



# Comparison of Direct and Indirect Extraction Methods for Deep Metaproteomics of the Gut Microbiota

Kristen L. Corrier<sup>1,2</sup>; Alison R. Erickson<sup>1,2</sup>; Regina Lamendella<sup>4</sup>; Brandi L. Cantarel<sup>3</sup>; Manesh Shah<sup>1</sup>; Claire Fraser-Liggett<sup>3</sup>; Robert L. Hettich<sup>1</sup>; Janet Jansson<sup>4</sup>; Marcelo Sztein<sup>3</sup>; Nathan C. VerBerkmoes<sup>1</sup>  
<sup>1</sup>Oak Ridge National Laboratory, Oak Ridge, TN; <sup>2</sup>University of Tennessee, Knoxville, TN; <sup>3</sup>University of Maryland School of Medicine, Baltimore, MD; <sup>4</sup>Lawrence Berkeley National Laboratory, Berkeley, CA



## OVERVIEW

- Throughout the human gastrointestinal tract is a complex microbial community which is critical to the health of the human host.
  - These microbes provide many key metabolic functions lacking in the host.
  - It is hypothesized that disturbances in this natural microflora can lead to many disease states.
- The gastrointestinal tract can be attacked by food borne pathogens thus causing diverse disease states.
  - Many of these pathogens can be effectively eliminated through vaccination.
- Metaproteomic analyses of fecal material provides a snapshot of the entire gastrointestinal system and may provide insight into host-microbe interactions, host disease state and vaccination mechanisms.
- Our goal is to test, compare and develop an advanced method for proteome characterization of the host and microflora proteomes from fecal material.

## INTRODUCTION

The human GI tract represents one of the most direct interfaces between human immunity and microorganisms, both pathogens and commensal flora, as well as innocuous dietary antigens. The GI tract is, thus, a key area for the study of human mucosal immunology, particularly in exploring the immune response against enteric pathogens. The application of mass spectrometry-based proteomics has had a significant impact on our understanding of healthy and diseased states, but this effort has primarily focused on specific cell types and serum/plasma studies. While fecal material has long been known to be a good proxy for the underlying physiology of the human GI tract, there has been virtually no effort in defining a "healthy" GI tract via the analyses of the proteome of fecal material. This approach has primarily been hindered by the difficulty in successful extraction of proteins from fecal material in a format amenable to proteomic analyses and limited application of emerging advanced MS techniques that can handle the complexity and dynamic range problems posed by fecal components.

**Aim 1:** Development of sample preparation methods for unbiased and effective extraction of the entire fecal proteome as a proxy for activity in the gastrointestinal tract, with sample preparation optimization focused on maximizing measurement of immunological components.

**Aim 2:** Development of advanced liquid chromatography-mass spectrometry methods for shotgun proteome analyses and deep coverage of the "healthy" fecal proteome from several human subjects, as well as an elderly population, in order to define core fecal immunological components.

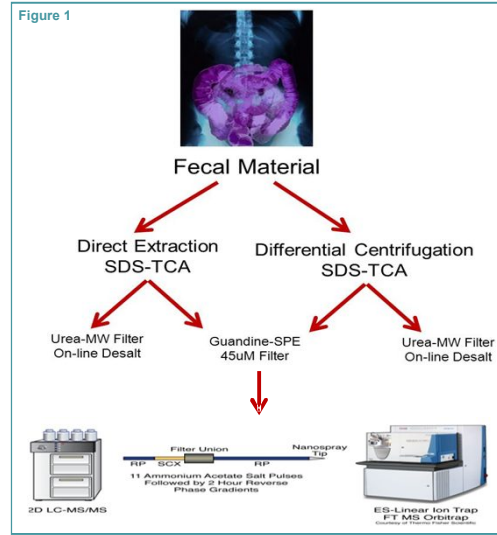
## METHODOLOGY

### Samples and Sample Preparation

- Fecal samples were obtained from adult healthy volunteers and prepared as below (Figure 1).
- Ten grams of fecal material was mixed with 10mL phosphate buffered saline (PBS), placed on ice and homogenized.
- Half of the sample was prepared via a direct extraction method with SDS/boiling for protein extraction followed by TCA prep. to clean up the proteome.
- The remaining half of the sample was processed via differential centrifugation, which enriches formicrobial cells. Pellets were then treated with SDS/boiling followed TCA prep.
- For both extraction methods, the pellets were divided and processed by the following:
  - Guanidine/DTT denaturation, trypsin digestion and solid phase extraction for desalting followed by 45uM filter to remove particulates.
  - Urea/DTT denaturation, trypsin digestion, 10kDa MW filter to remove particulates and undigested proteins followed by on-line desalting on a RP-SCX back column.

### LC/LC-MS/MS and Informatics

- All samples were analyzed via two-dimensional (2D) nano-LC MS/MS system with a split-phase column (Urea method RP-SCX-RP or Guanidine method SCX-RP) on a LTQ-Orbitrap, LTQ-Velos or Velos Orbitrap (Thermo Fisher Scientific) with 22 hour runs per sample as shown in Figure 1.
- The Orbitrap settings were as follows: 30K resolution on full scans in Orbitrap, all data-dependent MS/MS in LTQ (top five), 2 microscans for both full and MS/MS scans, centroid data for all scans. Velos analyses were run in the exact same method except top ten data dependent MS/MS.
- A protein database was built by downloading Human proteins from NCBI (HrefSeq) and appended with 21 microbial isolate genomes from the Human Microbiome Genome efforts (JGI/MG/HMP), common contaminants were also added.
- All MS/MS spectra were searched with the SEQUEST algorithm and filtered with DTASelect/Contrast at the peptide level [Xcorr of at least 1.8 (+1), 2.5 (+2) 3.5 (+3)]. Only proteins identified with two fully tryptic peptides were considered for further analyses.



## RESULTS AND DISCUSSION

### Compare Results For All Methods Via:

- A) Number of Human proteins' spectra to bacterial proteins' spectra
- B) Total number of proteins and spectra assigned
- C) Number of immunological and other human proteins
- D) Any difference in microbial abundances (i.e., is any microbe being differentially affected)
- E) Quality of Full Scans and MS/MS scans (i.e., assessment of non-peptide interference)

### General Note:

- The Urea-MW/SPE method is easier to automate.
- The Urea-MW Filter method provides better protein denaturation and digestions thus better deeper dynamic range at the cost of rapid sample prep.

We first compared raw LC-MS/MS spectra from both the direct and differential method (Figure 2 shows a representative raw file). Generally both methods gave thick chromatograms with large quantities of peptides. The direct approach had higher backgrounds (presumably from contaminants) and some extremely abundant peptides, mainly arising from abundant human proteins (Table 2 immunological proteins highlighted in yellow).

Table 1 illustrates proteins and spectra arising from human and bacterial components. This table clearly shows the differential method provides more bacterial spectra while the direct method results in greater number of human spectra. (Data collected on Velos Orbitrap.)

Figure 3 illustrates that not all bacteria are equally represented in the direct and differential method. Here three representative bacterial were chosen. Bacteroides shows large difference between the extraction methods while the other two microbes are largely unaffected.

Figure 4 illustrates reproducibility by comparing two runs using the same method.

Figure 5 illustrates the large difference in spectral counts between the methods.

Table 1: Protein and Spectra totals from both methods

Sample Name	Bacterial Protein Count	Bacterial Spectral Count	Human Protein Count	Human Spectral Count	Total Proteins	Total Spectra
Diff Cent Guanidine SPE Run1	2411	44839	322	8216	2733	53055
Diff Cent Guanidine SPE Run2	2172	39099	287	8508	2459	47607
Diff Cent Urea-MW Filter Run1	2621	34105	279	3873	2900	37978
Diff Cent Urea-MW Filter Run2	2494	33911	253	4087	2747	37998
Direct Ext Guanidine SPE Run1	1553	17599	344	14355	1897	31954
Direct Ext Guanidine SPE Run2	1724	19550	382	14357	2106	33907
Direct Ext Urea-MW Filter Run1	1652	21491	332	12481	1984	33972
Direct Ext Urea-MW Filter Run2	1958	19699	389	10631	2347	30330

All previous efforts in the HMP project with fecal material involved proteome analyses of enriched microbial fractions (VerBerkmoes, ISME, 2008).

### Potential Disadvantages:

- Enriched microbial fractions may not be a direct representation of the microbiota in the gastrointestinal tract. (i.e. might enrich one microbe over another).
- The enrichment process might "change" the proteomes of the microbial component and might not include the entire host proteome.

### Potential Advantages:

- Enriched microbial fractions may actually contain specifically interacting host immune and other proteins.
- The enrichment process removes large amounts of interfering small molecules and other non-protein materials.
- Proteomic methods have already developed for enriched pellet.

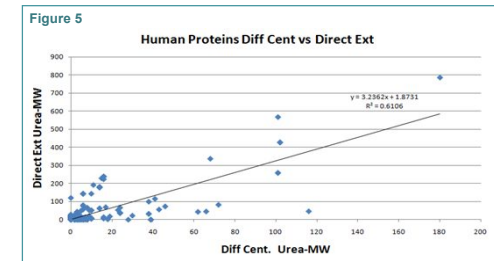
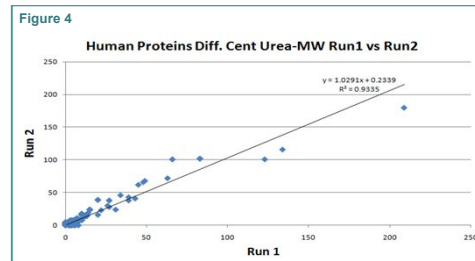
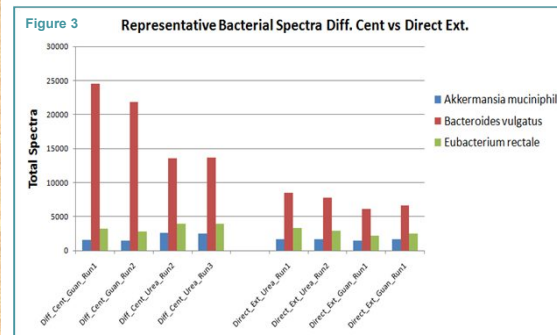
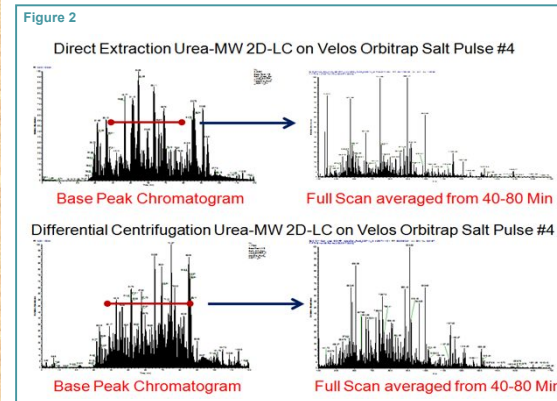


Table 2: 20 Abundant Human Proteins across both methods raw spectra.

Protein	Diff. Cent. Urea-MW	Diff. Cent. Guanidine	Direct Ext. Urea-MW	Direct Ext. Guanidine	Average	Description					
HP_HG_14833842_ref_seq_015568.2	811	1046	81	202	2109	deleted in malignant brain tumors 1 protein isoform 1 precursor (Homo sapiens)					
HP_HG_14833844_ref_seq_00549.2	811	1046	81	202	2109	deleted in malignant brain tumors 1 protein isoform 1 precursor (Homo sapiens)					
HP_HG_14833846_ref_seq_00579.4	811	1046	81	202	2109	deleted in malignant brain tumors 1 protein isoform 1 precursor (Homo sapiens)					
HP_HG_14833848_ref_seq_00497.2	811	1046	81	202	2109	deleted in malignant brain tumors 1 protein isoform 1 precursor (Homo sapiens)					
HP_HG_14833850_ref_seq_00988.2	491	499	125	307	366	IGFBP-binding protein precursor (Homo sapiens)					
HP_HG_14833852_ref_seq_01105.1	200	75	209	42	364	calreticulin precursor (Homo sapiens)					
HP_HG_14833854_ref_seq_01121.1	300	105	12	14	330	alpha-1-antitrypsin precursor (Homo sapiens)					
HP_HG_14833856_ref_seq_01382.2	81	85	49	48	54	calreticulin precursor (Homo sapiens)					
HP_HG_14833858_ref_seq_01178.1	225	229	104	208	197	chymotrypsin-like elastase family member 8B precursor (Homo sapiens)					
HP_HG_14833860_ref_seq_00865.2	114	140	38	14	71	34	polymerase (immunoglobulin receptor precursor (Homo sapiens)				
HP_HG_14833862_ref_seq_00429.2	19	10	10	17	26	140	alpha-amylase 1 precursor (Homo sapiens)				
HP_HG_14833864_ref_seq_00809.2	81	10	7	11	30	141	multiple glutathionase, intestinal (Homo sapiens)				
HP_HG_14833866_ref_seq_01137.1	79	17	1	1	126	112	46	49	58.25	immunoglobulin lambda-like polypeptide 3 (Homo sapiens)	
HP_HG_14833868_ref_seq_00362.1	81	20	0	0	42	58	102	102	51.125	transferrin precursor (Homo sapiens)	
HP_HG_14833870_ref_seq_00019.1	388	195	0	0	27	25	4	5	54.625	phospholipase A2 precursor (Homo sapiens)	
HP_HG_14833872_ref_seq_00129.2	81	42	11	24	79	89	37	37	51.125	calcium-activated chloride channel regulator 1 precursor (Homo sapiens)	
HP_HG_14833874_ref_seq_00689.4	4	4	7	10	42	42	142	142	52.25	iron-sulfur NS-A (Homo sapiens)	
HP_HG_14833876_ref_seq_00176.2	21	29	10	17	71	71	49	49	45.125	alpha-1-antichymotrypsin precursor (Homo sapiens)	
HP_HG_14833878_ref_seq_04808.2	46	85	34	46	121	121	70	50	44.625	reticulonucleotide pyrophosphatase/phosphodiesterase family member 7 precursor (Homo sapiens)	
HP_HG_14833880_ref_seq_00695.2	81	48	48	42	35	41	49	49	48	48	reticulonucleotide pyrophosphatase/phosphodiesterase family member 1 precursor (Homo sapiens)

## CONCLUSIONS

- The goal of this study was to compare two different methods for extracting bacterial and human proteins from human fecal material.
- Generally the LC-MS raw spectra are cleaner from the differential method with less non-peptide interference. The raw spectra from the direct method tend to have more dominant peptide peaks arising from very abundant human proteins.
- The differential centrifugation method provides a deeper look into the bacterial community.
- From all 20 microbes included in the database search Bacteroides was the most heavily impacted by the extraction type with almost twice the number of spectra from the differential method (Figure 6). Many bacterial species were not effected by the extraction method.



- As each method was shown to extract a somewhat unique human protein set both methods should be employed if human proteins are the focus.
- If bacterial proteins are the focus then the differential method appears to be superior.
- These results must be shown to be consistent across numerous samples from different volunteer types including the elderly and vaccinated.

## ACKNOWLEDGMENTS

We thank Dr. David Tabb and the Yates Proteomics Laboratory at Scripps Research Institute for DTASelect/Contrast software, and the Institute for Systems Biology for proteome informatics tools used in analysis of the MS data.

This research sponsored by the U.S. National Institute of Health and by the U.S. Department of Energy, with Oak Ridge National Laboratory, managed and operated by UT-Battelle, LLC.