

Proteomic Characterization of the Gut Microbiomes of Gnotobiotic Mice and Colonization with Single or Dual Bacterial Strains

A. Russell¹, N. C. VerBerkmoes¹, M. Shah¹, M. Mahowald², J. Gordon², R. L. Hettich¹

¹ Oak Ridge National Laboratory, Oak Ridge, TN ² Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO



OVERVIEW

- A deep understanding of human genetic and physiologic diversity requires characterization of our microbiome (i.e., the genomes of our microbial symbionts), focusing on factors that influence its assembly, stability, functions, and functional variations.
- The distal gut harbors most of our microbes and is dominated by bacterial members of just two phyla - the Bacteroidetes and the Firmicutes.
- To examine how prominent representatives of these two phyla interact *in vivo*, we colonized germ-free mice (Figure 1) with sequential human gut-derived type strains of *B. thetaioatoamicon* (Figure 2) or *E. rectale*, or both.

INTRODUCTION

- Gnotobiotic mice are germ-free mice that have lived in sterile conditions (Figure 3) since birth. These mice were ideal for these studies considering their low level of microbial complexity compared to that of a normal mouse or human gut microbiome.
- Bacteroidetes harbor large paralogous sets of polysaccharide sensing, acquisition and degradation genes; gut Firmicutes have smaller genomes, and a disproportionately small number of genes involved in polysaccharide degradation.
- From genome analyses it is hypothesized that *E. rectale* has a narrower niche than prominent gut Bacteroidetes and lacks a variety of enzymes to degrade host-derived glycans, and a large number of predicted alpha amylases.
- To further define the responses of these organisms to their gut habitat, we used a 2D-LC-MS/MS shotgun proteomics approach to perform proteomic measurements of the protein products produced by these symbionts *in vivo*, and obtain insights about their functional activities in a well defined, low complexity gut ecosystem.

Figure 1. Gnotobiotic Mouse



EXPERIMENTAL

Samples and Sample Preparation

- Germ-free mice were colonized with either *B. thetaioatoamicon* or *E. rectale*, or both or none (control). Fourteen days after inoculation, both microbial species colonized the distal gut (ceca) of recipient mice at roughly similar levels, whether in mono- or bi-association.
- Two biological replicates of four cecal samples were received at different times, but processed with the same protocol as shown below (Figure 4).
- For each sample set, one mouse was not colonized with any bacteria (germ-free control); one was colonized only with *B. thetaioatoamicon*, one was colonized with both *B. thetaioatoamicon* and *E. rectale*, and the last one was colonized only with *E. rectale*. All eight samples were coded, and the MS measurements conducted without prior sample knowledge.
- Luminal contents were collected from the ceca of eight gnotobiotic mice and were lysated via a single tube cell lysis and protein digestion. Briefly, the cell pellet was re-suspended in 6M Guanidinium OH DTT to lyse cells and denature proteins. The Guanidine concentration was adjusted to 1M with Tris buffer, and then sequencing grade trypsin was added to digest proteins to peptides. The complex peptide isolation was de-salted via C18 solid phase extraction, concentrated, filtered, and aliquot.

LC-MS/MS and Informatics

- All samples were analyzed via two-dimensional (2D) nano-LC MS/MS system with a split-phase column (SCX-RP) on a linear ion trap (Thermo Fisher Scientific), with a 22 hour run per sample. The first sample set was analyzed in technical duplicates while the second sample set was analyzed in technical triplicates.
- The LTO settings were as follows: dynamic exclusion set at one, five data-dependent MS/MS in a LTO, 2 microscans were averaged for both Full and MS/MS scans, centroid data was collected for all scans.
- Database searches were performed with SEQUEST (Eng, 1994) against the entire mouse genome, *B. thetaioatoamicon*, *E. rectale*, rice, and yeast genomes (common contaminants such as keratin and trypsin were included).
- The SEQUEST settings were as follows: enzyme type, trypsin; Parent Mass Tolerance, 3.0; Fragment Ion Tolerance, 0.5; up to 4 missed cleavages allowed (general lysine and arginine residues), and fully tryptic peptides only (both ends of the peptide must have arisen from a tryptic specific cut, except N and C-terminal of proteins). All datasets were filtered at the individual run level with DTASelect (Tabb, 2002) [cutoffs of at least 1.8 (+/-), 2.5 (-) < 3.5 (+/-)]. Only proteins identified with two fully tryptic peptides were considered for further biological study. Previous studies with reverse database searching have shown this filter level to generally give a false positive rate less than 1% even with large databases (Ram, 2005; Lo, 2007).

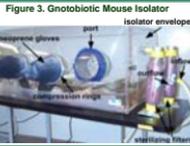
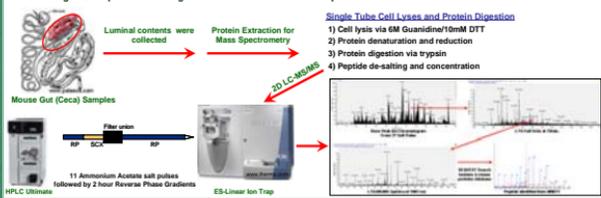


Figure 3. Gnotobiotic Mouse Isolator

Table 1. Breakdown of Database

Database	Protein Size	MSI
<i>B. thetaioatoamicon</i>	1,121	1,121
<i>E. rectale</i>	3,121	3,121
Mouse	13,121	13,121
Rice	14,121	14,121
Yeast	15,121	15,121
Control	16,121	16,121

Figure 4. Experimental Design for Gnotobiotic Mice Ceca Sample Collection and Proteomic MS Characterization



Proteomics of the Mouse Cecum

- The long term goal of these studies is to apply proteomics techniques to understand the physiology of complex microbial communities in the gut microbiome.
- The short term goal of these studies was to test current shotgun proteomics techniques from sample collection and cell lysis to MS characterization and informatics on low complexity mouse gut (ceca) samples.
- Eight samples, from a double blind study, were run in either duplicate or triplicate by 2D-MS/MS on a LTO. These samples contained chromatograms with 10,000s of peptides, as illustrated in Figure 4. Each replicate was run for 24 hrs.
- The MS/MS spectra searched against one database (Table 1), which includes the human gut-derived type strains of *B. thetaioatoamicon*, *E. rectale*, rice, yeast, and common contaminants as well as the mouse genome. Table 2 and 3 illustrate the total number of proteins, peptides, and spectra identified from each sample set with each sample, run number and the species that was inoculated in the cecum specified.
- For peptide identifications, the comprehensiveness of the database employed is very important; thus this was designed to include not only proteins from the relevant colonizing bacteria, but also from mouse and food. To find potential food proteins, yeast and rice databases were included (the mice were fed yeast extracts).
- Table 1 shows the breakdown of the database with each entry and it's size.
- The measured proteomes had high reproducibility in terms of total number of proteins observed and spectra matching to each species (Figures 5 and 6), while biological or experimental variability between samples was higher for both total proteins and abundances via spectral counts.

Table 2. Sample Set 1 Preliminary Identifications

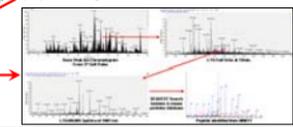
Sample	Protein	Peptide	Species	Identification
700 Run	152	348	792	None (control)
701 Run	152	348	792	None (control)
702 Run	152	348	792	<i>B. thetaioatoamicon</i>
703 Run	152	348	792	<i>B. thetaioatoamicon</i>
704 Run	152	348	792	<i>B. thetaioatoamicon</i> and <i>E. rectale</i>
705 Run	152	348	792	<i>B. thetaioatoamicon</i> and <i>E. rectale</i>
706 Run	152	348	792	<i>E. rectale</i>
707 Run	152	348	792	<i>E. rectale</i>

Table 3. Sample Set 2 Preliminary Identifications

Sample	Protein	Peptide	Species	Identification
708 Run	152	348	792	None (control)
709 Run	152	348	792	None (control)
710 Run	152	348	792	<i>B. thetaioatoamicon</i>
711 Run	152	348	792	<i>B. thetaioatoamicon</i>
712 Run	152	348	792	<i>B. thetaioatoamicon</i> and <i>E. rectale</i>
713 Run	152	348	792	<i>B. thetaioatoamicon</i> and <i>E. rectale</i>
714 Run	152	348	792	<i>E. rectale</i>
715 Run	152	348	792	<i>E. rectale</i>

Single Tube Cell Lysis and Protein Digestion

- Cell lysis via 6M Guanidinium OH DTT
- Protein denaturation and reduction
- Protein digestion via trypsin
- Peptide de-salting and concentration



Abundant Proteins and Microbial Species in Gut (Cecal) Samples

- Distinct microbial species were obtained from the intestinal ceca areas of eight gnotobiotic mice. All eight samples were coded, and the MS measurements were conducted without any prior knowledge of the microbial constitution of each sample.
- From this initial study, several 1000s of proteolytic peptides were identified in each sample by the 2D-LC-MS/MS method and ~600-1500 proteins were identified per sample.
- The resultant microbial species distributions were exactly as expected from the coded samples (Figures 5 and 6).
- Table 1 lists some of the representative proteins identified in all samples (not including the control's 700 and 799).
- Of the identified proteins from all samples, common metabolic proteins such as ribosomal proteins, glycolysis/TA proteins, and chaperone proteins were routinely identified.
- The level of protein hits to "food" proteins was fairly constant between all samples, and virtually no peptides from the other bacterium were found in the single colonized samples.
- Figure 7. *B. thetaioatoamicon* and *E. rectale* proteins identified in sample #810 from the database search were broken down into COG categories and compared. We found that several COG categories were more highly represented in *B. thetaioatoamicon* (except for cell motility).
- Several hypothetical proteins were identified with high spectral counts, such as one very large hypothetical protein (2090Da), with a known function annotated as a mouse protein "deleted in malignant brain tumor 1" (DMBT1) (Figure 8). Studies have proposed that this protein family plays a role in the mucosal and cellular immune response in the brain, digestive tract and lung. A consequence of alterations and/or the loss of DMBT1 function may lead to cell carcinomas. With limited statistics available in this study it is difficult to draw any conclusions.
- Figure 9 represents the total number of unique *B. thetaioatoamicon* and *E. rectale* spectra compared to that of the non-unique spectra in sample #810

Figure 5. Total Unique Spectra Count for Each Entry in Sample Set 1

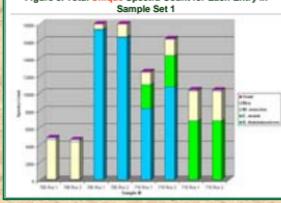


Figure 6. Total Unique Spectra Percentages for Each Entry in Sample Set 2

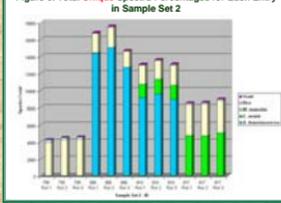


Table 4. Top 6 Abundant Protein Entries Found In All Samples

Protein	Accession	Size	MW	PI
Deleted in malignant brain tumor 1 (DMBT1)	U08001	1,188	2090	5.1
Protein	U08001	1,188	2090	5.1
Protein	U08001	1,188	2090	5.1
Protein	U08001	1,188	2090	5.1
Protein	U08001	1,188	2090	5.1
Protein	U08001	1,188	2090	5.1

Figure 7. Cecal Sample #810 Protein Counts Represented by COG Functions

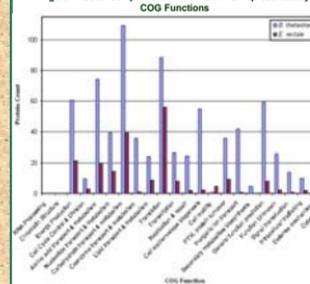


Figure 8. Deleted Malignant Brain Tumors 1 (DMBT1) Identified with High Spectral Counts

Sample	Scan	Score	Protein	Length	Mass	Charge	Mod
700	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
701	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
702	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
703	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
704	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
705	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
706	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0
707	1000	10.0	Deleted in malignant brain tumor 1	1188	2090	5	0

Sample#700 Run 1 DMBT1 Sequence Coverage

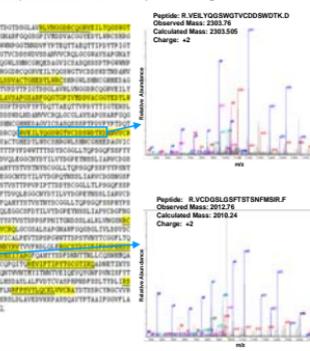


Figure 9. Unique *B. thetaioatoamicon* and *E. rectale* vs Non-unique spectra comparison



DISCUSSION

- To explore the interactions between the Bacteroidetes and Firmicutes *in vivo*, we colonized germ-free mice with *B. thetaioatoamicon* or *E. rectale*, or both species together.
- B. thetaioatoamicon* adapts to the presence of *E. rectale* by upregulating its expression of a variety of polysaccharide utilization genes as shown in Figure 7 with a high protein abundance in carbohydrate transport and metabolism.
- In contrast, microarray analysis has shown that *E. rectale*'s response to *B. thetaioatoamicon* is to downregulate glycoside hydrolases genes and to upregulate peptide and sugar transporters, suggesting that *E. rectale* is adapting to the presence of *B. thetaioatoamicon* by deriving more of its energy from simple sugars and peptides liberated by enzymes produced by *B. thetaioatoamicon*.

CONCLUSIONS

- Preliminary data shows great promise for this technique, with excellent technical reproducibility for both microbial species and proteins identified.
- We have applied shotgun proteomics to 8 mouse ceca samples and the mass spectrometric results match the inoculated microbial species distributions as expected.
- Most abundant identified proteins from both microbes included common metabolic proteins: ribosomal proteins, elongation factors, chaperones, and proteins involved in energy metabolism.
- MS allows for confident semi-quantitative species identification.
- By using low complexity gnotobiotic mice and mass spectrometry, we can simultaneously monitor host and microbial responses to colonization in mice and begin to understand how both species cooperate and compete in the presence of each other.

ACKNOWLEDGMENTS

We thank Dr. David Tabb and the Yates Proteomics Laboratory at Scripps Research Institute for DTASelect/Contrast software, the Institute for Systems Biology for proteomic bioinformatics tools used in analysis of the MS data. The ORNL part of this work was sponsored in part by U.S. Department of Energy under Contract DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.