AEM
Journals.ASM.org

Application of Metagenomic Sequencing to Food Safety: Detection of Shiga Toxin-Producing *Escherichia coli* on Fresh Bagged Spinach

Susan R. Leonard, Mark K. Mammel, David W. Lacher, Christopher A. Elkins

Division of Molecular Biology, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, Maryland, USA

Culture-independent diagnostics reduce the reliance on traditional (and slower) culture-based methodologies. Here we capitalize on advances in next-generation sequencing (NGS) to apply this approach to food pathogen detection utilizing NGS as an analytical tool. In this study, spiking spinach with Shiga toxin-producing *Escherichia coli* (STEC) following an established FDA culture-based protocol was used in conjunction with shotgun metagenomic sequencing to determine the limits of detection, sensitivity, and specificity levels and to obtain information on the microbiology of the protocol. We show that an expected level of contamination (~10 CFU/100 g) could be adequately detected (including key virulence determinants and strain-level specificity) within 8 h of enrichment at a sequencing depth of 10,000,000 reads. We also rationalize the relative benefit of static versus shaking culture conditions and the addition of selected antimicrobial agents, thereby validating the long-standing culture-based parameters behind such protocols. Moreover, the shotgun metagenomic approach was informative regarding the dynamics of microbial communities during the enrichment process, including initial surveys of the microbial loads associated with bagged spinach; the microbes found included key genera such as *Pseudomonas*, *Pantoea*, and *Exiguobacterium*. Collectively, our metagenomic study highlights and considers various parameters required for transitioning to such sequencing-based diagnostics for food safety and the potential to develop better enrichment processes in a high-throughput manner not previously possible. Future studies will investigate new species-specific DNA signature target regimens, rational design of medium components in concert with judicious use of additives, such as antibiotics, and alterations in the sample processing protocol to enhance detection.

Shiga toxin-producing *Escherichia coli* (STEC) is a foodborne pathogen that can cause severe illness such as hemorrhagic colitis and hemolytic-uremic syndrome (HUS) in humans (1). Fresh leafy vegetables have been implicated in many foodborne disease outbreaks, including an STEC O157:H7 outbreak associated with consumption of fresh ready-to-eat bagged spinach in 2006 that caused 100 hospitalizations and five deaths (2, 3). In an unrelated outbreak, bagged baby spinach was recalled in 2010 due to STEC O157:H7 contamination (4). The Centers for Disease Control and Prevention reported that between 1973 and 2012, STEC was second only to norovirus in the number of outbreaks associated with leafy vegetables in the United States, of which 46 of the 49 STEC outbreaks were caused by a serogroup O157 strain (2). Although STEC did not cause the greatest number of outbreaks, it did lead to by far the greatest number of hospitalizations for a given microorganism, with 46% of the total hospitalizations due to all leafy-vegetable-associated outbreaks attributed to STEC. Illness due to ingestion of STEC-contaminated leafy greens led to nine deaths, including six from spinach (2). With a calculated infectious dose of only 10 to 100 CFU (5–7), the ability to detect very low levels of STEC contamination on fresh leafy vegetables is important.

Currently, the U.S. Food and Drug Administration's *Bacteriological Analytical Manual* (FDA BAM) protocol is used for the detection of STEC on leafy vegetables (8). In this procedure, the produce is enriched in medium for approximately 24 h and then screened by multiplex real-time PCR for the stx_1, stx_2, and *uidA* (+93 single nucleotide polymorphism [SNP] mutation indicative of O157 strains) genes. Unlike other foodborne pathogens such as *Salmonella* and *Listeria*, it is not sufficient to detect merely the presence of the bacterial species when determining whether *E. coli* contamination presents a risk to public health. Many *E. coli* strains are not pathogenic to humans, thus necessitating at least partial characterization of an *E. coli* strain associated with fresh produce to determine pathogenicity. In fact, highly virulent hybrid pathotype strains have been reported, for example, the *E. coli* O104:H4 STEC/enteroaggregative hybrid associated with sprouts that caused an outbreak with a high percentage of HUS in Germany in 2011 (9). A detection method that incorporates the ability to simultaneously determine the presence of multiple virulence genes would be ideal. If virulence characterization of a contaminating STEC strain is desired beyond the presence of stx_1 and/or stx_2 in the genome, the strain must be isolated and used in conventional PCR assays (as practiced in the BAM protocol) or sequenced. These are multistep processes that trade time for strain-level information.

Pyrosequencing of 16S rRNA genes has been utilized for direct detection of microorganisms and their proportions in microbial communities without the time-consuming step of culturing individual bacterial species for identification. This method of determining the composition of bacterial communities has been applied to surfaces of ready-to-eat vegetables, including spinach

Received 10 August 2015 **Accepted** 15 September 2015

Accepted manuscript posted online 18 September 2015

Citation Leonard SR, Mammel MK, Lacher DW, Elkins CA. 2015. Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing *Escherichia coli* on fresh bagged spinach. Appl Environ Microbiol 81:8183–8191. doi:10.1128/AEM.02601-15.

Editor: H. L. Drake

Address correspondence to Christopher A. Elkins, chris.elkins@fda.hhs.gov.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02601-15.

(10–13). Using such an approach to determine the presence of foodborne pathogens in the bacterial community associated with fresh bagged leafy greens may provide the ability to detect pathogens more quickly than current methods permit. However, high-throughput metagenomic shotgun sequencing has been used to investigate the taxonomic composition of microbial communities and can provide higher-resolution taxonomic information than 16S rRNA sequencing, as it makes use of either clade-specific marker genes (14) or matches to k-mers unique to a database entry (15, 16). The discriminative k-mer methods of metagenomic sequence data classification have increased the speed at which data can be processed, which is an important aspect of a pathogen detection method. While 16S rRNA sequencing would allow detection of a possibly contaminating *E. coli* strain, metagenomic shotgun sequencing might allow classification at a higher resolution, e.g., *E. coli* phylogroup, and only shotgun metagenomic sequencing, provided adequate sequencing depth was achieved, would provide the data for virulence gene characterization needed to deduce the potential pathogenicity of the strain.

The objective of this study was to explore the sensitivity of a metagenomic shotgun sequencing method of detecting contaminating STEC on bagged spinach. Specifically, we addressed the questions of how much contamination is necessary for detection without enriching the spinach sample and how long an enrichment process is necessary for detection of a very low level STEC contamination with deference to the established FDA BAM culture protocol (8). In addition, because we were able to determine the bacterial community associated with bagged spinach using this method, we focused on bacterial community changes with regard to enrichment time and changes to the enrichment process.

MATERIALS AND METHODS

Spiked *E. coli* strain. STEC strain Sakai (chromosome [BA000007.2], pOSAK1 [AB011548.2], pO157 [AB011549.2] [GenBank accession numbers shown in brackets]), used for spiking experiments, was grown overnight in modified buffered peptone water with pyruvate (mBPWp) medium. The overnight culture was diluted appropriately to add 274 μl containing the target number of spiked cells. Dilutions of the overnight culture were spread on Luria-Bertani plates to determine the level of artificial contamination of the spinach. STEC strain Sakai was chosen for spiking experiments, as a closed genome is available and this strain carries both stx_1 and stx_2 along with the other virulence factors utilized in this work. In addition, STEC Sakai is a serotype O157:H7 strain, the pathogenic STEC serotype causing the greatest number of outbreaks associated with fresh produce (2). The Sakai strain is associated with *E. coli* phylogroup E and is an O157:H7 clade 1 archetype (17).

Sample processing. Packages of ready-to-eat spinach (227 g) were purchased from a grocery store and stored at 4°C until use. Experiments were carried out using a slightly modified version of the BAM protocol (Fig. 1) (8) that included a greater volume of medium at the start of the experiment to allow for removal of aliquots during the course of the enrichment process. To 1,627-ml filter whirlpak bags containing 100 g spinach, 250 ml of mBPWp medium was added and then spiked with the appropriate CFU of STEC Sakai. The bags were manually massaged and then shaken at 37°C and 185 rpm for 30 min. For preenrichment samples, the bags were again manually massaged, and 20-ml aliquots were removed and pelleted. The bags were then incubated statically at 37°C for 5 h and manually massaged, and 5-ml aliquots were removed and pelleted. Three antimicrobial agents, acriflavin, cefsulodin, and vancomycin, were added to the remaining 225 ml medium in each bag to final concentrations of 10 mg/liter, 10 mg/liter, and 8 mg/liter, respectively. The bags were then incubated statically at 42°C for 18 h and manually massaged, and 1- or 2-ml aliquots were removed and pelleted. For samples in which aliquots were not taken at the preenrichment time point or the 5-h time point, the starting volume of the medium was adjusted appropriately. For samples taken between 5 and 23 h of enrichment, 2-ml aliquots were removed and pelleted. All pellets were stored at −20°C prior to genomic DNA (gDNA) extraction.

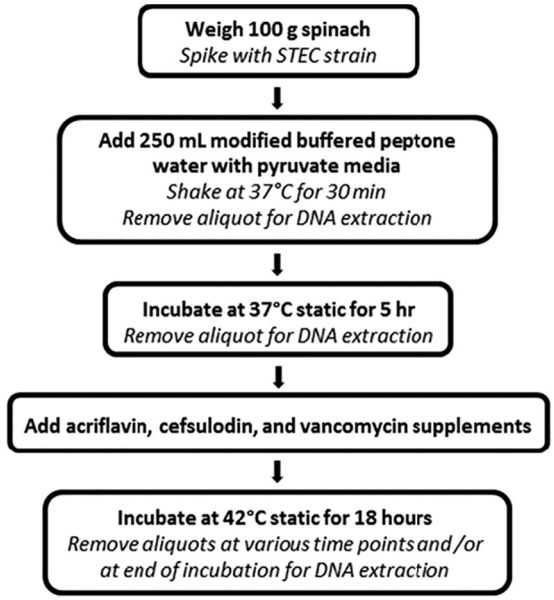

FIG 1 Flow diagram of the U.S. Food and Drug Administration's *Bacteriological Analytical Manual* enrichment protocol for detection of Shiga toxin-producing *E. coli* on leafy vegetables. Procedural steps shown in italics are additions to the protocol used in this work. The initial volume of medium was modified as needed to allow for removal of aliquots during the enrichment process. Preenrichment samples were removed after the 30-min shaking step.

DNA isolation, sequencing, and mapping. Genomic DNA was extracted using the DNeasy blood and tissue kit (Qiagen, Germantown, MD, USA) following the protocol for Gram-positive bacteria to ensure the inclusion of both Gram-positive and Gram-negative bacteria present on the spinach samples. Sequencing libraries were prepared with the Nextera DNA sample preparation kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina MiSeq platform, generating paired-end 250-bp reads. The number of reads per sample averaged 11,596,284 ± 3,571,586. Raw reads were trimmed and mapped to the STEC Sakai genome using CLC Genomics Workbench v7.5.1, v8.0, or v8.0.1 (CLC bio, Boston, MA, USA). Mapping parameters used were 98% identity for 100% of the trimmed read length. The same mapping parameters were used for reference-based assembly. The coverages of the chromosome and two plasmids harbored by the STEC Sakai strain were determined from the mappings using CLC Genomics Workbench. For selected samples, reads were also mapped to the *Spinacia oleracea* cultivar SynViroflay genome (GenBank accession no. AYZV01). In this case, the mapping parameters used were 80% identity for 80% of the read length.

Determination of bacterial community. Metagenomic analysis for microbial composition was conducted using custom C++ programs developed to compile a k-mer signature database containing multiple unique 25-bp sequences per target entry (bacterial species, phylogroup, or serovar) and then identify bacterial species in a sample by the unique 25-mers using the trimmed sequencing reads as input. Determination of database entries was based on previous work (18), BLAST results using some sequence reads generated in this work as queries, and results from several of the samples utilizing MetaPhlAn version 2.0.0 (14). For each species of interest, each nonduplicated 25-mer from a reference whole-genome sequence was placed into a database. We removed any k-mers not found in at least two-thirds of a set of additional genome sequences of the

same species, and we removed any k-mer found in genomes of other species. The resulting k-mer database used in this work contains 155 target entries, each consisting of approximately 40,000 (range, 967 to 155,819) unique k-mers. The coverage of each species' 25-mer set was determined by testing each possible 100-bp read in a reference genome against the 25-mer set and tallying the number of reads in the genome that are matched by a 25-mer. The database includes 52 different bacterial genera and can differentiate between *E. coli* phylogroups A, B1, B2, D, and E. An attempt to identify each read from the sequencing run was made by matches to the 25-mer database. Normalization is performed to correct for bias due to different numbers of k-mers used per database entry, and the results are tabulated as the percentage of identified reads (contribution to the microbial population of identified species) for each database entry. For those samples for which MetaPhlAn was utilized in determining additional bacterial genera/species to add to our database, the results from our custom programs for the percentages of bacterial genera/species in the microbial community were in agreement with those obtained from MetaPhlAn. However, we were able to perform classification at a higher level for some genera, and particularly important, within the *E. coli* species.

Molecular serotyping and virulence gene detection. The numbers of trimmed reads mapping to the stx_{1a}, stx_{2a}, γ-*eae*, and *ehxA* virulence genes, as well as the *wzm*, *wzt*, *wzx*, *wzy*, and *fliC* molecular serotyping loci were determined using SeqMan NGen v12.2.0 using the default reference-based assembly parameters for Illumina reads (DNASTAR Inc., Madison, WI, USA). The two-tailed *t* test was used to calculate *P* values.

Metagenomic data accession number. Metagenomic sequence data from this study were deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP062607.

RESULTS

Bacterial community associated with bagged spinach. The metagenomic sequencing data were analyzed in three ways, namely, by mapping the reads to the Sakai genome, determining the ability to serotype and discover the virulence profile of the spiked strain from the data, and determining the microbial community using a k-mer signature database, including detection of the spiked STEC (as percent *E. coli* phylogroup E) within the community analysis. For unspiked preenriched spinach samples, the background values for percent reads mapping to the Sakai genome and percentage of bacterial community as determined from the k-mer analysis were 0.03% and 0.004%, respectively. Metagenomic data analysis was performed on unspiked spinach samples at preenrichment, 5-h, and 23-h time points to determine the native microbial community associated with bagged spinach and how the composition changes throughout the enrichment process. Spiked samples that were collected preenrichment also revealed information about the indigenous microbial community. Considering a contribution of greater than 0.1% to the identified bacterial population, analysis on preenrichment samples indicated 16 bacterial genera associated with the prepackaged spinach, including six different species in the genera *Pseudomonas* (*P. aeruginosa*, *P. alcaligenes*, *P. fluorescens*, *P. mendocina*, *P. putida*, and *P. stutzeri*) and three *Exiguobacterium* species (*E. sibiricum*, *E. antarcticum*, and *E. undae*). Most of the genera displayed low abundance, as only *Pseudomonas* spp., *Pantoea agglomerans*, *Exiguobacterium* spp., and *Sutterella wadsworthensi* were present at greater than 1% of the identified community in at least one sample. Preenrichment, the predominant bacterial species was *P. fluorescens* (more than 60% of bacterial population), with either *P. agglomerans* or *E. sibiricum* the second most abundant species. However, the microbial composition was observed to change with enrichment time so that after 5 h of enrichment, *P. agglomerans* became the predominant species in some samples, while *Exiguobacterium* spp. became the predominant species in others, *Aeromonas* spp. grew in abundance in some samples, and the percent *P. fluorescens* contributing to the bacterial community was greatly diminished. After 23 h of enrichment, more than 75% of the reads associated with identified species were attributed to *Enterobacter cloacae*, *P. agglomerans* had a lower percent abundance than at preenrichment, and the contribution of *Pseudomonas* spp. was less than 1%. *E. coli* phylogroup B2 was discovered in one bagged spinach sample at a low level of abundance, 0.109% of the population of identified species. Molecular serotyping identified *E. coli* O45 and O101 loci in this sample; therefore, there were two different strains in this sample. These strains did not possess the common STEC virulence factors *stx*, *eae*, and *ehxA*. In addition to the metagenomic community analysis, the reads from unspiked samples were mapped to the spinach genome. The results reveal that 41%, 6.3%, and 1.7% of the total reads could be attributed to spinach DNA for the samples at preenrichment, 5 h, and 23 h of enrichment, respectively.

TABLE 1 Results of spiking spinach with STEC Sakai[a]

Sampling time and spike (CFU)	*E. coli* phylogroup E (%)[b]	Mapped reads (%)[c]	Coverage (fold)[d]
Preenrichment			
10	0.003	0.01	0.04
1,000	0.003	0.05	0.11
100,000	0.119	0.03	0.10
1,000,000	0.356	0.13	0.41
10,000,000[e]	4.251	0.84	3.16
5-h enrichment			
10[f]	0.008	0.08	0.24
1,000[f]	0.066	0.12	0.39
10,000[e]	7.418	3.30	13.3
100,000	30.449	17.09	70.9
23-h enrichment			
10[f]	66.954	49.73	184
1,000[f]	84.330	61.77	252

[a] Samples of 100 g spinach were spiked with STEC Sakai.
[b] Percentage of the bacterial population belonging to *E. coli* phylogroup E.
[c] Percentage of total reads mapping to the STEC Sakai genome.
[d] Average chromosome coverage. The values were normalized to 10,000,000 total reads per sample. The coverage for the two plasmids was greater than that of the chromosome for all samples.
[e] Average of triplicate biological samples.
[f] Average of duplicate biological samples.

Detection limit for preenriched samples. Spinach samples were spiked with STEC Sakai at levels ranging from 10 to 10,000,000 CFU/100 g spinach and processed preenrichment (Fig. 1). At the lower levels of contamination, 10 and 1,000 CFU, the spiked Sakai strain could not be detected (Tables 1 and 2). At the 10 and 1,000 CFU spike levels, the few reads mapping to the Sakai genome were at the rRNA-encoding loci, thus possibly due to very small amounts of an *E. coli* strain in the native bacterial community. At the 100,000 CFU spike level, the spiked strain was detected in the bacterial community analysis, and the reads mapping to the Sakai genome were more dispersed over the length of the genome, although the percent mapping was not above the background level (0.03%). The Sakai strain could be detected with greater confidence at the 1,000,000 CFU spike level;

TABLE 2 Number of reads mapping to virulence genes and serotyping loci for spinach samples spiked with STEC Sakai

Sampling time and spike (CFU)	No. of reads mapping to the following virulence gene or serotyping locus[a]:						
	stx$_{1a}$	*stx*$_{2a}$	γ-*eae*	*ehxA*	O157 *wzx*	O157 *wzy*	H7 *fliC*
Preenrichment							
10	0	0	0	0	0	0	0
1,000	0	0	0	0	0	0	0
100,000	1	1	1	1	0	0	0
1,000,000	2	0	7	4	3	0	2
10,000,000[b]	21	17	79	44	15	12	12
5-h enrichment							
10	0	0	0	0	0	1	0
1,000	0	2	4	5	0	0	0
10,000[b]	57	65	329	231	36	22	43
100,000	373	275	1,676	1,382	194	126	232
23-h enrichment							
10[c]	1,225	1,361	3,467	4,013	988	850	978
1,000	1,715	1,647	4,773	4,529	1,419	1,110	1,298

[a] Number of reads mapping to virulence genes and serotyping loci for 100-g spinach samples spiked with STEC Sakai. For comparison, the numbers are normalized to 10,000,000 total reads per sample.
[b] Average of triplicate biological samples.
[c] Average of duplicate biological samples.

however, the virulence profile could not be reliably determined, as no reads mapped to the *stx*$_{2a}$ gene or one of the loci used for serotyping (Table 2). The Sakai strain was easily detected, and the serotype and virulence profile were determined at a spike level of 10,000,000 CFU. Some variation in detection sensitivity was observed between different bags of spinach. For example, the number of reads mapping to the *stx*$_{2a}$ gene varied between 6 and 26 reads among the three triplicate samples. Although the bacterial community found on the bagged spinach preenrichment varied in the percentages of minor species found, it was so dominated by *P. fluorescens* that biological replicates taken from different bags appeared quite similar (Fig. 2). Minor quantities, between 0.1 and 1.0%, of other species contributing to the overall community in at least one of the samples included *Enterococcus faecalis*, *Lactobacillus sanfranciscensis*, *Listeria monocytogenes*, *Salinibacter ruber*, *Alistipes* spp., *Clostridium botulinum* C, *Lactococcus garvieae*, *Akkermansia muciniphila*, *Stenotrophomonas maltophilia*, and *Edwardsiella tarda* (see Table S1 in the supplemental material).

Detection limit for samples enriched for 5 h. Spinach samples were spiked with STEC Sakai at levels ranging from 10 to 100,000 CFU/100 g spinach and processed at 5-h enrichment time (Fig. 1). Analysis of the metagenomic data revealed that the spiked strain could be detected in the community analysis at 1,000 CFU, and although many reads mapping to the Sakai genome were located at rRNA-encoding loci, there were also reads distributed over the length of the genome. However, information about serotype and full virulence profile could not be determined at this spike level (Table 2). At a spike level of 10,000 CFU, the metagenomic data contained sufficient reads associated with the Sakai strain to fully characterize the strain. In fact, an average chromosome coverage of 13× was obtained (Table 1). As was observed for the preenrichment samples, there was variable detection sensitivity between spinach bags. The number of reads mapping to the *stx*$_{2a}$ gene ranged from 22 to 128 for a spike level of 10,000 CFU. At a contamination level of 100,000 CFU/100 g spinach, significant chromosomal coverage was obtained from the metagenomic data, and the consensus sequence of the Sakai genome from a reference-based assembly contained only seven gaps totaling 265 bp. There was some variation between biological replicates in the composition of the bacterial community found in the samples after 5-h enrichment time and in the percent contribution from the spiked

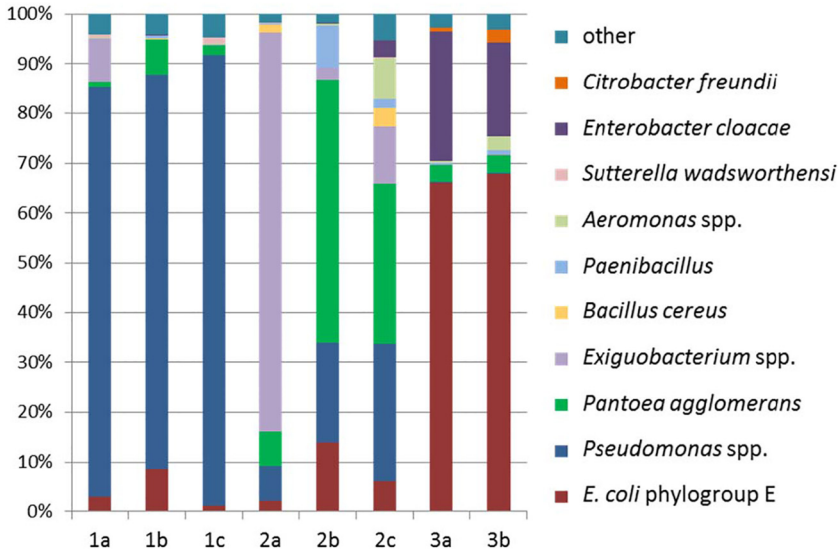

FIG 2 Relative abundance of bacterial species associated with spinach samples spiked with STEC Sakai. Samples of 100 g spinach from different bags were spiked with 10,000,000 CFU and processed preenrichment (bars 1a to 1c), 10,000 CFU and enriched 5 h (bars 2a to 2c), or 10 CFU and enriched 23 h (bars 3a and 3b). Sequencing reads from metagenomic samples were analyzed using a custom k-mer signature database. Bacterial genera or species contributing 1% or greater to the total identified population are shown, and the sum of all other species identified in lower amounts is included in "other." For species with minor abundances (between 0.1 and 1.0%), see Tables S1, S2, and S3 in the supplemental material.

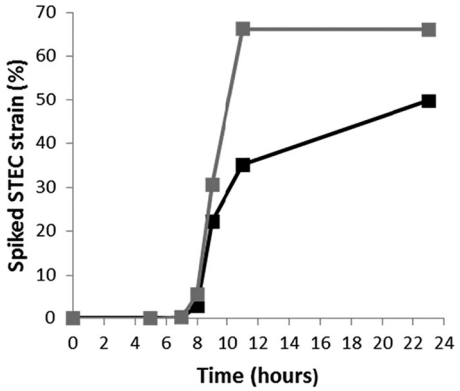

FIG 3 Detection of spiked STEC Sakai at different total enrichment times. Spinach samples were spiked with 10 CFU/100 g spinach and processed for the indicated enrichment time. The results are plotted as percentage of total reads mapping to the STEC Sakai genome (black) and percentage of identified bacterial population belonging to *E. coli* phylogroup E (gray).

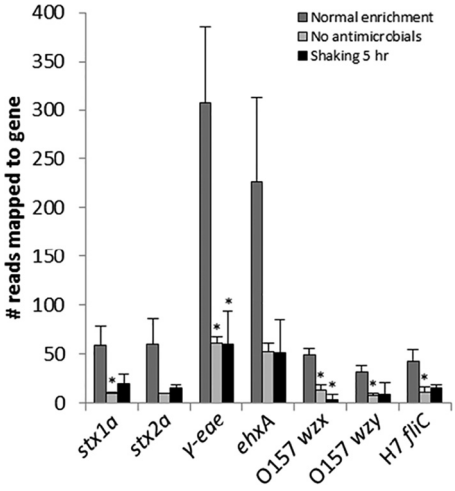

FIG 4 Change in detection sensitivity with changes in enrichment processing. Spinach samples were spiked with 10 CFU/100 g spinach and processed with the indicated changes. Samples were taken after an 8-h enrichment time, and the number of reads mapping was normalized to 10,000,000 total reads per sample. The results obtained from samples where the enrichment process was changed (duplicate biological samples) were compared to those from the normal process (triplicate biological samples). The values that were statistically significantly different ($P < 0.05$ by the two-tailed t test) from the value for the normal enrichment process are indicated by an asterisk.

Sakai strain (Fig. 2). In addition, the metagenomic data revealed greater variation in bacterial composition in the samples enriched for 5 h in comparison to preenrichment. Along with the bacterial species included in Fig. 2, one of the samples (Fig. 2, column 2c) was found to contain 1.8% *Yersinia enterocolitica* contributing to the sample community. Although each contributed between only 0.1 and 1.0% to the overall population, other bacterial species, *Bacillus* spp. (other than *B. cereus*), *Alistipes* spp., *L. garvieae*, *Leuconostoc lactis*, and *E. tarda* were present in at least one sample (see Table S2 in the supplemental material).

Detection limit for samples enriched for 23 h. Metagenomic sequencing data were obtained from spiked spinach samples that were incubated for an additional 18 h at 42°C after the 5-h enrichment at 37°C for a total of 23 h (Fig. 1). Samples were prepared at STEC Sakai contamination levels of 10 and 1,000 CFU/100 g spinach. After 23-h enrichment, the spiked strain constitutes approximately half the bacterial community for a spike level of 10 CFU, resulting in a chromosomal coverage of 184× (Table 1). The metagenomic data contains a far greater number of reads than are necessary to determine the molecular serotype and virulence profile with confidence (Table 2). The consensus sequence of a reference-based assembly of the STEC Sakai genome was determined to contain only six gaps totaling 200 bp. A large shift in bacterial composition of the sample was observed, resulting in *E. coli* phylogroup E (spiked strain) and *E. cloacae* as the predominant species (Fig. 2). One of the samples (Fig. 2, column 3b) was found to contain 1.1% *Proteus mirabilis* contributing to the sample community along with the bacterial species included in Fig. 2. Other bacterial species, contributing between only 0.1 and 1.0% to the overall population, were present in at least one sample including *Cronobacter* spp., *E. tarda*, *Klebsiella oxytoca*, *E. coli* phylogroup B1 (serogroup determined to be O101), and *Salmonella enterica* serovar Heidelberg (see Table S3 in the supplemental material).

Determination of the shortest enrichment time to detect the lowest limit of contamination. Clearly, 5-h enrichment time is not long enough to detect the STEC Sakai strain at the low spike level of 10 CFU/100 g spinach, but the full subsequent 18-h incubation at 42°C is more than sufficient time. We sought to determine the minimum enrichment time necessary to detect 10 CFU. Sequencing samples were prepared at intermediate enrichment times, namely, 2, 3, 4, and 6 h of enrichment at 42°C after the 5-h incubation at 37°C. The spiked STEC Sakai strain appears to be transitioning between lag and exponential-phase growth at 7-h enrichment, at which point 0.22% of the reads map to the Sakai genome, and 0.322% of the bacterial community is attributed to *E. coli* phylogroup E (Fig. 3). The serotype and virulence profile cannot be reliably determined at 7-h enrichment, but by 9 h, there are hundreds of reads mapping to each virulence gene, and 22% of the metagenomic sample reads map to the Sakai genome. Intermediate to these time points, 8 h of enrichment yields an average of 2.85% reads mapping to the Sakai genome and average genome coverage of 10.5×. Furthermore, 5.647% of the bacterial community consists of *E. coli* phylogroup E, and the spiked strain is present in sufficient quantity to reliably determine the molecular serotype and virulence profile (Fig. 3 and 4). Enrichment beyond 11 h does not increase the ability to detect the spiked strain appreciably or to gain additional information about the strain, as it appears to enter stationary growth phase (Fig. 3). Although the quantitative results show disparity between the two methods of determining the contribution of STEC Sakai to the total sample, especially at 11 and 23 h, the time courses are consistent with each other. The bacterial population in the samples enriched for 8 h exhibited minimal variation and was dominated by *P. agglomerans* and *Exiguobacterium* spp. (Fig. 5). Although each contributed between only 0.1 and 1.0% to the overall population, other bacterial species were present in at least one sample; these species included *Bacillus coagulans*, *L. garvieae*, *E. cloacae*, *E. tarda*, *S. wadsworthensi*, and *Citrobacter freundii* (see Table S4 in the supplemental material).

Effects of changes in the enrichment process. After determining that 8 h is the shortest enrichment time for detection of STEC Sakai at a spike level of 10 CFU/100 g spinach, we sought to investigate how changes in the enrichment process would affect the detection sensitivity. To determine the effect of shaking, samples

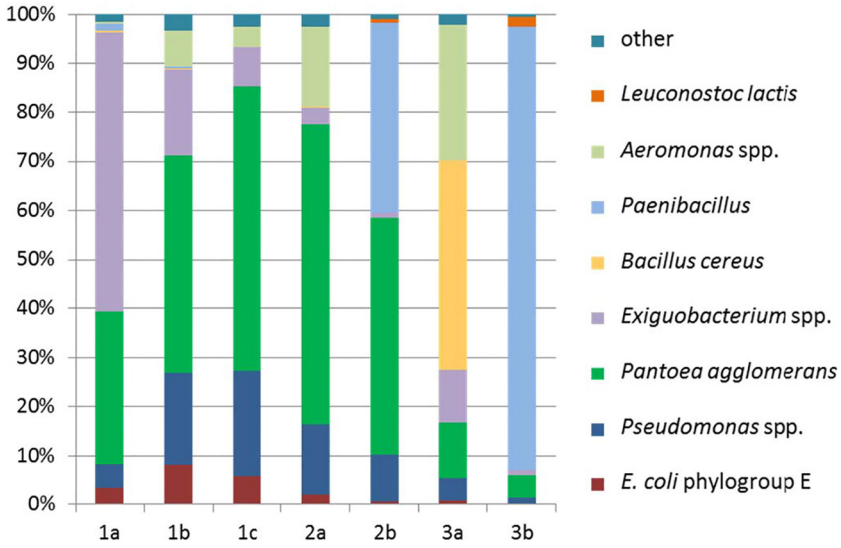

FIG 5 Relative abundance of bacterial species associated with spinach samples spiked with STEC Sakai. Samples of 100 g spinach were spiked with 10 CFU and enriched for 8 h. Samples were either enriched according to the described normal process (bars 1a to 1c), shaken for the first 5 h (bars 2a and 2b), or grown without the addition of antimicrobials after the first 5 h (bars 3a and 3b). Sequencing reads from metagenomic samples were analyzed using a custom k-mer signature database. For comparison, samples shown by bars 1c and 2a were prepared from the same bag of spinach, as were samples shown by bars 1b and 3a. Bacterial genera or species contributing 1% or greater to the total identified population are shown, and the sum of all other species identified in lower amounts is included in "other." For species with minor abundances (between 0.1 and 1.0%), see Table S4 in the supplemental material.

were shaken for 5 h at 37°C before increasing to 42°C for the remaining 3 h instead of incubating samples statically for 5 h at 37°C (Fig. 1). Shaking the sample had an effect on the proportional growth of the STEC Sakai strain compared to other bacterial species. While an average of 2.85% of the reads mapped to the Sakai genome from metagenomic samples that were incubated statically for 8 h, only 0.75%, significantly less ($P = 0.042$), mapped from samples that were shaken for the first 5 h of the 8-h enrichment process. Although the serotype and virulence profile could be determined, there were much fewer reads associated with the relevant genes (Fig. 4). For direct comparison, a pair of samples, one incubated statically and the other with shaking, were processed from the same bag of spinach (Fig. 5, bars 1c and 2a). The bacterial community analysis resulting from the metagenomic data suggests that *Aeromonas* spp. may grow preferentially at the expense of the spiked STEC strain when the sample is shaken. In addition, shaking appears to favor the growth of *Paenibacillus*. Other changes in population abundance were minor (see Table S4 in the supplemental material).

Another variation to the protocol investigated was the addition of antimicrobial agents after the first 5 h of incubation (Fig. 1). The bacterial community of a sample enriched for the entire 23 h, but without the addition of the antimicrobials, was determined to be comprised of 63% *E. coli* phylogroup E (see Table S3 in the supplemental material) compared to 67% processed with antimicrobials (Table 1); thus, the spiked STEC Sakai could be as easily detected and genomically characterized without the addition of antimicrobials as with them. In addition, this sample was determined to have 10% and 6.9% of the bacterial population comprised of *E. coli* phylogroup B1 (serogroup nontypeable) and *Klebsiella pneumoniae*, respectively (Table S3). Other *E. coli* phylogroups were detected in lower amounts, namely, *E. coli* phylogroups A (0.383%, serogroup O9) and D (0.102%). We sought to determine whether it was advantageous to add the antimicrobials if the enrichment time was only 8 h. The results indicate that the addition of the antimicrobials greatly affects the detection sensitivity at spike levels of 10 CFU and enrichment time of 8 h (Fig. 4 and 5). Only an average of 0.53% of the reads from the metagenomic samples prepared without addition of antimicrobials mapped to the Sakai genome, a statistically significant difference ($P = 0.030$) compared to 2.85% when antimicrobials were added. Relevant genes for serotyping and virulence profile determination had sufficient coverage; however, the coverage was significantly reduced (Fig. 4). A pair of samples was processed from the same bag of spinach for direct comparison of the associated bacterial community (Fig. 5, bars 1b and 3a). The results indicate a dramatic shift in bacterial population, with *P. agglomerans*, *Exiguobacterium* spp., and *Pseudomonas* spp. crowded out by *B. cereus* and *Aeromonas* spp. Another sample for which antimicrobials were not added is almost entirely comprised of *Paenibacillus*. *K. pneumoniae* or other *E. coli* phylogroups were not detected in the samples enriched for 8 h without antimicrobials. Other minor changes in population abundance were detected (see Table S4 in the supplemental material).

DISCUSSION

Variables associated with defining STEC detection limit. Although microbial community analysis of metagenomic samples using 16S rRNA sequencing has been applied to the study of bacterial communities associated with fresh vegetables, including spinach (10-13, 19), our study was aimed at expanding this concept to determining the sensitivity of metagenomic shotgun sequencing for detection of a pathogen contaminant associated with the spinach. We were able to define limits for detection of STEC on fresh bagged spinach at preenrichment and at 5-h and 23-h time points during the enrichment process following the BAM protocol (8) and a sequencing depth of 10,000,000 reads per sample (Tables 1 and 2). Our results demonstrate that the level of

contaminating STEC necessary for detection preenrichment is 10,000,000 CFU/100 g spinach. This suggests that the overall microbial load associated with the ready-to-eat bagged spinach is too great for utilizing a metagenomic approach without some enrichment time to select for the contaminating STEC. In one study, plate counts from conventionally grown ready-to-eat packaged baby spinach determined the viable counts of culturable bacteria to be 10^8 to 10^9 CFU/g spinach (12), while in another study, total bacterial counts were found to vary widely from $10^{3.9}$ to $10^{8.2}$ CFU/g (20). In yet another study, Babic et al. found the initial populations of mesophilic and psychrotrophic aerobic microflora were each 10^6 to 10^7 CFU/g spinach but increased to 10^{10} CFU/g each during 8 days of storage at 10°C (21). Not only does this high indigenous bacterial load present difficulty in detecting very low levels of STEC contamination, but for preenriched samples, we determined that approximately 41% of the total metagenomic sample sequencing reads map to the spinach genome, thus necessitating even deeper sequencing to detect low levels of a contaminating pathogen. Similar to preenriched samples, our results indicate that the metagenomic method would be impractical at 5-h enrichment without deeper sequencing. However, following the complete 23-h enrichment process described in the BAM protocol (8), we were able to demonstrate that a low level of STEC contamination, 10 CFU/100 g spinach, could be detected. In fact, a reference-based assembly produced an almost complete genome, thus allowing full characterization of the contaminating STEC strain. Our results show that after 11-h enrichment, the STEC had completed the exponential growth phase, and therefore, further enrichment time is of no advantage (Fig. 3). By 8 h, the STEC strain could be detected, and the virulence profile could be determined (Fig. 4). In fact, so many reads mapped to the virulence genes at a sequencing depth of 10,000,000 total reads that it would be possible to detect and characterize the STEC without as great a sequencing depth, thus allowing for more samples to be processed in a single sequencing run. Alternatively, if time were not a factor, samples could be enriched longer than 8 h, and it would be possible to sequence an even greater number of samples contaminated with as little as 10 CFU/100 g spinach and still gain a significant amount of strain-level information. Our results illustrate the interplay between time, cost, and strain-level information gained when utilizing a metagenomic approach to pathogen detection. The limit of detection for a given enrichment time and number of total reads is dependent on the goal, i.e., whether mere knowledge of the presence of STEC is satisfactory or an assembly is required.

Assembly of the STEC genome from metagenomic sequencing data. The determination that a bag of spinach is contaminated with STEC does not require a draft whole-genome sequence of the contaminating strain; however, a genome sequence would provide essential information during an outbreak or traceback analysis. We performed reference-based assemblies using the metagenomic sequencing reads from several of our spiked samples. Our data suggest that the contamination level would need to be a very high value, 100,000,000 CFU/100 g spinach, to achieve ~30× coverage of the spiked strain in preenriched samples. However, we were able to attain genome consensus sequences averaging only six gaps totaling 200 bp for samples spiked with 10 CFU/100 g spinach and enriched for 23 h. As with virulence gene detection discussed above, our results demonstrate that if samples spiked with 10 CFU/100 g spinach were enriched for 23 h, adequate genome coverage could be obtained for assembly with less than 10,000,000 reads per sample. Eliminating the need to first isolate a contaminating STEC from a sample enriched for 23 h by plating on selective medium and incubating overnight, followed by growth of a single colony isolate overnight for genome sequencing, can shorten the time necessary for obtaining substantial strain-level characterization by 2 days. At 8-h enrichment time, determined to be the shortest time to obtain the serotype and virulence profile of the spiked strain at an initial level of 10 CFU/100 g spinach, the average genome coverage was determined to be 10.5×. This result indicates that for 8-h enrichment, a contamination level of ~30 CFU/100 g spinach would be sufficient to yield a genome sequence with ~30× coverage. Because the spiked strain is just entering log phase growth at 8-h enrichment (Fig. 3), adding even 30 min to the enrichment time would make a significant difference as to whether a draft genome sequence of sufficient quality could be extracted from the metagenomic data for these very low contamination levels. Since the genome sequence of the contaminating STEC strain would be unknown in a produce screening or outbreak analysis situation, ultimately, it would be desirable to assemble the genome *de novo*, rather than as a reference-based assembly, as was possible in the present study. Extracting reads for the *de novo* assembly of a single genome from metagenomic sample reads is challenging, but it has been attempted with some success from a fecal sample containing STEC from an outbreak (22). These investigators used a method whereby they subtracted sequencing reads identified as either human or the microbial community found in healthy human fecal samples. Our results reveal that the percentage of total metagenomic sequencing reads matching the spinach genome declines quickly, from 41% to 6.3%, during the first 5 h of enrichment, thus for shorter enrichment times, subtracting the reads identified as matching the spinach genome could be useful. We have identified the bacterial species most often found associated with fresh bagged spinach at different time points throughout the enrichment process, and this information may prove valuable for future *de novo* assemblies if a subtractive approach is used to extract reads associated with the contaminating STEC strain.

***E. coli* virulence gene identification and typing.** We detected native *E. coli* associated with the bagged spinach in three of the samples that had been enriched for 23 h. The fact that native *E. coli* was detected only after 23-h enrichment suggests that the initial number of *E. coli* bacterial cells on the bagged spinach was very low. The genomes of the *E. coli* O45 and O101 strains identified in an unspiked sample did not carry Shiga toxin genes or other common STEC virulence factors. The other indigenous *E. coli* strains identified in spiked samples belonged to *E. coli* phylogroups B1 (serogroups O101 and nontypeable), A (serogroup O9), and D (insufficient reads to determine the serogroup). Because these *E. coli* strains were found in spiked samples and the quantity was low, it cannot be definitively determined whether they carry the STEC virulence factors included in this study. The discovery of nonpathogenic *E. coli* associated with one or some of the spinach samples reinforces the need to identify virulence factors. This necessitates a greater number of reads mapping to the *E. coli* genome than would be necessary for detection relying solely on 16S rRNA sequencing of metagenomic samples. In the present investigation, only one STEC strain was used for spiking; however, we were able to determine the molecular serotype and identify the intimin gene allele type along with both Shiga toxin gene allele types carried by this strain. Determination of gene allele types and serotype is im-

portant, as they can be used to aid in determining the risk a particular contaminating strain poses to human health. Other virulence factors known to contribute to STEC pathogenicity could be included, and although we focused on STEC in this work, molecular markers for all of the *E. coli* pathotypes (1) could easily be incorporated into the analysis of the metagenomic data.

Indigenous microbial community associated with bagged spinach. Our study revealed that considering bacterial genera or species that contribute at least 1% to the population, the diversity of the bacterial community associated with unenriched fresh bagged spinach is low and that *Pseudomonas* spp., and in particular *P. fluorescens*, is the dominant bacterial species (Fig. 2). By pyrosequencing of the 16S rRNA gene, Jackson et al. also found that *Pseudomonas* spp. dominated (93.8%) the bacterial community associated with conventionally grown bagged baby spinach (12). Additionally, that study reported a lower operational taxonomic unit richness for baby spinach than for four types of leaf lettuce. Another study also reported that the most numerous 16S rRNA sequences associated with fresh washed spinach were identified as *Pseudomonas* spp. (9.2%); however, 13.7% of the sample was classified in the family *Enterobacteriaceae* (13). In contrast to our results, Leff and Fierer reported that although the microbial community on prepackaged spinach was dominated by *Gammaproteobacteria*, more than 50% of the sample consisted of bacteria belonging to the family *Enterobacteriaceae* rather than *Pseudomonadaceae* (4%) (10). In that study, *Pantoea* sp. was found with a relative abundance of 32.4%, and a low abundance of *Paenibacillaceae* and *Exiguobacteraceae* was discovered. Although the relative abundances are different, this is consistent with several of the bacterial genera we found in our work. Previous to the widespread use of molecular methods for determining bacterial species associated with fresh packaged spinach, Babic et al. used plating on media to differentiate and enumerate bacterial species (21). These investigators report that *P. fluorescens* was the most abundant species and that *Pseudomonadaceae* were found at numbers of approximately 2 log units greater than those of *Enterobacteriaceae*. In general, our community analysis results obtained from metagenomic shotgun sequencing analysis using discriminative k-mers are in agreement with microbiological and 16S rRNA sequencing methods as to the major genera associated with ready-to-eat spinach. However, we were able to obtain higher taxonomic resolution by differentiating presence and abundance contributions to the species level, and we were also able to identify low-abundance species. Lopez-Velasco et al. report changes in the spinach microbiome with storage, even at 4°C (13), Babic et al. report growth of both *Pseudomonadaceae* and *Enterobacteriaceae* with storage at 10°C (21), and others report changes in the microbial communities in ready-to-eat packaged vegetable salads with storage (19). Storage time and conditions may be factors that explain some relative abundance differences between studies, as well as season and type of irrigation as was reported for lettuce microbiomes (11).

Shifts in the microbial community affect STEC detection sensitivity. Enriching the spinach samples shifts the population in the microbial community so that initially low-abundance bacterial species may become more abundant and small differences in initial abundance may become magnified as exemplified by our results (Fig. 2). Along with some variation in abundance of species contributing at least 1% of the identified sample population between different bags of spinach, there was further minor variation revealed in the presence/absence of very low abundance bacterial species as noted by our results for bacterial species contributing between 0.1 and 1.0% to the total population. Clearly, 5-h enrichment at 37°C causes a dramatic shift in the relative abundances within the bacterial community (Fig. 2). The growth of some genera or species is favored over others, and there is greater variation in the abundances between biological replicate samples. It is expected that a complicated interplay between the presence and abundance of a particular member of the bacterial population in the sample and the other members will influence the resulting community at any given time during enrichment. The spiked STEC Sakai strain, as a member of the bacterial community, is also subject to the influence of the growth pressures from competing bacterial species. During the first 5 h, the spiked STEC is growing very slowly (Fig. 3), while it appears *Pantoea agglomerans* and *Exiguobacterium* spp. are multiplying at a higher rate and/or *Pseudomonas* spp. are unable to survive (Fig. 2). Despite the different bacterial population compositions of samples spiked with 10,000 CFU/100 g spinach and enriched for 5 h, the spiked STEC strain could be detected and characterized for serotype and virulence factors (Fig. 2 and Table 2). Not surprisingly, given the similar proportion of STEC Sakai to other bacteria in the samples and the fact that the STEC strain is just transitioning to exponential growth phase at 8 h, a comparison of the compositions of the bacterial communities in spinach samples spiked with 10,000 CFU and enriched for 5 h and that of samples spiked with 10 CFU and enriched for 8 h reveals considerable similarity (Fig. 2 and 5). However, by the end of 23-h enrichment, even at a low spike level of 10 CFU/100 g spinach, the STEC Sakai strain was able to outcompete other bacteria in the microbial community (Fig. 2). This may be partly due to the fact that the 18-h incubation is performed at 42°C, since some bacterial species, *P. fluorescens* included (23), do not grow as well as STEC does at this temperature.

Along with growth competition due to other bacterial species in the sample community, the growth of the STEC strain is affected intrinsically by the enrichment procedure. Baylis investigated the growth of 20 STEC strains under various conditions and reported that the growth of the STEC strains during enrichment was at least to some extent strain dependent and depends on a combination of factors, including temperature, media, and the concentration and combination of antibiotics (24). Ideally, these conditions would be optimized for the selective growth of STEC compared to the rest of the sample community. One change to the protocol we investigated in this work was the lack of addition of antimicrobials. Since a sample spiked with 10 CFU/100 g spinach and enriched for 23 h without the addition of antimicrobials was comprised of 63% *E. coli* phylogroup E, similar to the samples processed with antimicrobials (67%), the STEC Sakai strain appears to be able to outcompete other members of the community even without the aid of the antimicrobials if given enough time. Although the lack of addition of antimicrobials during sample processing did not appear to affect the ability to detect and characterize the STEC Sakai strain after 23-h enrichment, the lack of antimicrobials did greatly affect the ability to detect and characterize the spiked strain after only 8 h of enrichment (Fig. 4 and 5). Our results indicate that the antimicrobials inhibit the growth of the Gram-positive bacteria *Bacillus cereus* and *Paenibacillus*. Another change to the protocol that we tested, since *E. coli* grows well when shaken, was to shake the spiked samples rather than incubating statically for the first 5 h. Interestingly, this change led to a

reduction in the percentage of the bacterial community identified as the spiked STEC strain when samples were taken at a total enrichment time of 8 h (Fig. 4 and 5). It appears that shaking favors the growth of *Aeromonas* spp. at the expense of the STEC strain. This exploratory work will be extended in future studies to investigate enrichment medium composition and incubation times and temperatures to ascertain the effect on the composition of the microbial community and the STEC detection level.

In conclusion, our work highlights the potential for utilizing a metagenomic approach to detect STEC on fresh produce, thereby eliminating the need to first isolate and culture a contaminating STEC strain in order to detect and characterize it. Assessment of the detection sensitivity using this approach has shown that strain-level characterization can be achieved even for very low levels (10 CFU/100 g spinach) of STEC contamination with less than the approximately 24-h enrichment currently used in the FDA BAM protocol (8). Along with determining STEC detection limits following the BAM protocol for enrichment, our metagenomic sequencing study was aimed at examining how the microbial community associated with spinach changes with modifications to the enrichment process, and we were able to determine this to the species level. We anticipate that this initial work will aid future work on optimizing the enrichment medium and/or protocol to enhance detection of STEC with a reduction in enrichment time. Knowledge gained in the present study about the bacterial community associated with spinach and how the population shifts during the BAM protocol will be a useful guide. In addition, this metagenomic shotgun sequencing approach can be applied to other foodborne pathogens.

ACKNOWLEDGMENT

This research was funded in part by the Oak Ridge Institute for Science Education.

REFERENCES

1. **Kaper JB, Nataro JP, Mobley HL.** 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol **2:**123–140. http://dx.doi.org/10.1038/nrmicro818.
2. **Herman KM, Hall AJ, Gould LH.** 20 February 2015. Outbreaks attributed to fresh leafy vegetables, United States, 1973–2012. Epidemiol Infect http://dx.doi.org/10.1017/S0950268815000047.
3. **Centers for Disease Control and Prevention.** 2006. Ongoing multistate outbreak of *Escherichia coli* serotype O157:H7 infections associated with consumption of fresh spinach—United States, September 2006. MMWR Morb Mortal Wkly Rep **55:**1045–1046.
4. **US Food and Drug Administration.** 2010. Ready Pac Foods, Inc. announces voluntary product recall because of possible health risk. http://www.fda.gov/Safety/Recalls/ucm218325.htm.
5. **Tuttle J, Gomez T, Doyle MP, Wells JG, Zhao T, Tauxe RV, Griffin PM.** 1999. Lessons from a large outbreak of *Escherichia coli* O157:H7 infections: insights into the infectious dose and method of widespread contamination of hamburger patties. Epidemiol Infect **122:**185–192. http://dx.doi.org/10.1017/S0950268898001976.
6. **Tilden J, Jr, Young W, McNamara AM, Custer C, Boesel B, Lambert-Fair MA, Majkowski J, Vugia D, Werner SB, Hollingsworth J, Morris JG, Jr.** 1996. A new route of transmission for *Escherichia coli*: infection from dry fermented salami. Am J Public Health **86:**1142–1145. http://dx.doi.org/10.2105/AJPH.86.8_Pt_1.1142.
7. **Teunis PF, Ogden ID, Strachan NJ.** 2008. Hierarchical dose response of *E. coli* O157:H7 from human outbreaks incorporating heterogeneity in exposure. Epidemiol Infect **136:**761–770.
8. **Feng P, Weagant SD, Jinneman K.** 2013. Diarrheagenic *Escherichia coli*. *In* Bacteriological analytical manual (BAM). US Food and Drug Administration, Silver Spring, MD. http://www.fda.gov/Food/FoodScienceResearch/LaboratoryMethods/ucm070080.htm.
9. **Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin CS, Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J, Rank D, Redman JC, Steyert SR, Frimodt-Moller J, Struve C, Petersen AM, Krogfelt KA, Nataro JP, Schadt EE, Waldor MK.** 2011. Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med **365:**709–717. http://dx.doi.org/10.1056/NEJMoa1106920.
10. **Leff JW, Fierer N.** 2013. Bacterial communities associated with the surfaces of fresh fruits and vegetables. PLoS One **8:**e59310. http://dx.doi.org/10.1371/journal.pone.0059310.
11. **Williams TR, Moyne AL, Harris LJ, Marco ML.** 2013. Season, irrigation, leaf age, and *Escherichia coli* inoculation influence the bacterial diversity in the lettuce phyllosphere. PLoS One **8:**e68642. http://dx.doi.org/10.1371/journal.pone.0068642.
12. **Jackson CR, Randolph KC, Osborn SL, Tyler HL.** 2013. Culture dependent and independent analysis of bacterial communities associated with commercial salad leaf vegetables. BMC Microbiol **13:**274. http://dx.doi.org/10.1186/1471-2180-13-274.
13. **Lopez-Velasco G, Welbaum GE, Boyer RR, Mane SP, Ponder MA.** 2011. Changes in spinach phylloepiphytic bacteria communities following minimal processing and refrigerated storage described using pyrosequencing of 16S rRNA amplicons. J Appl Microbiol **110:**1203–1214. http://dx.doi.org/10.1111/j.1365-2672.2011.04969.x.
14. **Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C.** 2012. Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods **9:**811–814. http://dx.doi.org/10.1038/nmeth.2066.
15. **Wood DE, Salzberg SL.** 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol **15:**R46. http://dx.doi.org/10.1186/gb-2014-15-3-r46.
16. **Ounit R, Wanamaker S, Close TJ, Lonardi S.** 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative *k*-mers. BMC Genomics **16:**236. http://dx.doi.org/10.1186/s12864-015-1419-2.
17. **Manning SD, Motiwala AS, Springman AC, Qi W, Lacher DW, Ouellette LM, Mladonicky JM, Somsel P, Rudrik JT, Dietrich SE, Zhang W, Swaminathan B, Alland D, Whittam TS.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc Natl Acad Sci U S A **105:**4868–4873. http://dx.doi.org/10.1073/pnas.0710834105.
18. **Patro JN, Ramachandran P, Lewis JL, Mammel MK, Barnaba T, Pfeiler EA, Elkins CA.** 2015. Development and utility of the FDA 'GutProbe' DNA microarray for identification, genotyping and metagenomic analysis of commercially available probiotics. J Appl Microbiol **118:**1478–1488. http://dx.doi.org/10.1111/jam.12795.
19. **Rudi K, Flateland SL, Hanssen JF, Bengtsson G, Nissen H.** 2002. Development and evaluation of a 16S ribosomal DNA array-based approach for describing complex microbial communities in ready-to-eat vegetable salads packed in a modified atmosphere. Appl Environ Microbiol **68:**1146–1156. http://dx.doi.org/10.1128/AEM.68.3.1146-1156.2002.
20. **Kase JA, Borenstein S, Blodgett RJ, Feng PC.** 2012. Microbial quality of bagged baby spinach and romaine lettuce: effects of top versus bottom sampling. J Food Prot **75:**132–136. http://dx.doi.org/10.4315/0362-028X.JFP-11-097.
21. **Babic I, Roy S, Watada AE, Wergin WP.** 1996. Changes in microbial populations on fresh cut spinach. Int J Food Microbiol **31:**107–119. http://dx.doi.org/10.1016/0168-1605(96)00969-5.
22. **Loman NJ, Constantinidou C, Christner M, Rohde H, Chan JZ, Quick J, Weir JC, Quince C, Smith GP, Betley JR, Aepfelbacher M, Pallen MJ.** 2013. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic *Escherichia coli* O104:H4. JAMA **309:**1502–1510. http://dx.doi.org/10.1001/jama.2013.3231.
23. **Tsuji A, Kaneko Y, Takahashi K, Ogawa M, Goto S.** 1982. The effects of temperature and pH on the growth of eight enteric and nine glucose non-fermenting species of gram-negative rods. Microbiol Immunol **26:**15–24. http://dx.doi.org/10.1111/j.1348-0421.1982.tb00149.x.
24. **Baylis CL.** 2008. Growth of pure cultures of verocytotoxin-producing *Escherichia coli* in a range of enrichment media. J Appl Microbiol **105:**1259–1265. http://dx.doi.org/10.1111/j.1365-2672.2008.03859.x.