proteins was identified by using QTOF and the Mascot search engine. Gene ontology and pathway analyses were carried out to reveal functions of the identified proteins in normal situations. This study was focused on an understanding of the microbial composition in the healthy chicken gut. Having established these baselines, future studies on quantitative analysis of fecal samples in healthy and diseased chickens could potentially reveal any changes in composition and bacterial protein expression as a result of a changed health/disease status or changed feed regimens. The effects of diet are important topics for the poultry industry. By using tools such as proteomics and 16S rRNA gene sequencing together with metabolomics, the state of gut microbiota can be described more accurately.

ACKNOWLEDGMENTS

We thank Shaun Cawthraw, Emma Kennedy for sample collection, Mark Weeks for QTOF analysis, and Maurice Sauer for scientific discussions. 16S rRNA sequencing was done by the AHVLA central sequencing unit.

This work was supported by funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement 222633.

REFERENCES

- Windhorst HW. 2006. Changes in poultry production and trade worldwide. *Worlds Poultry Sci. J.* 62:585–602. <http://dx.doi.org/10.1017/S0043933906001140>.
- Schiffrin EJ, Blum S. 2002. Interactions between the microbiota and the intestinal mucosa. *Eur. J. Clin. Nutr.* 56(Suppl 3):S60–S64. <http://dx.doi.org/10.1038/sj.ejcn.1601489>.
- Prakash S, Rodes L, Coussa-Charley M, Tomaro-Duchesneau C. 2011. Gut microbiota: next frontier in understanding human health and development of biotherapeutics. *Biologics* 5:71–86. <http://dx.doi.org/10.2147/BTT.S19099>.
- Chambers JR, Gong J. 2011. The intestinal microbiota and its modulation for Salmonella control in chickens. *Food Res. Int.* 44:3149–3159. <http://dx.doi.org/10.1016/j.foodres.2011.08.017>.
- Amit-Romach E, Sklan D, Uni Z. 2004. Microflora ecology of the chicken intestine using 16S ribosomal DNA primers. *Poult. Sci.* 83:1093–1098. <http://ps.fass.org/content/83/7/1093.long>.
- Uni Z, Tako E, Gal-Garber O, Sklan D. 2003. Morphological, molecular, and functional changes in the chicken small intestine of the late-term embryo. *Poult. Sci.* 82:1747–1754. <http://ps.fass.org/content/82/11/1747.long>.
- Yegani M, Korver DR. 2008. Factors affecting intestinal health in poultry. *Poult. Sci.* 87:2052–2063. <http://dx.doi.org/10.3382/ps.2008-00091>.
- Lu J, Idris U, Harmon B, Hofacre C, Maurer JJ, Lee MD. 2003. Diversity and succession of the intestinal bacterial community of the maturing broiler chicken. *Appl. Environ. Microbiol.* 69:6816–6824. <http://dx.doi.org/10.1128/AEM.69.11.6816-6824.2003>.
- Wong JM, de Souza R, Kendall CW, Emam A, Jenkins DJ. 2006. Colonic health: fermentation and short chain fatty acids. *J. Clin. Gastroenterol.* 40:235–243. <http://dx.doi.org/10.1097/00004836-200603000-00015>.
- Sekirov I, Russell SL, Antunes LC, Finlay BB. 2010. Gut microbiota in health and disease. *Physiol. Rev.* 90:859–904. <http://dx.doi.org/10.1152/physrev.00045.2009>.
- Apajalahti JH, Sarkilahti LK, Maki BR, Heikkinen JP, Nurminen PH, Holben WE. 1998. Effective recovery of bacterial DNA and percent-guanine-plus-cytosine-based analysis of community structure in the gastrointestinal tract of broiler chickens. *Appl. Environ. Microbiol.* 64:4084–4088.
- Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261–5267. <http://dx.doi.org/10.1128/AEM.00062-07>.
- Wang Y, Qian PY. 2009. Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* 4:e7401. <http://dx.doi.org/10.1371/journal.pone.0007401>.
- Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* 26:266–267. <http://dx.doi.org/10.1093/bioinformatics/btp636>.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <http://dx.doi.org/10.1093/bioinformatics/btq461>.
- Stintzi A. 2003. Gene expression profile of *Campylobacter jejuni* in response to growth temperature variation. *J. Bacteriol.* 185:2009–2016. <http://dx.doi.org/10.1128/JB.185.6.2009-2016.2003>.
- Wang H, Chang-Wong T, Tang HY, Speicher DW. 2010. Comparison of extensive protein fractionation and repetitive LC-MS/MS analyses on depth of analysis for complex proteomes. *J. Proteome Res.* 9:1032–1040. <http://dx.doi.org/10.1021/pr900927y>.
- Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71:8228–8235. <http://dx.doi.org/10.1128/AEM.71.12.8228-8235.2005>.
- Egan S, Lanigan M, Shiell B, Beddome G, Stewart D, Vaughan J, Michalski WP. 2008. The recovery of *Mycobacterium avium* subspecies paratuberculosis from the intestine of infected ruminants for proteomic evaluation. *J. Microbiol. Methods* 75:29–39. <http://dx.doi.org/10.1016/j.mimet.2008.04.008>.
- Wisniewski JR, Zougman A, Mann M. 2009. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* 8:5674–5678. <http://dx.doi.org/10.1021/pr900748n>.
- Farrah T, Deutsch EW, Omenn GS, Campbell DS, Sun Z, Bletz JA, Mallick P, Katz JE, Malmstrom J, Ossola R, Watts JD, Lin B, Zhang H, Moritz RL, Aebersold R. 2011. A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* 10:M110.006353. <http://dx.doi.org/10.1074/mcp.M110.006353>.
- Bessarabova M, Ishkin A, JeBailey L, Nikolskaya T, Nikolsky Y. 2012. Knowledge-based analysis of proteomics data. *BMC Bioinformatics* 13(Suppl 16):S13. <http://dx.doi.org/10.1186/1471-2105-13-S16-S13>.
- Sennels L, Bukowski-Wills JC, Rappsilber J. 2009. Improved results in proteomics by use of local and peptide-class specific false discovery rates.

- BMC Bioinformatics 10:179. <http://dx.doi.org/10.1186/1471-2105-10-179>.
24. Volchenboum SL, Kristjansdottir K, Wolfgeher D, Kron SJ. 2009. Rapid validation of Mascot search results via stable isotope labeling, pair picking, and deconvolution of fragmentation patterns. *Mol. Cell. Proteomics* 8:2011–2022. <http://dx.doi.org/10.1074/mcp.M800472-MCP200>.
 25. Hongzhan H, Shukla HD, Cathy W, Satya S. 2007. Challenges and solutions in proteomics. *Curr. Genomics* 8:21–28. <http://dx.doi.org/10.2174/138920207780076910>.
 26. Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, Lee A, van Sluyter SC, Haynes PA. 2011. Less label, more free: approaches in label-free quantitative mass spectrometry. *Proteomics* 11:535–553. <http://dx.doi.org/10.1002/pmic.201000553>.
 27. Bahl H, Muller H, Behrens S, Joseph H, Narberhaus F. 1995. Expression of heat shock genes in *Clostridium acetobutylicum*. *FEMS Microbiol. Rev.* 17:341–348. <http://dx.doi.org/10.1111/j.1574-6976.1995.tb00217.x>.
 28. Arsene F, Tomoyasu T, Bukau B. 2000. The heat shock response of *Escherichia coli*. *Int. J. Food Microbiol.* 55:3–9. [http://dx.doi.org/10.1016/S0168-1605\(00\)00206-3](http://dx.doi.org/10.1016/S0168-1605(00)00206-3).
 29. De Angelis M, Di Cagno R, Huet C, Crecchio C, Fox PF, Gobetti M. 2004. Heat shock response in *Lactobacillus plantarum*. *Appl. Environ. Microbiol.* 70:1336–1346. <http://dx.doi.org/10.1128/AEM.70.3.1336-1346.2004>.
 30. Len AC, Harty DW, Jacques NA. 2004. Stress-responsive proteins are upregulated in *Streptococcus mutans* during acid tolerance. *Microbiology* 150:1339–1351. <http://dx.doi.org/10.1099/mic.0.27008-0>.
 31. Di Cagno R, De Angelis M, Limitone A, Fox PF, Gobetti M. 2006. Response of *Lactobacillus helveticus* PR4 to heat stress during propagation in cheese whey with a gradient of decreasing temperatures. *Appl. Environ. Microbiol.* 72:4503–4514. <http://dx.doi.org/10.1128/AEM.01829-05>.
 32. Carrete R, Reguant C, Bordons A, Constanti M. 2005. Relationship between a stress membrane protein of *Oenococcus oeni* and glyceraldehyde-3-phosphate dehydrogenases. *Appl. Biochem. Biotechnol.* 127:43–51. <http://dx.doi.org/10.1385/ABAB:127:1:043>.
 33. Wilkins JC, Homer KA, Beighton D. 2001. Altered protein expression of *Streptococcus oralis* cultured at low pH revealed by two-dimensional gel electrophoresis. *Appl. Environ. Microbiol.* 67:3396–3405. <http://dx.doi.org/10.1128/AEM.67.8.3396-3405.2001>.
 34. Mailloux RJ, Singh R, Brewer G, Auger C, Lemire J, Appanna VD. 2009. Alpha-ketoglutarate dehydrogenase and glutamate dehydrogenase work in tandem to modulate the antioxidant alpha-ketoglutarate during oxidative stress in *Pseudomonas fluorescens*. *J. Bacteriol.* 191:3804–3810. <http://dx.doi.org/10.1128/JB.00046-09>.
 35. Caldas T, Laalami S, Richarme G. 2000. Chaperone properties of bacterial elongation factor EF-G and initiation factor IF2. *J. Biol. Chem.* 275:855–860. <http://dx.doi.org/10.1074/jbc.275.2.855>.
 36. Kobayashi H, Miyamoto T, Hashimoto Y, Kiriki M, Motomatsu A, Honjoh K, Iio M. 2005. Identification of factors involved in recovery of heat-injured *Salmonella Enteritidis*. *J. Food Prot.* 68:932–941.
 37. Caldas TD, El Yaagoubi A, Richarme G. 1998. Chaperone properties of bacterial elongation factor EF-Tu. *J. Biol. Chem.* 273:11478–11482. <http://dx.doi.org/10.1074/jbc.273.19.11478>.
 38. Di Cagno R, De Angelis M, Coda R, Minervini F, Gobetti M. 2009. Molecular adaptation of sourdough *Lactobacillus plantarum* DC400 under co-cultivation with other lactobacilli. *Res. Microbiol.* 160:358–366. <http://dx.doi.org/10.1016/j.resmic.2009.04.006>.
 39. Jurgen B, Breitenstein A, Urlacher V, Buttner K, Lin H, Hecker M, Schweder T, Neubauer P. 2010. Quality control of inclusion bodies in *Escherichia coli*. *Microb. Cell Fact.* 9:41. <http://dx.doi.org/10.1186/1475-2859-9-41>.
 40. Cortesby B. 2007. Roundtrip ticket for secretory IgA: role in mucosal homeostasis? *J. Immunol.* 178:27–32. <http://www.jimmunol.org/content/178/1/27.long>.
 41. Lu J, Santo Domingo J, Shanks OC. 2007. Identification of chicken-specific fecal microbial sequences using a metagenomic approach. *Water Res.* 41:3561–3574. <http://dx.doi.org/10.1016/j.watres.2007.05.033>.
 42. Zhu XY, Zhong T, Pandya Y, Joerger RD. 2002. 16S rRNA-based analysis of microbiota from the cecum of broiler chickens. *Appl. Environ. Microbiol.* 68:124–137. <http://dx.doi.org/10.1128/AEM.68.1.124-137.2002>.
 43. Holben WE, Feris KP, Kettunen A, Apajalahti JH. 2004. GC fractionation enhances microbial community diversity assessment and detection of minority populations of bacteria by denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* 70:2263–2270. <http://dx.doi.org/10.1128/AEM.70.4.2263-2270.2004>.
 44. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* 6:639–641. <http://dx.doi.org/10.1038/nmeth.1361>.
 45. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl. Environ. Microbiol.* 71:7724–7736. <http://dx.doi.org/10.1128/AEM.71.12.7724-7736.2005>.
 46. Morgan JL, Darling AE, Eisen JA. 2010. Metagenomic sequencing of an in vitro-simulated microbial community. *PLoS One* 5:e10209. <http://dx.doi.org/10.1371/journal.pone.0010209>.
 47. Kalyuzhnaya MG, Lapidus A, Ivanova N, Copeland AC, McHardy AC, Szeto E, Salamov A, Grigoriev IV, Suci D, Levine SR, Markowitz VM, Rigoutsos I, Tringe SG, Bruce DC, Richardson PM, Lidstrom ME, Chistoserdova L. 2008. High-resolution metagenomics targets specific functional types in complex microbial communities. *Nat. Biotechnol.* 26:1029–1034. <http://dx.doi.org/10.1038/nbt.1488>.
 48. Shah N, Tang H, Doak TG, Ye Y. 2011. Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. *Pac. Symp. Biocomput.* 2011:165–176. http://dx.doi.org/10.1142/9789814335058_0018.
 49. Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J, Lefsrud MG, Apajalahti J, Tysk C, Hettich RL, Jansson JK. 2009. Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* 3:179–189. <http://dx.doi.org/10.1038/ismej.2008.108>.
 50. Ehrlich PR, Dobkin DS, Wheye D. 1988. The birder's handbook. A field guide to the natural history of North American birds, including all species that regularly breed north of Mexico. Simon & Schuster, New York, NY.
 51. Videnska P, Faldynova M, Juricova H, Babak V, Sisak F, Havlickova H, Rychlik I. 2013. Chicken faecal microbiota and disturbances induced by single or repeated therapy with tetracycline and streptomycin. *BMC Vet. Res.* 9:30. <http://dx.doi.org/10.1186/1746-6148-9-30>.
 52. Lundstrom K. 2004. Structural genomics on membrane proteins: mini review. *Comb. Chem. High Throughput Screen.* 7:431–439. <http://dx.doi.org/10.2174/1386207043328634>.
 53. Rockstroh M, Müller SA, Jende C, Kerzhne A, von Bergen M, Tomm JM. 2011. Cell fractionation—an important tool for compartment proteomics. *J. Integr. OMICS* 1:135–143. <http://dx.doi.org/10.5584/jiomics.v1i1.52>.
 54. Wisniewski JR, Zougman A, Nagaraj N, Mann M. 2009. Universal sample preparation method for proteome analysis. *Nat. Methods* 6:359–362. <http://dx.doi.org/10.1038/nmeth.1322>.
 55. Kolmeder CA, de Been M, Nikkila J, Ritamo I, Matto J, Valmu L, Salojärvi J, Palva A, Salonen A, de Vos WM. 2012. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS One* 7:e29913. <http://dx.doi.org/10.1371/journal.pone.0029913>.
 56. Dicksved J, Halfvarson J, Rosenquist M, Järnerot G, Tysk C, Apajalahti J, Engstrand L, Jansson JK. 2008. Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J.* 2:716–727. <http://dx.doi.org/10.1038/ismej.2008.37>.