**Title: Metagenomic analysis of viruses associated with field-grown and retail lettuce
identifies human and animal viruses**

Authors: Tiong Gim Aw*, Samantha Wengert, Joan B. Rose

Department of Fisheries and Wildlife, Michigan State University, East Lansing, Michigan 48824,
United States

*Corresponding author: Tiong Gim Aw

Mailing address: 303 Manly Miles, 1405 South Harrison Road, Michigan State University, East
Lansing MI 48824

Phone: (517) 432-1930

Fax: (517) 353-9807

E-mail: tgaw@msu.edu

Abstract

The emergence of culture- and sequence-independent metagenomic methods has not only provided great insight into the microbial community structure in a wide range of clinical and environmental samples but has also proven to be powerful tools for pathogen detection. Recent studies of the food microbiome have revealed vast genetic diversity of bacteria associated with fresh produce. However, no work has been done to apply metagenomic methods to tackle viruses associated with fresh produce for addressing food safety. Thus, there is a little knowledge about the presence and diversity of viruses associated with fresh produce from farm-to-fork. To address this knowledge gap, we assessed viruses on commercial romaine and iceberg lettuce in fields and a produce distribution center using a shotgun metagenomic sequencing targeting both RNA and DNA viruses. Commercial lettuce harbors an immense assemblage of viruses that infect a wide range of hosts. As expected, plant pathogenic viruses dominated these communities. Sequences of rotaviruses and picobirnaviruses were also identified in both field-harvest and retail lettuce samples, suggesting an emerging foodborne transmission threat that has yet to be fully recognized. The identification of human and animal viruses in lettuce samples in the field emphasizes the importance of preventing viral contamination on leafy greens starting at the field. Although there are still some inherent experimental and bioinformatics challenges in applying viral metagenomic approaches for food safety testing, this work will facilitate further application of this unprecedented deep sequencing method to food samples.

Keywords

Metagenomics, Foodborne viruses, Fresh produce, Next generation sequencing, Rotavirus, Picobirnavirus

Introduction

Fresh produce, particularly leafy greens, is an important component of a healthy and balanced diet, yet one that is increasingly being recognized as an important vehicle for transmission of human pathogens including bacteria, viruses and parasites. Most recently, fresh produce has been identified as a major contributor to foodborne outbreaks worldwide (Berger et al., 2010; Lynch et al., 2009; Newell et al., 2010).

The number of foodborne outbreaks in the United States (U.S.) has continued to serve as an indicator of the vulnerability of our food systems to contamination and public health risk. The Centers for Disease Control and Prevention (CDC) has estimated, based on the data from 2000 to 2008, that each year 31 major pathogens caused 9.4 million episodes of foodborne illness in the U.S., with most illnesses caused by viruses (59%), particularly noroviruses (Scallan et al., 2011b). Although norovirus is currently the leading foodborne virus of concern, all pathogenic viruses of fecal origin can potentially cause foodborne illnesses ranging from acute gastroenteritis to chronic disease complications such as myocarditis, meningitis, liver disease and neurological symptoms (Newell et al., 2010). In addition, CDC's estimates that approximately 80% of foodborne disease cases in the U.S. are caused by unspecified agents (Scallan et al., 2011a) which suggests that a better foodborne disease surveillance system is needed to fill a current knowledge gap concerning unknown and unidentified foodborne agents.

While numerous studies of bacterial contamination of fresh produce have been reported (Allen et al., 2013; Tango et al., 2014; Wijnands et al., 2014; Wood et al., 2015), our understanding of viral contamination of commercial fresh produce remains very limited (Kokkinos et al., 2012; Mattison et al., 2010), partly due to the difficulty in detecting viruses in food matrices. Detection of foodborne viruses has undergone a significant transition from

70 traditional cell culture to molecular techniques, particularly PCR-based assays (Ceuppens et al.,
71 2014; De Medici et al., 2015). However, PCR-based detection is largely restricted to known and
72 well-characterized viruses, which have been sequenced, and there is currently no universal PCR
73 assay that can target all viruses in a single sample. The rapid development of culture- and
74 sequence-independent metagenomics coupled with next generation sequencing (NGS) offers
75 novel and exciting opportunities to enhance our understanding of complex, diverse and dynamic
76 microbial communities including viruses associated with food environments, and thus improve
77 foodborne pathogen identification and potential contamination source tracking (Bergholz et al.,
78 2014). Previous studies investigating microbial communities associated with fresh produce using
79 the NGS, mostly 16 S rRNA pyrosequencing, have primarily focused on bacteria (Jackson et al.,
80 2013; Leff and Fierer, 2013; Talias et al., 2011) and no study has been conducted to assess the
81 composition of virus communities of raw fresh produce. Until recently, viral metagenomic
82 approaches have only been used to address bacteriophages in fermented food (Park et al., 2011)
83 and plant pathogenic viruses in diseased plants (Coetzee et al., 2010).

84 A firm understanding of the virus ecology is needed for better defining the key risk
85 associated with fresh produce vegetables, which are usually consumed raw or minimally
86 processed. In this study, we describe the diversity and composition of virus communities (or
87 viromes), with a particular focus on potential foodborne viruses, associated with lettuce collected
88 from commercial production fields and a produce distribution center using shotgun metagenomic
89 sequencing.

Material and Methods

Lettuce sample collection for viral metagenomics

Samples of field-grown iceberg and romaine lettuce were collected directly from the fields in the U.S. during December of 2013. A total of 42 (21 iceberg and 21 romaine) lettuce heads were collected at different stages of farm-level production. A detailed sample description is provided in Table 1. A total of 13 iceberg and 5 romaine lettuce samples were hand cut by field workers and placed on a packaging machine for bagging prior to sampling. A total of 8 romaine lettuce samples were collected after washing and cutting by workers on a processing machine in the field for bagged salad production. In addition, a total of 8 iceberg and 8 romaine lettuce samples were hand cut at ground level by the research team using gloves and a sterile harvesting knife as control samples. The outer leaflets were removed before placing the heads in large sterile Whirl-pak bags (Nasco). The purpose of this was to determine potential sources of viral contamination in lettuce field pre- and post-harvest.

A total of 54 (27 iceberg and 27 romaine) lettuce samples were also collected from a produce distribution center monthly from January to March 2014. Fresh produce vegetables are usually stored in the distribution center no longer than three days before they are distributed to retail stores. During each sampling event, three cartons of each iceberg and romaine lettuce were opened and three samples were taken from each carton. The iceberg lettuce collected in this study was wrapped and sealed in a plastic bag at harvest. All lettuce samples were collected using sterile, disposable gloves and placed into large Whirl-pak bags. All Samples were immediately transported to the laboratory for processing. A total of 96 lettuce samples were processed for viral metagenomics.

Recovery of viruses from lettuce

The recovery of viruses from lettuce was performed using a concentration method as previously described with modifications (Dubois et al., 2006). Briefly, the outer leaflets were cut with a scalpel 2.5 to 5 cm from the core under sterile conditions. Fifty grams of each sample was washed in a Whirl-pak bag with 250 ml sterile 100 mM Tris- 50 mM glycine buffer at a pH of 9.5 and gently mixed for 20 min at room temperature. The wash solution was recovered and immediately adjusted to neutral pH (7.2 ± 0.2). Viral particles contained in the wash solution were further concentrated and purified by polyethylene glycol (PEG) precipitation. The samples were mixed with 10% (wt/vol) molecular biology grade PEG 8000 (Promega Corporation, Madison WI) and 0.3 M NaCl (w/v). The samples were incubated at 4°C for approximately 18 hours before centrifuging the sample at 10,800 x g (8000 rpm) for 30 min at 4°C. The supernatant was then carefully poured off and the remaining pellet was dissolved in 20 mL of sterile phosphate buffered saline (PBS) by soaking at room temperature for 1 h. An equal volume of chloroform was added to each PEG precipitate to remove the PEG and purify the sample. The solution was then vortexed gently for 30 s and centrifuged at 3000 x g (4300 rpm) for 15 min at 4°C to collect the supernatant containing virus particles. The remaining supernatant was then passed through 0.45- and 0.22-µm filters and further concentrated to approximately 1 ml by Amicon centrifugal ultrafiltration (30kDa, Millipore, Billerica, MA).

Nucleic acid extraction and sequencing

The final concentrates (1 ml) were treated with 100 units of DNase-I (Roche) for 1 hour at 37°C before nucleic acid extraction to remove free nucleic acids from the concentrated virus samples. Viral DNA and RNA were simultaneously extracted using PureLink Viral RNA/DNA Mini Kit

(Life Technologies) following the manufacturer's instructions. For each viral concentrate, three individual nucleic acid extracts were prepared to minimize nucleic acid extraction bias. Following extraction, the samples were screened by 16S ribosomal DNA (rDNA) PCR with 27F/1492R universal primers to ensure the absence of any contaminating microbial DNA. To obtain a sufficient DNA and cDNA (for RNA viruses) for metagenomics sequencing, the viral nucleic acids were reverse transcribed and amplified as previously described (Wang et al., 2002; Wang et al., 2003). Briefly, RNA was reverse transcribed with Primer A (5'-GTTTCCCAGTCACGATCNNNNNNNNN-3') using Superscript III reverse transcriptase (Life Technologies). Sequenase 2.0 (USB/Affymetrix, Cleveland, OH, USA) was used for second-strand cDNA synthesis and for random-primed amplification of viral DNA. Each sample was then subjected to 40 cycles of PCR amplification with Primer B (5'-GTTTCCCAGTCACGATC-3') using AmpliTaq Gold (Life Technologies). Three PCR reactions were performed from the same nucleic acid extract to minimize amplification bias and the PCR products were pooled. PCR products were purified using Promega Wizard SV Gel and PCR Clean-Up System (Promega Corporation).

Libraries from each sample were prepared using Rubicon ThruPLEX DNA-seq kit (Rubicon Genomics) with a unique dual index adapter pair for each sample. Samples were sequenced in a 2x100- base pair (bp) paired end format using two lanes of an Illumina HiSeq 2500 Rapid Run flow cell at the Research Technology Support Facility, Michigan State University.

Bioinformatics analysis of viromes

Prior to assembly and annotation of the metagenomic dataset, the quality of the Illumina sequencing data was checked using FastQC. Each data was screened for the 17-bp Primer B sequence and any reads homologous to the Primer B sequence at their 5' ends were removed using cutadapt with a maximum error rate of 0.2 and minimum overlap of 10 bases (Martin, 2011). Trimmomatic was used for sequencing adapter removal and quality trimming with parameters included: a maximum mismatch count value of 2 allowed for a full match (seed mismatch), a palindrome clip threshold of 30, a simple clip threshold of 10, a minimum adapter length of 8 with both the forward and reverse read kept, removal of low quality leading and trailing bases below a quality of 3, a 4-base sliding window scan that cuts when the average quality is below 15, and removal of reads less than 30 bases long (Bolger et al., 2014). Following filtering and trimming of raw reads, paired-end reads were subjected to *de novo* assembly into a longer contiguous sequence (contig) using IDBA-UD (Peng et al., 2012). Contigs larger than 200-bp were then queried against the National Center for Biotechnology Information (NCBI) Viral Reference Sequence (RefSeq) database for taxonomic assignment using BLASTX with an E-value cutoff of 10^{-5} . The BLASTX output was parsed using the MEtaGenome Analyzer (MEGAN) version 5.6.6 with the following parameters for the Lowest Common Ancestor (LCA) algorithm: Min Score = 50.0, Max Expected = $1.0E^{-5}$, Top Percent = 10.0, Min Support Percent = 0.1, Min Support = 1, and LCA Percent = 100.0 (Huson et al., 2007).

We focused on the identification of human and animal viruses, which have been suggested as potential foodborne viruses. Contigs identified as viral pathogens of human and animal were extracted and further used as the queries in BLASTX against the NCBI non-redundant (nr) sequence database. For rotavirus contigs, BLASTN was used (Matthijnssens et al.,

2008). To determine relative abundance of a phylogenetic group, read mapping to contigs was performed using Bowtie 2 version 2.1.0 with default settings (Langmead and Salzberg, 2012). Relative abundance for each contig, the number of reads aligned to a contig divided by the contig length was calculated (Kim et al., 2015). The relative abundance of each phylogenetic group was calculated by summing the abundance of each contig classified in a particular group.

The data from Illumina sequencing have been deposited in the NCBI Sequence Read Archive (www.ncbi.nlm.nih.gov/sra) under accession number SRP066046 (BioProject PRJNA295085).

Results

Identification of human and animal viruses in lettuce virome using metagenomics

Ninety-six lettuce samples were collected and sequenced using the Illumina HiSeq 2500 instrument, generating a total of 770.2 and 785.2 million paired-end 100-bp sequence reads, with an average of 25.7 ± 7.8 and 26.2 ± 8.1 (mean $\pm$ s.d.) million reads per sample for romaine and iceberg lettuce, respectively (after quality trimming and filtering of reads). The paired-end reads were assembled into contigs, producing a total of 53 978 and 65 431 assembled contigs for romaine and iceberg lettuce, respectively (Supplementary material: Table S1 and S2). The mean contig size among the contigs across all viromes was 680-bp. Taxonomic analysis using the BLASTX search against the NCBI viral RefSeq database and LCA algorithm assigned 10.0 to 19.7% of the contigs to viruses (Table 1).

Of the contigs with similarity to known viruses, 1.5 to 3.0% of the sequences were similar to viruses that infect vertebrates (Figure 1). Contigs belonging to viral pathogens of human and animal particularly in the family of *Reoviridae* and *Picobirnaviridae* were detected in

all sample categories including field-grown lettuce prior to harvest, produce harvested by workers and from produce distribution center (Table 1). Specifically, these viral pathogens were more abundant in lettuce samples collected from produce distribution center with 56 and 80 contigs for romaine and iceberg lettuce, respectively (Table 1).

In lettuce fields, rotavirus A and rotavirus C were identified in one sample harvested by workers and one control sample harvested by the research team, respectively (Table 2). The rotavirus A contig shared a 99% nucleotide identity to the VP1 gene of a human G8P[14] rotavirus strain (Genbank accession number KP006506). Rotavirus C identified in the romaine lettuce control sample shared a 98% nucleotide identity to the NSP2 gene of a bovine isolate. Rotaviruses (Group A, D, F and G) were identified in 8 viromes (representing 8 cartons of lettuce with 3 samples from each carton) of the lettuce samples collected from produce distribution center with high nucleotide identity (79 to 98%) to the rotavirus sequences in the NCBI database (Table 2). Group A rotaviruses of human, bovine and porcine were abundant and multiple double-stranded RNA genome segments (11 segments) of rotaviruses were identified in lettuce.

Viral contigs related to double-stranded RNA viruses in the family of *Picobirnaviridae* were identified in 6 field-grown lettuce samples, including romaine lettuce for value-added production (Table 3). In addition, both human and animal picobirnaviruses were also identified in 9 viromes (representing 9 cartons of lettuce with 3 samples from each carton) of the lettuce samples collected from produce distribution center (Table 3). Both genomic segments of picobirnaviruses (segment 1 encodes for capsid protein and segment 2 encodes for the RNA-dependent RNA polymerase) were detected. However, almost all individual contigs related to picobirnavirus capsid protein possessed less than 30% amino acid identity to existing picobirnavirus sequences in the NCBI database. Contigs related to picobirnavirus RNA-

dependent RNA polymerase had 68-78% amino acid identity to known picobirnavirus sequences in the NCBI database.

Papillomavirus was identified in one field-grown iceberg lettuce sample harvested by the research team. BLAST analysis of this contig showed a nucleotide identity of 98% (query coverage of 91% and E-value of 0) to human papillomavirus type 3 and 100% of amino acid identity to the partial late protein gene of human papillomavirus (Table 3). Other contigs related to animal viruses including cyclovirus and porcine bocavirus were also detected in lettuce viromes (Table 3).

Other viruses associated with commercial lettuce

Besides viral pathogens of human and animal, viruses infecting a wide range of hosts, including plants, bacteria, invertebrate, amoeba, fungi and alga were also identified in lettuce viromes (Figure 1). Plant pathogenic viruses were dominant, with relative abundance of 66.7 and 64.4% in romaine and iceberg lettuce, respectively. Relative abundance of plant viral families showed that the majority of assigned viral contigs were homologous to *Caulimoviridae* and *Closteroviridae* (Supplementary material: Table S3). Of the contigs with similarity to bacteriophages, the *Siphoviridae* and *Myoviridae* families were abundant (Supplementary material: Table S4). Bacteriophage families in the lettuce virome were associated with 63 different bacterial hosts, with the majority being *Enterobacteria* phages. Several contigs showed homology to bacteriophages whose hosts belong to genera that also include human pathogens such as *Pseudomonas* phages, *Escherichia* phages, *Salmonella* phages and *Vibrio* phages.

Discussion

To our knowledge, this is the first report of the characterization of the commercial lettuce virome using metagenomics that demonstrates the presence of human and animal viruses in field-grown and retail lettuce. The ability to identify viral pathogens directly in lettuce samples using metagenomics demonstrates the advancement of this approach for food safety monitoring without a priori knowledge of what viruses may be present. The majority of human and animal viruses identified in lettuce viromes were segmented double-stranded RNA viruses including rotaviruses and picobirnaviruses. As the current foodborne viruses of concern are single-stranded RNA viruses such as noroviruses, hepatitis A and E viruses, this study suggests a potential risk for foodborne transmission of rotaviruses and picobirnaviruses from raw lettuce. Noroviruses were not identified in this study even though human noroviruses are often associated with foodborne illnesses. This could be due to the insufficient sequencing depth, which ultimately results in higher method detection limits for certain foodborne viruses. Metagenomics measures a relative abundance, which means that even in samples with a high absolute abundance of a particular pathogen, the identification may be obscured by even higher concentrations of other viruses (Bibby and Peccia, 2013) such as plant pathogenic viruses, bacteriophages and double-stranded RNA viruses as observed in this study. In addition, the absence of other foodborne viruses in this study could be due to the seasonal variations in the prevalence of these viruses. Further studies are needed to determine the seasonal profile of viruses in fresh produce using metagenomics and the effect of sequencing depth on detection of foodborne viruses, particularly single-stranded RNA viruses in food samples.

Rotavirus is a genus in *Reoviridae* family with a genome of 11 segments of double-stranded RNA encoding six structural viral proteins (VP1, VP2, VP3, VP4, VP6, VP7) and six non-structural proteins (NSP1-NSP6) (Matthijssens et al., 2008). In this study, all rotavirus RNA segments were detected. Rotaviruses particularly group A are the major cause of severe acute gastroenteritis in children and the elderly as well as adults with immunosuppression (Ruggeri and Fiore, 2013). Interestingly, a recent survey in the U.S. showed unexpectedly high prevalence of rotaviruses among adults with acute gastroenteritis in emergency departments (Bresee et al., 2012). Foodborne illnesses and outbreaks associated with rotavirus A have been reported, for example in the District of Columbia, U.S in 2000 after college students eating tuna or chicken salad sandwiches from dining hall (Fletcher et al., 2001) and most recently in Japan in 2012 after high school students eating raw sliced cabbage (Mizukoshi et al., 2014). Although rotavirus infections are usually acquired by ingesting contaminated water or food, little is known about the presence of rotaviruses in food commodities that is consumed raw. Recently, rotaviruses have been detected in fresh vegetables in Mexico (Felix-Valenzuela et al., 2012; Quiroz-Santiago et al., 2014), packaged leafy greens in Canada (Mattison et al., 2010) and field-grown strawberries (Brassard et al., 2012). This study, which focused on lettuce, further suggests that fresh produce could serve as a vehicle for wide dissemination of rotaviruses. There is now increasing evidence that strains of animal rotavirus particularly bovine and porcine origin can infect humans, suggesting the potential for zoonotic transmission of rotaviruses (Midgley et al., 2012; Papp et al., 2013). However, avian rotaviruses of groups A, D, F and G have not been detected in humans.

Picobirnaviruses are small, bisegmented double-stranded RNA viruses of vertebrate hosts including human. The identification of picobirnaviruses associated with a broad range of host is

of interest because of their potential of zoonotic transmission (Ganesh et al., 2012). Although picobirnaviruses have been detected in sporadic episodes of diarrhea and outbreaks of acute gastroenteritis, the pathogenicity of these viruses remain unclear (Ganesh et al., 2012). There is currently no documented evidence of foodborne transmission of picobirnaviruses. Picobirnaviruses have been shown to be prevalent in environmental water and wastewater samples (Hamza et al., 2011; Symonds et al., 2009; Zhang et al., 2015). Taken together, the results of this study and other studies on environmental water surveillance suggest the potential for waterborne and foodborne transmission of picobirnaviruses. The detection of picobirnaviruses with low levels of amino acid similarity to known picobirnavirus sequences in this study suggests limited knowledge of the genome of the *Picobirnaviridae* family and the role it plays in food systems. Genome sequences of *Picobirnaviridae* are currently highly underrepresented in sequence databases (Malik et al., 2014). At present, cell culture system or animal model for picobirnaviruses has not been established.

Recent studies have shown the prevalence and wide dissemination of oncogenic potential viruses including human papillomavirus in water and wastewater environments (Iaconelli et al., 2015; La Rosa et al., 2013) and bioaerosol in animal slaughterhouses (Hall et al., 2013). In this study, human papillomavirus type 3 was identified in an iceberg lettuce control sample (control associated with harvesting by the research team), indicating irrigation water could be a possible contamination source. Although human papillomavirus infections are transmitted primarily through direct skin-to-skin contact and are considered as sexually transmitted diseases, non-sexual transmission of the virus is still under debate (Ryndock and Meyers, 2014). However, human papillomaviruses have been shown to be highly stable outside the host and have the potential to be transmitted via fomites and surfaces (Ryndock and Meyers, 2014). This study

shows the potential exposure of papillomaviruses to harvest workers in the field and consumers by contaminating food and environmental fomites or surfaces at home.

Fresh produce, such as lettuce, which is subjected to intense production practices, has many opportunities for pathogen contamination from farm-to-fork. The main routes of contamination could be through contaminated irrigation water, the use of animal manure as fertilizer and animals in the field as well as human handling during and post-harvest. The identification of human and animal viruses in both romaine and iceberg lettuce samples in this study showed that the contamination occurred at the pre-harvest stage of production from the field through to the distribution center prior to moving to the store. In this study, rotavirus A was found in an iceberg lettuce sample collected by harvest workers and found to have a high percentage of sequence identity match to a human isolate, which suggests this enteric virus was present on the lettuce either prior or following field harvest. In addition, rotavirus A is a well-known waterborne pathogen and possible sources of contamination could be from fecally contaminated irrigation water or poor worker hygiene. Bovine rotavirus C was identified in a romaine lettuce control sample prior to harvest, indicating fecal contamination from cattle with possible sources including contaminated irrigation water or improperly treated manure used as fertilizer. The detection of enteric viruses in lettuce samples collected from a produce distribution center suggests that viruses are highly persistent and stable once attached to lettuce and being transported. However, metagenomic approach detects viral nucleic acid and further study using cell culture is needed to determine whether these viruses are infectious after the transportation to produce distribution center and retail stores. In addition, further study using viral metagenomic approaches to determine sources of contamination in lettuce field is warranted.

Our metagenomic analysis detected not only sequences belonged to human and animal viruses, but also demonstrated that commercial lettuce harbors viruses that infect a wide range of hosts prior and following field harvest. As expected, plant pathogenic viruses were dominant in lettuce viromes. While most viruses identified in these lettuce viromes are likely to be non-pathogenic to humans, this study demonstrates that metagenomics can widely detect and characterize human and animal viral pathogens that present at low abundance in complex food matrices. These pathogenic viruses can become internalized in lettuce once the produce is contaminated (DiCaprio et al., 2015; Urbanucci et al., 2009) and pose a potential risk to consumer safety as fresh produce vegetables are usually consumed without the use of preparation methods that would remove or kill the associated virus communities. A better knowledge about what viruses may be present in our food, even if they do not associated with human diseases, can be valuable for the surveillance of potential new public health threats.

In conclusion, the findings of this study provide new insights into viruses including human and animal viruses in commercial lettuce, emphasizing the importance of preventing viral contamination on leafy greens starting at the field. The modern food system is complex and new approaches that use microbial genetic data to better understand and improve food safety are emerging (Bergholz et al., 2014; Stasiewicz et al., 2015). Recently, whole genome sequencing has already transformed food safety through improved surveillance of bacterial foodborne pathogens and outbreak investigation (Stasiewicz et al., 2015). Although metagenomic sequencing has not been widely adapted for evaluating food virome to date, this study suggests that viral metagenomic sequencing could be a promising monitoring tool for food safety, which allows for sequence-independent detection and identification of nonculturable viruses and previously unknown pathogens in both food matrices and clinical samples, and could be used

particularly for outbreak investigation.

Acknowledgements

This project was supported by Agriculture and Food Research Initiative Competitive Grant no. 2013-67005-21188 from the National Institute of Food and Agriculture, U.S. Department of Agriculture.

We acknowledge Channah Rock and Kurt Nolte at University of Arizona for helping with the field sampling. We thank the Research Technology Support Facility and the High Performance Computing Center at Michigan State University for sequencing and computational supports.

References

Allen, K.J., Kovacevic, J., Cancarevic, A., Wood, J., Xu, J.Q., Gill, B., Allen, J.K., Mesak, L.R., 2013. Microbiological survey of imported produce available at retail across Canada. International Journal of Food Microbiology 162, 135-142.

Berger, C.N., Sodha, S.V., Shaw, R.K., Griffin, P.M., Pink, D., Hand, P., Frankel, G., 2010. Fresh fruit and vegetables as vehicles for the transmission of human pathogens. Environmental Microbiology 12, 2385-2397.

Bergholz, T.M., Switt, A.I.M., Wiedmann, M., 2014. Omics approaches in food safety: fulfilling the promise? Trends in Microbiology 22, 275-281.

391 Bibby, K., Peccia, J., 2013. Identification of viral pathogen diversity in sewage sludge by
392 metagenome analysis. *Environmental Science & Technology* 47, 1945-1951.

393 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina
394 sequence data. *Bioinformatics* 30, 2114-2120.

395 Brassard, J., Gagne, M.J., Genereux, M., Cote, C., 2012. Detection of Human Food-Borne and
396 Zoonotic Viruses on Irrigated, Field-Grown Strawberries. *Applied and Environmental*
397 *Microbiology* 78, 3763-3766.

398 Bresee, J.S., Marcus, R., Venezia, R.A., Keene, W.E., Morse, D., Thanassi, M., Brunett, P.,
399 Bulens, S., Beard, R.S., Dauphin, L.A., Slutsker, L., Bopp, C., Eberhard, M., Hall, A., Vinje, J.,
400 Monroe, S.S., Glass, R.I., Etiology, U.S.A.G., 2012. The Etiology of Severe Acute
401 Gastroenteritis Among Adults Visiting Emergency Departments in the United States. *Journal of*
402 *Infectious Diseases* 205, 1374-1381.

403 Ceuppens, S., Li, D., Uyttendaele, M., Renault, P., Ross, P., Van Ranst, M., Cocolin, L., Donaghy,
404 J., 2014. Molecular Methods in Food Safety Microbiology: Interpretation and Implications of
405 Nucleic Acid Detection. *Comprehensive Reviews in Food Science and Food Safety* 13, 551-577.

406 Coetzee, B., Freeborough, M.J., Maree, H.J., Celton, J.M., Rees, D.J.G., Burger, J.T., 2010. Deep
407 sequencing analysis of viruses infecting grapevines: Virome of a vineyard. *Virology* 400, 157-
408 163.

409

410 De Medici, D., Kuchta, T., Knutsson, R., Angelov, A., Auricchio, B., Barbanera, M., Diaz-Amigo,
 411 C., Fiore, A., Kudirkiene, E., Hohl, A., Tomic, D.H., Gotcheva, V., Popping, B., Prukner-
 412 Radovcic, E., Scaramaglia, S., Siekel, P., To, K.A., Wagner, M., 2015. Rapid Methods for
 413 Quality Assurance of Foods: the Next Decade with Polymerase Chain Reaction (PCR)-Based
 414 Food Monitoring. *Food Analytical Methods* 8, 255-271.

415 DiCaprio, E., Purgianto, A., Lia, J.R., 2015. Effects of Abiotic and Biotic Stresses on the
 416 Internalization and Dissemination of Human Norovirus Surrogates in Growing Romaine Lettuce.
 417 *Applied and Environmental Microbiology* 81, 4791-4800.

418 Dubois, E., Hennechart, C., Deboosere, N., Merle, G., Legeay, O., Burger, C., Le Calve, M.,
 419 Lombard, B., Ferre, V., Traore, O., 2006. Intra-laboratory validation of a concentration method
 420 adapted for the enumeration of infectious F-specific RNA coliphage, enterovirus, and hepatitis A
 421 virus from inoculated leaves of salad vegetables. *International Journal of Food Microbiology* 108,
 422 164-171.

423 Felix-Valenzuela, L., Resendiz-Sandoval, M., Burgara-Estrella, A., Hernandez, J., Mata-Haro, V.,
 424 2012. Quantitative detection of hepatitis A, rotavirus and genogroup I Norovirus by RT-qPCR in
 425 fresh produce from packinghouse facilities. *Journal of Food Safety* 32, 467-473.

426 Fletcher, M., Levy, M.E., Griffin, D.D., 2001. Foodborne outbreak of group A rotavirus
 427 gastroenteritis among college students - District of Columbia, March-April 2000 (Reprinted from
 428 MMWR, vol 49, pg 1131-1133, 2000). *JAMA-Journal of the American Medical Association* 285,
 429 405-406.

430 Ganesh, B., Banyai, K., Martella, V., Jakab, F., Masachessi, G., Kobayashi, N., 2012.

431 Picobirnavirus infections: viral persistence and zoonotic potential. *Reviews in Medical Virology*

432 22, 245-256.

433 Hall, R.J., Leblanc-Maridor, M., Wang, J., Ren, X.Y., Moore, N.E., Brooks, C.R., Peacey, M.,

434 Douwes, J., McLean, D.J., 2013. Metagenomic Detection of Viruses in Aerosol Samples from

435 Workers in Animal Slaughterhouses. *PLoS ONE* 8, e72226.

436 Hamza, I.A., Jurzik, L., Uberla, K., Wilhelm, M., 2011. Evaluation of pepper mild mottle virus,

437 human picobirnavirus and Torque teno virus as indicators of fecal contamination in river water.

438 *Water Research* 45, 1358-1368.

439 Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C., 2007. MEGAN analysis of metagenomic data.

440 *Genome Research* 17, 377-386.

441 Iaconelli, M., Petricca, S., Libera, S.D., Di Bonito, P., La Rosa, G., 2015. First Detection of

442 Human Papillomaviruses and Human Polyomaviruses in River Waters in Italy. *Food and*

443 *Environmental Virology* 7, 309-315.

444 Jackson, C.R., Randolph, K.C., Osborn, S.L., Tyler, H.L., 2013. Culture dependent and

445 independent analysis of bacterial communities associated with commercial salad leaf vegetables.

446 *BMC Microbiology* 13, 274.

447 Kim, Y., Aw, T.G., Teal, T.K., Rose, J.B., 2015. Metagenomic Investigation of Viral Communities

448 in Ballast Water. *Environmental Science & Technology* 49, 8396-8407.

449 Kokkinos, P., Kozyra, I., Lazic, S., Bouwknecht, M., Rutjes, S., Willems, K., Moloney, R.,
 450 Husman, A.M.D., Kaupke, A., Legaki, E., D'Agostino, M., Cook, N., Rzezutka, A., Petrovic, T.,
 451 Vantarakis, A., 2012. Harmonised Investigation of the Occurrence of Human Enteric Viruses in
 452 the Leafy Green Vegetable Supply Chain in Three European Countries. Food and Environmental
 453 Virology 4, 179-191.

454 La Rosa, G., Fratini, M., Accardi, L., D'Oro, G., Della Libera, S., Muscillo, M., Di Bonito, P.,
 455 2013. Mucosal and Cutaneous Human Papillomaviruses Detected in Raw Sewages. PLoS ONE 8,
 456 e52391.

457 Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods
 458 9, 357-U354.

459 Leff, J.W., Fierer, N., 2013. Bacterial Communities Associated with the Surfaces of Fresh Fruits
 460 and Vegetables. PLoS ONE 8, e59310.

461 Lynch, M.F., Tauxe, R.V., Hedberg, C.W., 2009. The growing burden of foodborne outbreaks due
 462 to contaminated fresh produce: risks and opportunities. Epidemiology and Infection 137, 307-
 463 315.

464 Malik, Y.S., Kumar, N., Sharma, K., Dhama, K., Shabbir, M.Z., Ganesh, B., Kobayashi, N.,
 465 Banyai, K., 2014. Epidemiology, Phylogeny, and Evolution of Emerging Enteric
 466 Picobirnaviruses of Animal Origin and Their Relationship to Human Strains. Biomed Research
 467 International 2014, 1-13.

468

469 Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
 470 EMBnet.journal 17, 10-12.

471 Matthijnssens, J., Ciarlet, M., Rahman, M., Attoui, H., Banyai, K., Estes, M.K., Gentsch, J.R.,
 472 Iturriza-Gomara, M., Kirkwood, C.D., Martella, V., Mertens, P.P.C., Nakagomi, O., Patton, J.T.,
 473 Ruggeri, F.M., Saif, L.J., Santos, N., Steyer, A., Taniguchi, K., Desselberger, U., Van Ranst, M.,
 474 2008. Recommendations for the classification of group A rotaviruses using all 11 genomic RNA
 475 segments. Archives of Virology 153, 1621-1629.

476 Mattison, K., Harlow, J., Morton, V., Cook, A., Pollari, F., Bidawid, S., Farber, J.M., 2010.
 477 Enteric Viruses in Ready-to-Eat Packaged Leafy Greens. Emerging Infectious Diseases 16, 1815-
 478 1817.

479 Midgley, S.E., Hjulsager, C.K., Larsen, L.E., Falkenhorst, G., Bottiger, B., 2012. Suspected
 480 zoonotic transmission of rotavirus group A in Danish adults. Epidemiology and Infection 140,
 481 1013-1017.

482 Mizukoshi, F., Kuroda, M., Tsukagoshi, H., Sekizuka, T., Funatogawa, K., Morita, Y., Noda, M.,
 483 Katayama, K., Kimura, H., 2014. A food-borne outbreak of gastroenteritis due to genotype G1P8
 484 rotavirus among adolescents in Japan. Microbiology and Immunology 58, 536-539.

485 Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., Opsteegh,
 486 M., Langelaar, M., Threfall, J., Scheutz, F., van der Giessen, J., Kruse, H., 2010. Food-borne
 487 diseases - The challenges of 20 years ago still persist while new ones continue to emerge.
 488 International Journal of Food Microbiology 139, S3-S15.

489 Papp, H., Borzak, R., Farkas, S., Kisfali, P., Lengyel, G., Molnar, P., Meleg, B., Matthijnsens,
 490 J., Jakab, F., Martella, V., Banyai, K., 2013. Zoonotic transmission of reassortant porcine G4P 6
 491 rotaviruses in Hungarian pediatric patients identified sporadically over a 15 year period.
 492 *Infection Genetics and Evolution* 19, 71-80.

493 Park, E.J., Kim, K.H., Abell, G.C.J., Kim, M.S., Roh, S.W., Bae, J.W., 2011. Metagenomic
 494 Analysis of the Viral Communities in Fermented Foods. *Applied and Environmental*
 495 *Microbiology* 77, 1284-1291.

496 Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: a de novo assembler for
 497 single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28,
 498 1420-1428.

499 Quiroz-Santiago, C., Vazquez-Salinas, C., Natividad-Bonifacio, I., Barron-Romero, B.L.,
 500 Quinones-Ramirez, E.I., 2014. Rotavirus G2P 4 Detection in Fresh Vegetables and Oysters in
 501 Mexico City. *Journal of Food Protection* 77, 1953-1959.

502 Ruggeri, F.M., Fiore, L., 2013. Advances in understanding of rotaviruses as food- and
 503 waterborne pathogens and progress with vaccine development, *Viruses in Food and Water: Risks,*
 504 *Surveillance and Control*, pp. 362-400.

505 Ryndock, E.J., Meyers, C., 2014. A risk for non-sexual transmission of human papillomavirus?
 506 *Expert Review of Anti-infective Therapy* 12, 1165-1170.

507 Scallan, E., Griffin, P.M., Angulo, F.J., Tauxe, R.V., Hoekstra, R.M., 2011a. Foodborne Illness
 508 Acquired in the United States-Unspecified Agents. *Emerging Infectious Diseases* 17, 16-22.

509 Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., Jones, J.L.,
510 Griffin, P.M., 2011b. Foodborne Illness Acquired in the United States-Major Pathogens.
511 Emerging Infectious Diseases 17, 7-15.

512 Stasiewicz, M.J., den Bakker, H.C., Wiedmann, M., 2015. Genomics tools in microbial food
513 safety. Current Opinion in Food Science 4, 105-110.

514 Symonds, E.M., Griffin, D.W., Breitbart, M., 2009. Eukaryotic Viruses in Wastewater Samples
515 from the United States. Applied and Environmental Microbiology 75, 1402-1409.

516 Tango, C.N., Choi, N.J., Chung, M.S., Oh, D.H., 2014. Bacteriological Quality of Vegetables
517 from Organic and Conventional Production in Different Areas of Korea. Journal of Food
518 Protection 77, 1411-1417.

519 Telias, A., White, J.R., Pahl, D.M., Ottesen, A.R., Walsh, C.S., 2011. Bacterial community
520 diversity and variation in spray water sources and the tomato fruit surface. BMC Microbiology
521 11, 81.

522 Urbanucci, A., Myrmel, M., Berg, I., von Bonsdorff, C.H., Maunula, L., 2009. Potential
523 internalisation of caliciviruses in lettuce. International Journal of Food Microbiology 135, 175-
524 178.

525 Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D., DeRisi, J.L., 2002.
526 Microarray-based detection and genotyping of viral pathogens. Proceedings of the National
527 Academy of Sciences of the United States of America 99, 15687-15692.

528 Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R.,
529 Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L.,
530 2003. Viral discovery and sequence recovery using DNA microarrays. *Plos Biology* 1, 257-260.

531 Wijnands, L.M., Delfgou-van Asch, E.H.M., Beerepoot-Mensink, M.E., van der Meij-Florijn, A.,
532 Fitz-James, I., van Leusden, F.M., Pielaat, A., 2014. Prevalence and Concentration of Bacterial
533 Pathogens in Raw Produce and Minimally Processed Packaged Salads Produced in and for The
534 Netherlands. *Journal of Food Protection* 77, 388-394.

535 Wood, J.L., Chen, J.C., Friesen, E., Delaquis, P., Allen, K.J., 2015. Microbiological Survey of
536 Locally Grown Lettuce Sold at Farmers' Markets in Vancouver, British Columbia. *Journal of*
537 *Food Protection* 78, 203-208.

538 Zhang, S.W., Bai, R., Feng, R., Zhang, H.X., Liu, L.X., 2015. Detection and evolutionary
539 analysis of picobirnaviruses in treated wastewater. *Microbial Biotechnology* 8, 474-482.

540

541

542

543

544

545

546

547

548

549

Table 1 Description of the lettuce samples collected for viral metagenomics, the percentage of viral contigs and number of contig identified as viral pathogens of human and animal for each sample category.

Lettuce type	Sample description	No. of sample	No. (%) of contigs assigned to virus ^a	No. of contig identified as viral pathogens of human and animal
Romaine	Lettuce field (control), lettuce harvested by research team	8	2,366 (17.9)	1
	Lettuce field - lettuce harvested by field worker	5	1,175 (13.6)	4
	Lettuce field - wash and cut for bagged salad (value-added produce)	8	1,658 (10.5)	1
	Produce distribution center	27	3,110 (12.5)	56
Iceberg	Lettuce field (control), lettuce harvested by research team	8	2,122 (10.9)	2
	Lettuce field - lettuce harvested by field worker	13	2,598 (10.0)	1
	Produce distribution center	27	3,726 (19.7)	80

^a Based on the BLASTX analysis against the NCBI Viral Reference Sequence database.

566 **Table 2** BLASTN output statistics for the contigs matching rotaviruses in lettuce samples

Sampling location, lettuce sample type	Genus (host(s))	Gene segment	No. of contig	Average (longest) contig length (bp)	BLASTN output statistics for the best hit		
					Average nucleotide identity (%)	Average query coverage (%)	GenBank accession no. of the match(s)
Field, Iceberg (worker harvest)	A (human)	VP1	1	476	99	99	KP006506
Produce distribution center, Iceberg	A (avian)	VP3	1	415	94	99	FJ169855
	A (human, bovine, porcine)	VP1	4	603 (1257)	96	99	GU565085, KP006506, AB853890, EF554093
		VP2	3	652 (819)	97	98	KJ411433, AB374144
		VP3	6	628 (1584)	95	97	KP006508, AB853892, JX094030, AB573081, KF500220, KF500176
		NSP1	2	884 (1402)	96	99	GU808570, AB573083
		NSP4	1	447	98	100	GU183215, KP006515
	D (avian)	VP1	1	583	86	99	GU733443
Field, Romaine (control)	C (bovine)	NSP2	1	483	98	80	AB874651
Produce distribution center, Romaine	A (avian)	VP1	1	477	91	100	FJ169853
		VP2	1	444	92	100	FJ169854
		VP3	2	311 (321)	94	92	FJ169855
		VP4	1	355	97	89	EU486959
	A (human, bovine, porcine)	VP1	7	671 (1774)	95	99	KP752910, GU565041, KC443598, EF560615, KF541281, KJ752061, AB748576, AB748574
		VP2	6	657 (1284)	97	98	KJ411433, AB748581, AB748578, JF693027, AB374144, GQ225832, HQ714474

	VP3	7	552 (1063)	95	99	AB853892, KP752912, JX094030, KP006508, FJ560906, AB573081
	VP4	2	681 (1022)	95	99	GU181281, M92986, KP752913, JX470489
	VP7	1	701	97	99	FJ422136
	NSP1	3	536 (635)	92	99	U14996
	NSP2	1	293	97	100	AB748590, LC055553, KJ532804, AB748591
	NSP3	2	643 (643)	97	100	AB853897
	NSP5	1	749	95	100	KJ584855, AB748597, AB748600
A (rhesus)	VP6	1	969	97	99	M33606, EU200803
D (avian)	VP1	3	518 (642)	86	98	GU733443
	VP2	1	579	92	99	GU733444
	VP4	1	375	82	97	GU733445
	VP6	1	539	93	99	JN034680
	NSP3	1	363	89	99	GU733449
	NSP5	1	550	79	99	GU733453
F (avian)	VP2	1	419	90	99	JQ919995
	NSP1	1	331	89	99	JQ919999
	NSP2	1	433	80	89	JQ920000
G (avian)	VP1	1	394	81	97	JN596592
	VP4	1	360	88	92	KJ752087
Murine	VP3	1	355	85	99	KJ477261

567

568

569

570

571

572

573

574

575

576

577

578

579 **Table 3** BLASTX output statistics for the contigs matching picobirnavirus, papillomavirus,

580 bocavirus and cyclovirus in lettuce samples

Sampling location, lettuce sample type	Virus	No. of contig	Average (longest) contig length (bp)	BLASTX output statistics for the best hit		
				Amino acid identity (%)	Query coverage (%)	GenBank accession no. of the match(s)
Field, Iceberg (control)	Human papillomavirus type 3	1	1128	100	15	CAA52475
	Animal picobirnavirus	1	326	56	77	AIY31267
Produce distribution center, Iceberg	Human picobirnavirus	11	643 (1287)	28-76	39-99	BAJ53295, YP_239360, AIG71989
	Animal picobirnavirus	50	1020 (2664)	25-87	38-99	AIY31293, AIY31265, AKN50624, AKN50623, AGZ93689, AHI59991, AHY84700
	Animal cyclovirus (Equine)	1	611	81	54	AKN50608
Field, Romaine (worker harvest, wash and cut)	Human picobirnavirus	3	404 (459)	48-69	56-96	BAJ53287, AHX00956
	Animal picobirnavirus	2	364 (412)	42-72	81-86	AHH44583, ACT64131
Produce distribution center, Romaine	Animal picobirnavirus	4	400 (422)	34-51	63-97	ACT64132, AHZ46150, AHZ46148, AFJ79070
	Porcine bocavirus	2	587 (702)	46-73	87-100	AGS15045, AGS15081

581

582

583

584

Figure captions

Figure 1 Overview of the relative abundance of all viral contiguous sequences (contigs) for lettuce samples collected from fields and a produce distribution center based on the types of hosts these viruses can infect. Multiple hosts include plant & eukaryotic microorganisms (*Endornaviridae*); vertebrate and invertebrate (*Iridoviridae*, *Parvoviridae*); vertebrate, invertebrate, plant and fungi (*Reoviridae*) and fungi & protozoa (*Totitviridae*)

Figure 1