015

Whole Genome Metagenomics as a Tool for Probiotics Analysis

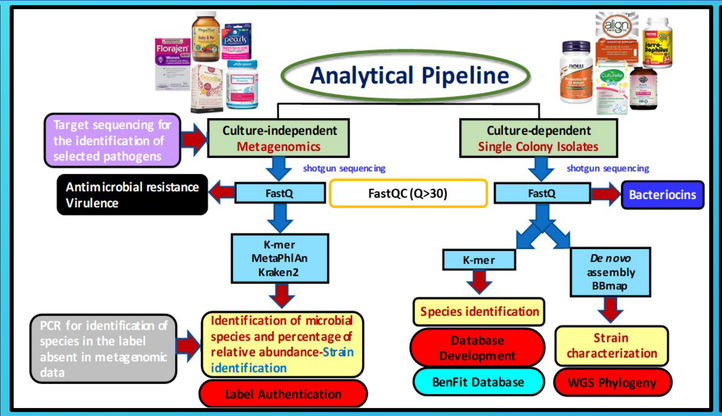
Carmen Tartera, Jayanthi Gangiredla, Mark Mammel, Isha Patel

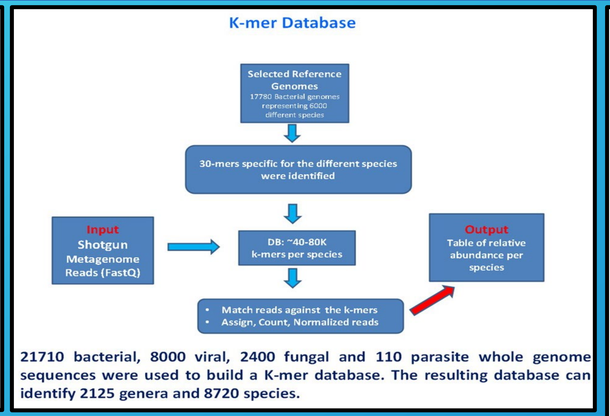
Center for Food Safety and Applied Nutrition, Office of Applied Research and Safety Assessment,
8301 Muirkirk Road, Laurel, MD 20708

Introduction

Many dietary products available to the consumer, such as dietary supplements, contain live microorganisms purported to confer a human health benefit. There are reports of label discrepancies of commercial probiotics where species are often misclassified or absent and contamination by microorganisms not listed on the label. The goal of this study is to provide a better understanding of the quality of probiotic products, the accuracy of their label, and the ability to identify potential contaminants and low-level constituents using Whole Genome Sequencing (WGS) metagenomics. In this study, DNA was purified from 123 probiotic products and metagenomic sequencing was performed using the Illumina MiSeq platform. The generated sequences were used to identify microbial constituents with unique species-specific signatures based on a novel in-house K-mer database. In parallel, using a culture-dependent approach, the microbial contents of the products were grown for single colony isolation followed by WGS to create a genome sequence database of beneficial microbes. Additionally, Resistance Gene Identifier and Virulence Finder bioinformatic tools were used to identify the presence of antibiotic resistance and virulence factor genes. A challenge in the metagenomic approach is the detection of microbes present in low numbers (constituents, contaminants, and pathogens) in the presence of high numbers of intentionally added species like *Lactobacillus* and *Bifidobacterium*. To circumvent this problem two approaches were evaluated: the use of specific phages and/or purified bacteriophage lysins to reduce the product's indigenous microorganisms to improve the detection of low-level microbial constituents and using target-amplicon sequencing. These studies will lead to a better understanding of the content of live microbial supplements available to the consumers and provide an analytical pathway to detect harmful pathogens present in low amounts.

Materials and Methods





21710 bacterial, 8000 viral, 2400 fungal and 110 parasite whole genome sequences were used to build a K-mer database. The resulting database can identify 2125 genera and 8720 species.

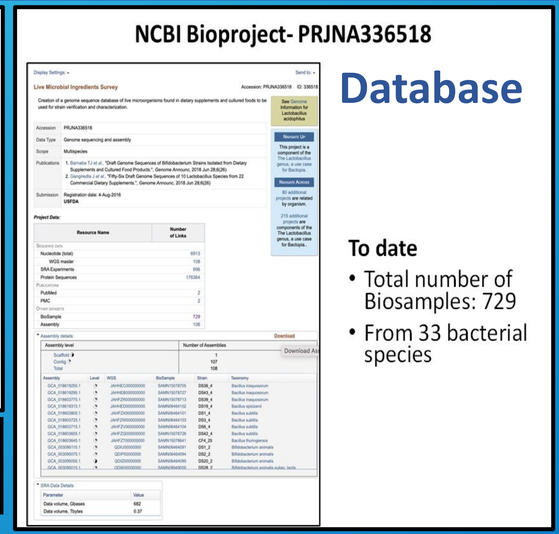
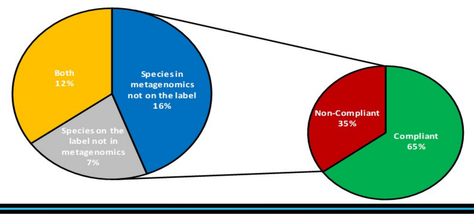


Figure 1. Microbial contents within dietary supplements and cultured foods were analyzed using a culture-independent, metagenomics-based approach. Shotgun genomic sequence data was used to identify microbial constituents in various products with unique species-specific signatures, based on a novel K-mer database. In parallel, using a culture-dependent approach, the microbial contents of the dietary products were grown for isolation under different temperature and atmospheric conditions (*i.e.,* anaerobic, microaerophilic, capnophilic, and aerobic) to allow for the growth of a broad spectrum of microbial species.

Shotgun Metagenomics a "One-Stop-Shop Results

- A convenience sample of 143 live microbial supplements, currently sold in the US market, were evaluated by WGS.
  - WGS failed for 17 products
  - Microbial components identified as proprietary blend in 3 products
- From the 123 products that generated data:
  - Label compliance for 80 products (65%)
  - Label discrepancies found for 43 products (35%):
    - 19 products : Species identified in the metagenomics data not listed on the label (16%)
    - 9 products: Species on the label not identified using shotgun metagenomics (7%)
    - 15 products: Both (12%)



Label Authentication:

- Relative abundance > 1%
- Relative abundance 0.1-1%
- Relative abundance < 0.1

LABEL:
Lactobacillus plantarum
Lactobacillus acidophilus
Lactobacillus salivarius
Lactobacillus gasseri
Bifidobacterium lactis
Bifidobacterium breve
*Bifidobacterium longum**
Bifidobacterium bifidum
*Bifidobacterium infantis**

**B. infantis* is a subspecies of *B. longum*

Species List-Product Label	Number of Reads	% Relative Abundance
Total Number of Reads	3,441,517	
no_id	2,539,633	
Lactobacillus_acidophilus	379,995	46.84
Lactobacillus_plantarum	221,152	35.03
Bifidobacterium_animalis	280,448	16.72
Bifidobacterium_longum	18,251	1.13
Bifidobacterium_breve	112	0.006
Lactobacillus_gasseri	11	0.001
Bifidobacterium_bifidum	4	0
Lactobacillus_salivarius	0	0

Figure 2. Absence of species indicated in the label at levels below 0.1% need to be confirmed by BLAST/or species-specific PCR/or targeted sequencing.

Identification of Species at Strain Level

- Strain Level Characterization of Species from Metagenomics Reads

LABEL
Bifidobacterium lactis HN019
Bifidobacterium lactis BI-04
Bifidobacterium lactis Bi-07
Lactobacillus acidophilus La-14
Lactobacillus salivarius Ls-33
Lactobacillus paracasei Lpc-37

More than 99.9987% similar!!!

Kmer Analysis	% of relative abundance
Bifidobacterium animalis subsp. lactis	92.45
Lactobacillus acidophilus	6.55
Lactobacillus casei/paracasei	0.64
Lactobacillus salivarius	0.31

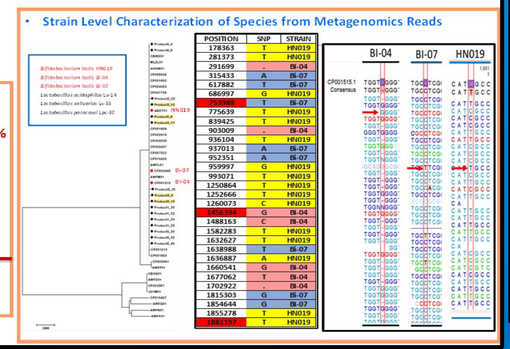


Figure 3. Size genome for *Bifidobacterium animalis* subsp. *lactis* is 1.94 Mb. Core genome total length of 1,364,751 bp that include 1,263 core genes. In products where different strains of the same species are present that are highly similar, strain-specific biomarkers based on SNP analysis can be used for identification.

Identification of Antibiotic Resistance Genes (ARG)

Product	AMR Gene	Drug Class	Resistance Mechanism	AMR Gene Family	% Length of Reference Sequence	Chromosome/Plasmid
Product 1	*lmrC*	licosamide antibiotic	antibiotic target protection	ABC-F ATP-binding cassette ribosomal protection protein	100	chromosome
	lmrD	licosamide antibiotic	antibiotic efflux	ATP-binding cassette (ABC) antibiotic efflux pump	100	chromosome
Product 9	*lnuA*	licosamide antibiotic	antibiotic inactivation	lincosamide nucleotidyltransferase (LNU)	100	plasmid

Product SCI	AMR Gene	Drug Class	Resistance Mechanism	AMR Gene Family	% Length of Reference Sequence	Species Identification
Product 1-SCI 39	*lmrC*	licosamide antibiotic	antibiotic target protection	ABC-F ATP-binding cassette ribosomal protection protein	100	*Lactococcus lactis*
	lmrD	licosamide antibiotic	antibiotic efflux	ATP-binding cassette (ABC) antibiotic efflux pump	100	
Product 9-SCI 10	*lnuA*	licosamide antibiotic	antibiotic inactivation	lincosamide nucleotidyltransferase (LNU)	100	*Lactobacillus reuteri*

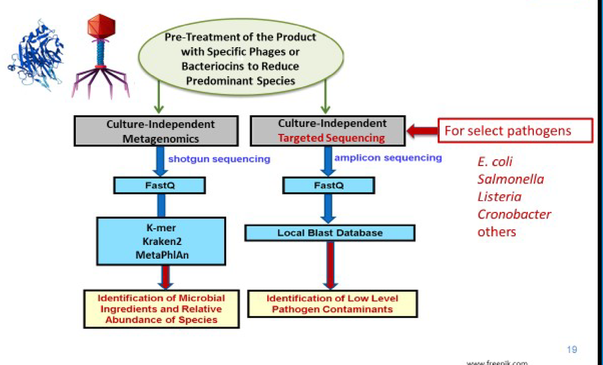
Identification of Virulence Factors Genes

Product	Virulence Factor Gene	Virulence Factor
Product 4	*bslA*	hydrophobin BslA (VF0419)
Product 10	*acm*	collagen adhesin precursor Acm (VF0419)
	sgrA	cell wall anchored protein SgrA (VF0540)

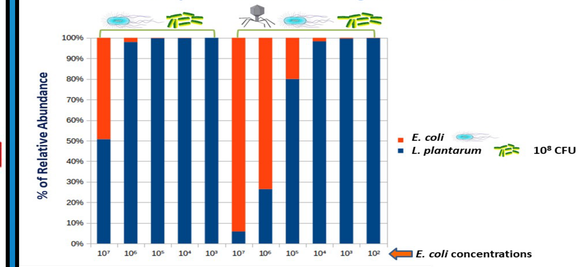
Product	Virulence Factor Gene	Virulence Factor	Species Identification
Product 4-SCI 1	*bslA*	hydrophobin BslA (VF0419)	***Bacillus subtilis***
Product 10-SCI 8	*acm*	collagen adhesin precursor Acm (VF0419)	***Enterococcus faecium***
	sgrA	cell wall anchored protein SgrA (VF0540)	

Figure 4. Antibiotic resistance was identified from metagenomic data and isolates using the Antibiotic Resistance Gene (ARG) Identifier bioinformatic tool. Virulence factors were identified with the Virulence Finder (VF) Database tool. Analysis of whole genome sequences derived from single colony isolates (SCI) with this tools identified the bacterial species harboring the ARG and VF traits.

Potential Use of Specific Phages or Bacteriocins to Increase Sensitivity of Detection of Low-Level Constituents and Contaminants



Improved Sensitivity of Detection of *E. coli* After Infection of *L. plantarum* with Phage B-1 for 3 Hours



E. coli CFU	10^7	10^6	10^5	10^4	10^3	10^2
Number of reads	50012	22433	8650	2767	88	11
stx 2A	pos	pos	pos	pos	pos	pos

Targeted Sequencing (*E. coli*)

Figure 5. Bacteriophages can be useful tools to facilitate the detection of low-level microbial species present in probiotics products. However, targeted sequencing might be more efficient for the identification of selected pathogens and presence of virulence genes.

Conclusions

- An enhanced database will improve species taxonomy, strain identification, and ultimately product development
- Whole Genome Metagenomic sequencing offers high potential for use in post-market surveillance and labeling verification
- Bacteriophages and bacteriocins could potentially be used to improve detection of labeled low-level constituents and contaminants

❖ This work was supported by the FDA Office of Women's Health. This project was supported in part by an appointment to the ORISE Research Participation Program in CFSAN, U.S. Food and Drug Administration, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and FDA/CFSAN.
❖ This project was supported in part by the Office of Dietary Supplement Programs, CFSAN, FDA.

Mission Relevance
The objective of this study is to provide a better understanding of the quality of probiotic products, the accuracy of their label, and the ability to identify potential contaminants and low-level constituents using Whole Genome Sequencing (WGS) metagenomics.