Metagenomic search strategies for interactions among plants and multiple microbes

Ulrich Melcher[1]*, Ruchi Verma[1] and William L. Schneider[2]

[1] Department of Biochemistry and Molecular Biology, Oklahoma State University, Stillwater, OK, USA
[2] Foreign Disease-Weed Science Research Unit, United States Department of Agriculture – Agricultural Research Service, Fort Detrick, MD, USA

Plants harbor multiple microbes. Metagenomics can facilitate understanding of the significance, for the plant, of the microbes, and of the interactions among them. However, current approaches to metagenomic analysis of plants are computationally time consuming. Efforts to speed the discovery process include improvement of computational speed, condensing the sequencing reads into smaller datasets before BLAST searches, simplifying the target database of BLAST searches, and flipping the roles of metagenomic and reference datasets. The latter is exemplified by the e-probe diagnostic nucleic acid analysis approach originally devised for improving analysis during plant quarantine.

Keywords: e-probes, BLAST, microbial consortia, taxonomic classification, databases, EDNA

BACKGROUND

A microbe entering a plant whether transmitted by a vector, through abrasion, or by wind-driven rain encounters an environment, the phytobiome, which consists of the plant and all microbes associated with it. Much has been learned about how an individual microbe interacts with a more-or-less pristine plant (Baker et al., 1997). Yet, investigations of microbes associated with plants often reveal the presence of multiple microbes. Multiple infections with multiple viruses are increasingly being discovered (Al Rwahnih et al., 2009; Villamor and Eastwell, 2013). Multiple species of bacteria are often found in endophytic association with plants (Ding et al., 2013; Ma et al., 2013). These same virus- or bacteria-infected plants may also harbor fungi or oomycetes. Interactions between phytobiome microbes have consequences for the plant. Microbe infection often induces systemic acquired resistance (Rojas et al., 2014), an alteration of the physiological status within the plant which alters the outcomes of arrival of other microbes. Virus infection by one virus can exacerbate disease symptoms in some cases, synergistic viral disease, or reduce the effects of introduction of a second-virus, cross-protection (Palukaitis, 2011) in others. Several microbes are known to be biocontrol agents that control the proliferation of other microbes associated with plants (Santoyo et al., 2012).

Increasingly, investigators want to consider all components of the phytobiome in their analyses. To this end, we consider here approaches based on next generation sequencing (NGS) to detect microbial components of the phytobiome. NGS has enabled large-scale metagenomics, which is a gene-based study of all organisms associated with a particular sample (Rucker et al., 2013; Wang et al., 2013). Time-efficient and -effective means of examining NGS databases to identify organisms that contribute to the metagenome are needed to study multiorganism consortia. Which organisms

are associated with one another? Which organisms exclude each other?

These questions fuel a need for the taxonomic classification of NGS DNA or RNA sequence reads. Such classification is important also for various other fields of study including ecology, diagnostics, and homeland security (Macdiarmid et al., 2013). The responses of marine ecosystems to climate change and anthropogenic pollution may be revealed by studies of the changing diversities of marine microbes (Coelho et al., 2013). Understanding the importance of the presence of certain microbial species in the human microbiome fuels attempts to adjust diets to achieve the most beneficial balance of bacterial species (Cox et al., 2013; Umu et al., 2013). Taxonomic profiles of microbes are important in understanding complex human diseases, such as inflammatory bowel disease, type 2 diabetes, and obesity (Segata et al., 2012; Cani, 2013). The balances of rhizosphere microbes among phytopathogenic bacteria, plant-growth-promoting bacteria, and, bacteria that can be pathogenic to animals and humans need further investigation (Mendes et al., 2013).

In diagnostic analysis, taxonomic classification of sequence reads is particularly important in the case of diseases whose etiology is unknown or whose symptoms could be produced by multiple species of infectious agents (Bernardo et al., 2013). Novel plant viruses have been identified by metagenomic means (Al Rwahnih et al., 2009; Roy et al., 2013a). The exploration of microbes associated with archaeological remains promises to enlighten the discussion of the emergence of infectious diseases in historic and prehistoric times (Gibbons, 2013; Smith et al., 2014).

Microbe–plant interaction studies should consider third (or higher order) partners when studying binary interactions of microbes with plants. The first step in such consideration is the now traditional metagenomic survey of the organismal

consortium including all microbes present in representative samples. This is followed by use of the sequence reads as queries of general databases, a time-consuming process. For ecological purposes, once particularly interesting organisms and their interrelations have been targeted, investigators may concentrate on the fluctuations of population sizes of particular taxa, simplifying the search, as described below.

TAXONOMIC CLASSIFICATION OF NGS READS: APPROACHES
OVERVIEW
With the advent of rapid, accurate, and less-expensive nucleic acid sequencing technology, phenomenal amounts of sequence data are being generated. The lengths of reads and the kinds and quantities of sequencing errors are characteristic of the sequencing methods (Dröge and McHardy, 2012). The rate of production for sequences is currently highest for Illumina technology, which can average 3.1×10^9 nt/h. The ability to analyze NGS data is growing also but at a much slower pace (Hunter et al., 2012), creating an analysis bottleneck in achieving many of the goals of metagenomic studies (Dröge and McHardy, 2012).

Taxonomic classification of NGS reads inherently consists of comparison of two datasets, the NGS reads, and a compilation of sequences of known taxonomic origin. The latter is frequently the non-redundant version of GenBank/EMBL/DDBJ nucleotide databases. The comparison is done typically using the BLASTn algorithm with the NGS reads as queries and the nr/nt database as target for the searches. Currently, the most typical analytical method for metagenomic data is to use sequence reads as queries of the general nucleotide databases to find the best matches to each query, followed by a taxonomic assignment of the read to an organism using software, such as MEGAN, Darkhorse, or Kirsten (Teeling and Glockner, 2012).

Four approaches to closing the gap between the generation of sequence reads and their analysis are being pursued: further improvement in computational speed; condensing the NGS reads dataset; simplifying the known sequence dataset; and flipping the roles of the two datasets. These are discussed below.

COMPUTATIONAL SPEED
Computational speed can be enhanced by using multiple compute nodes. However, facilities offering massively parallel computing are often not available at the location of the sequence generation unit. Thus, reads need to be transported to the computing unit either using large bandwidth communications or physically, by sending high-capacity hard drives. In addition, speed can be increased by breaking the total pool of reads into multiple subpools. The fragmentation may remove overlap possibilities, a process that could lead assembly into a non-justified sequence recombination. Considerable acceleration of taxonomic assignment at the generic level (and at the species level with lower sensitivity) can be obtained by restricting searches to finding only complete matches to k-mer words, as implemented in Kraken (Wood and Salzberg, 2014).

CONDENSING THE NGS READS DATASET
The sequences can first be subjected to an assembly process and the resulting contigs can be queries in BLASTn searches.

Assemblers such as Genovo, MetaIDBA, MetaVelvet, and MAP are used, but do themselves take considerable time to finish the assemblies of large datasets (Pell et al., 2012). Since fewer, but longer, sequences are used, such searches may be faster than searching with the raw data. However, the time required in assembly and the hazards of misassembly may negate the advantage. In addition, most assembly methods require a filtering of the read data to remove low-abundance reads which may come from minor community components. Recently, the use of graph theory on short k-mers using a Bloom filter was proposed (Pell et al., 2012) and reduced the memory requirements for large assemblies of metagenomic data and did not require the discarding of reads.

Another approach to simplifying sequence datasets is the use of bioinformatic or molecular approaches in pre-sequencing or post-sequencing steps that enrich for pathogen- or microbe-related sequences (Melcher et al., 2008). For example, multiple researchers have utilized the pool of small RNAs (sometimes called the degradome) as a target pool of nucleic acids that are enriched for viral sequences via plant defense responses (Donaire et al., 2009; Kreuze et al., 2009; Pantaleo et al., 2010; Kashif et al., 2012; Li et al., 2012; Loconsole et al., 2012; Roy et al., 2013a,b). Roy et al. (2013a,b) added subtractive bioinformatics approaches to the degradome sequence data to significantly reduce and simplify a metagenomic dataset for detection and assembly of complete genomes of plant viruses.

SIMPLIFYING KNOWN SEQUENCE DATASETS
Segata et al. (2012) have explored a strategy (MetaPhlAn) in which sets of marker genes specific to species or higher level taxa are placed in a database that is only 4% the size of the nr database. Their search strategy maps the reads to this reduced set of sequences without the prior assembly of the reads. It yields abundances of known organisms and does not need prior filtering to remove errors and does not require annotation of reads. Reads can be assigned at 450/s. An alternative, Phymm, is to generate oligonucleotides characteristic of specific taxonomic groupings by interpolated Markov models. In a strategy similar to MetaPHlAn, but less rapid, Phymm can be coupled to BLAST (PhymmBL; Brady and Salzberg, 2011). A preclassification of database targets according to k-mer word contents enhances speed (available in USEARCH, Edgar, 2010) by preventing exhaustive further searching once a good hit has been found. As a result, in effect, each query searches a less than full database. A condensed database consisting of the taxonomically most informative 18 or 20 k-mers from the raw genome database and associated with their NCBI taxonomic identifiers has also been constructed and used to speed analysis in Livermore Metagenomics Analysis Toolkit (Ames et al., 2013), which uses k-mer matching as in Kraken.

Protein sequence databases can be substituted for the nucleotide sequence databases, in which case a BLAST search will utilize the BLASTx option (Zhao et al., 2012). Alternatively, databases of conserved protein sequences, such as Pfam, have been searched with translated queries using a Hidden Markov Model in software tools such as CARMA (Krause et al., 2008) and Treephyler (Schreiber et al., 2010). Alphabet reduction (Zhao et al., 2012; Huson and Xie, 2014) can further accelerate the amino

acid sequence approaches. In these cases, non-coding sequences would be prevented from taking part in the taxonomic assignment of sequences and nucleotide variations, often important for finer taxonomic discriminations, are lost.

Whether protein or nucleotide sequence target databases are used, analysis of metagenomic datasets by BLAST search using NGS data as query is time consuming. With the large numbers of sequences currently being added to the databases, the prospects are for query times to lengthen rather than shorten for all. An additional problem for plant-based metagenomics is the likely presence of uncharacterized microbes of all types. Research on the human microbiome is aided by recent careful studies and characterization of pathogens and symbionts. There are virtually no data describing the microbiomes of plants in their many natural environments.

Martin et al. (2012) approached the problem by restricting their spectrum of organisms whose sequences were to be the targets of comparison. The whole genomes of the chosen targets for the human microbiome project were used as reference genomes against which six alignment programs mapped the reads (Martin et al., 2012). In another approach (Liu et al., 2011), the target sequence dataset was reduced considerably by focusing on a carefully chosen set of 31 marker genes that allow higher level taxonomic assignment. The reads were then mapped against these marker genes. The resulting Metaphyler software accomplished assignment in 8 h as opposed to 34 days for assignment using BLASTn and MEGAN.

FLIPPING THE SEARCH

Concerns for plant biosecurity motivated the development of e-probe diagnostic nucleic acid analysis (EDNA; Stobbe et al., 2013). For plant biosecurity, it is important to know that particularly hazardous organisms are not present in materials imported across borders (Macdiarmid et al., 2013). For example, Race 3 biovar 2 of *Ralstonia solanacearum* is thought to have entered the United States on imported geranium plants (Kim et al., 2002). Plant biosecurity includes not only invasion of pathogens from abroad but also internal bioterrorist attacks. A prime defense against such bioterrorism is an excellent microbial forensics ability (Fletcher et al., 2010a,b). The microbial profiles of crime scene objects should clarify which objects are associated with the crime and should lead to comparisons with objects in a suspect's hands (Smith, 2007). In plant biosecurity: the question asked is: which, if any, of a list of pathogenic organisms of concern are present.

EDNA simplifies answering these questions by presenting a complete reversal of the current standard procedure of operation. Instead of using the NGS sequences as queries of the ever-expanding general database, the NGS sequences are formatted to a BLAST searchable database, to be queried with panels of pretested probes specific for whatever taxonomic level is desired (Stobbe et al., 2013). By comparison of the target organism's sequence with that of near relatives, a set of oligonucleotide sequences of a specified length is generated and tested for specificity against a general database. The surviving sequences are designated as "e-probes." Such probes and their reversals, designated "decoy probes," are used in BLASTn searches of

unassembled, non-quality-checked metagenomic sequence reads formatted in a BLAST database. E-probes have been designed for a selected group of bacteria, viruses, fungi, and oomycetes. E-probe lengths of 80 or more nucleotides gave good discriminatory power. Statistical tests for comparing the results of e-probe searches with those of decoy-probe searches were devised to provide confidence levels in an identification of presence or absence of the target in the NGS dataset. EDNA analysis required no assembly or filtering, considered all portions of the NGS data (10–20 Mnt) and took only minutes to run on a typical laptop. EDNA was initially developed to aid in screening plant materials coming into quarantine for the presence or absence of pathogenic microorganisms of concern. It has applications also in phytopathological diagnostics. For example, metagenomes from three diseased plants were prepared and screened with plant virus electronic probes, resulting in the identification of a potexvirus in one of the samples and allowing further investigation of whether this virus had produced the disease (Stobbe, 2013).

EDNA is well suited to association–dissociation studies, and revealing endosymbionts and commensals. EDNA suffers from the requirement that the investigator needs to know, not only which organisms should be tested for, but also the nucleotide sequences of at least a large part of the genomes of those organisms, for the design of e-probes. However, it may be possible to design e-probe sets that recognize sequences specific at higher taxonomic levels than species. Such e-probe sets may lead to the recognition of previously unknown microbes, but only if they are related to known microbes. The design of e-probes that distinguish among viral strains has been demonstrated (Stobbe et al., 2014).

CONCLUSION

The understanding of microbe–plant interactions will be improved by the knowledge of how multiple microbes interact with each other and with their hosts. NGS has the potential to generate such knowledge but requires computational improvement to accelerate the discovery process. The development of multiple strategies to produce such improvements portends adoption of NGS as a major tool for phytobiome exploration. The strategies include increasing computing speed, condensing the NGS sequence dataset, enriching for microbe sequence, simplifying known sequence datasets and changing the direction of BLAST searches. The latter, a property of the EDNA strategy, using e-probes in BLAST searches, has the potential of assisting investigation of interactions of multiple microbes with each other and the plant.

Clearly, dissection of the molecular details of multimicrobe interactions with plants will require experimentation on model systems with known combinations of microbes in green houses and growth chambers. On the other hand, knowing which multimicrobe–plant interactions are in need of investigation can best be facilitated by a metagenomic approach that correlates the presence of specific sets of microbes with physiological and developmental phenotypes in field-grown crops or in naturally growing non-cultivated stands of plants.

AUTHOR CONTRIBUTIONS

Ulrich Melcher provided the concept for the article and created a draft; Ruchi Verma and William L. Schneider contributed improvements to the draft; William L. Schneider is the originator of the EDNA concept discussed in this article. All authors have contributed to the revision and editing of the article and approved its submission.

ACKNOWLEDGMENTS

REFERENCES

Al Rwahnih, M., Daubert, S., Golino, D., and Rowhani, A. (2009). Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387, 395–401. doi: 10.1016/j.virol.2009.02.028

Ames, S. K., Hysom, D. A., Gardner, S. N., Lloyd, G. S., Gokhale, M. B., and Allen, J. E. (2013). Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* 29, 2253–2260. doi: 10.1093/bioinformatics/btt389

Baker, B., Zambryski, P., Staskawicz, B., and Dinesh-Kumar, S. P. (1997). Signaling in plant-microbe interactions. *Science* 276, 726–733. doi: 10.1126/science.276.5313.726

Bernardo, P., Albina, E., Eloit, M., and Roumagnac, P. (2013). Pathology and viral metagenomics, a recent history. *Med. Sci.* (*Paris*) 29, 501–508. doi: 10.1051/medsci/2013295013

Brady, A., and Salzberg, S. (2011). PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nat. Methods* 8, 367. doi: 10.1038/nmeth0511-367

Cani, P. D. (2013). Gut microbiota and obesity: lessons from the microbiome. *Brief. Funct. Genomics* 12, 381–387. doi: 10.1093/bfgp/elt014

Coelho, F., Santos, A. L., Coimbra, J., Almeida, A., Cunha, A., Cleary, D. F. R., et al. (2013). Interactive effects of global climate change and pollution on marine microbes: the way ahead. *Ecol. Evol.* 3, 1808–1818. doi: 10.1002/ece3.565

Cox, M. J., Cookson, W., and Moffatt, M. F. (2013). Sequencing the human microbiome in health and disease. *Hum. Mol. Genet.* 22, R88–R94. doi: 10.1093/hmg/ddt398

Ding, T., Palmer, M., and Melcher, U. (2013). Community terminal restriction fragment length polymorphisms reveal insights into the diversity and dynamics of leaf endophytic bacteria. *BMC Microbiol.* 13:1. doi: 10.1186/1471-2180-13-1

Donaire, L., Wang, Y., Gonzalez-Ibeas, D., Mayer, K. F., Aranda, M. A., and Llave, C. (2009). Deep-sequencing of plant viral small RNAs reveals effective and widespread targeting of viral genomes. *Virology* 392, 203–214. doi: 10.1016/j.virol.2009.07.005

Dröge, J., and McHardy, A. C. (2012). Taxonomic binning of metagenome samples generated by next-generation sequencing technologies. *Brief. Bioinform.* 13, 646–655. doi: 10.1093/bib/bbs031

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Fletcher, J., Barnaby, N. G., Burans, J. P., Melcher, U., Nutter, F. W. Jr., Thomas, C., et al. (2010a). "Forensic plant pathology," in *Microbial Forensics*, 2nd Edn, eds B. Budowle, S. E. Schutzer, R. G. Breeze, P. S. Keim, and S. A. Morse (Amsterdam: Elsevier), 89–105.

Fletcher, J., Luster, D. G., Melcher, U., and Sherwood, J. L. (2010b). "Microbial forensics and plant pathogens: attribution of agricultural crime," in *Wiley Handbook of Science and Technology for Homeland Security*, ed. J. Voeller (New York: Wiley & Sons), 1880–1894.

Gibbons, A. (2013). The thousand-year graveyard. *Science* 342, 1306–1310. doi: 10.1126/science.342.6164.1306

Hunter, C. I., Mitchell, A., Jones, P., McAnulla, C., Pesseat, S., Scheremetjew, M., et al. (2012). Metagenomic analysis: the challenge of the data bonanza. *Brief. Bioinform.* 13, 743–746. doi: 10.1093/bib/bbs020

Huson, D. H., and Xie, C. (2014). A poor man's BLASTX – high-throughput metagenomic protein database search using PAUDA. *Bioinformatics* 30, 38–39. doi: 10.1093/bioinformatics/btt254

Kashif, M., Pietila, S., Artola, K., Jones, R. A. C., Tugume, A. K., Makinen, V., et al. (2012). Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. *Plant Dis.* 96, 1430–1437. doi: 10.1094/pdis-03-12-0268-re

Kim, S. H., Olson, R. N., and Schaad, N. (2002). *Ralstonia solanacearum* Biovar 2, Race 3 in geraniums imported from Guatemala to Pennsylvania in 1999. *Plant Dis.* 92, S42.

Krause, L., Diaz, N. N., Goesmann, A., Kelley, S., Nattkemper, T. W., Rohwer, F., et al. (2008). Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res.* 36, 2230–2239. doi: 10.1093/nar/gkn038

Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., et al. (2009). Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388, 1–7. doi: 10.1016/j.virol.2009.03.024

Li, R. G., Gao, S., Hernandez, A. G., Wechter, W. P., Fei, Z. J., and Ling, K. S. (2012). Deep sequencing of small RNAs in tomato for virus and viroid identification and srain differentiation. *PLoS ONE* 7:e37127. doi: 10.1371/journal.pone.0037127

Liu, B., Gibbons, T., Ghodsi, M., Treangen, T., and Pop, M. (2011). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* 12:S4. doi: 10.1186/1471-2164-12-s2-s4

Loconsole, G., Onelge, N., Potere, O., Giampetruzzi, A., Bozan, O., Satar, S., et al. (2012). Identification and characterization of Citrus yellow vein clearing virus, a putative new member of the genus Mandarivirus. *Phytopathology* 102, 1168–1175. doi: 10.1094/phyto-06-12-0140-r

Ma, B., Lv, X. F., Warren, A., and Gong, J. (2013). Shifts in diversity and community structure of endophytic bacteria and archaea across root, stem and leaf tissues in the common reed, *Phragmites australis*, along a salinity gradient in a marine tidal wetland of northern China. *Antonie Van Leeuwenhoek* 104, 759–768. doi: 10.1007/s10482-013-9984-3

Macdiarmid, R., Rodoni, B., Melcher, U., Ochoa-Corona, F., and Roossinck, M. (2013). Biosecurity implications of new technology and discovery in plant virus research. *PLoS Pathog.* 9:e1003337. doi: 10.1371/journal.ppat.1003337

Martin, J., Sykes, S., Young, S., Kota, K., Sanka, R., Sheth, N., et al. (2012). Optimizing read mapping to reference genomes to determine composition and species prevalence in microbial communities. *PLoS ONE* 7:e36427. doi: 10.1371/journal.pone.0036427

Melcher, U., Muthukumar, V., Wiley, G. B., Min, B. E., Palmer, M. W., Verchot-Lubicz, J., et al. (2008). Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from *Ambrosia psilostachya*. *J. Virol. Methods* 152, 49–55. doi: 10.1016/j.jviromet.2008.05.030

Mendes, R., Garbeva, P., and Raaijmakers, J. M. (2013). The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol. Rev.* 37, 634–663. doi: 10.1111/1574-6976.12028

Palukaitis, P. (2011). The road to RNA silencing is paved with plant–virus interactions. *Plant Pathol. J.* 27, 197–206. doi: 10.5423/ppj.2011.27.3.197

Pantaleo, V., Saldarelli, P., Miozzi, L., Giampetruzzi, A., Gisel, A., Moxon, S., et al. (2010). Deep sequencing analysis of viral short RNAs from an infected Pinot Noir grapevine. *Virology* 408, 49–56. doi: 10.1016/j.virol.2010.09.001

Pell, J., Hintze, A., Canino-Koning, R., Howe, A., Tiedje, J. M., and Brown, C. T. (2012). Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13272–13277. doi: 10.1073/pnas.1121464109

Rojas, C. M., Senthil-Kumar, M., Tzin, V., and Mysore, K. S. (2014). Regulation of primary plant metabolism during plant–pathogen interactions and its contribution to plant defense. *Front. Plant Sci.* 5:17. doi: 10.3389/fpls.2014.00017

Roy, A., Choudhary, N., Guillermo, L. M., Shao, J., Govindarajulu, A., Achor, D., et al. (2013a). A novel virus of the genus Cilevirus causing symptoms similar to citrus leprosis. *Phytopathology* 103, 488–500. doi: 10.1094/phyto-07-12-0177-r

Roy, A., Shao, J., Hartung, J. S., Schneider, W. L., and Brlansky, R. H. (2013b). A case study on discovery of novel Citrus leprosis virus cytoplasmic type 2 utilizing small RNA libraries by next generation sequencing and bioinformatic analyses. *J. Data Mining Genomics Proteomics* 4, 1–6. doi: 10.4172/2153-0602.1000129

Rucker, O., Dangel, A., and Klein, H. G. (2013). Developments and insights into the analysis of the human microbiome. *J. Lab. Med.* 37, 329–335. doi: 10.1515/labmed-2013-0018

Santoyo, G., Orozco-Mosqueda, M. D., and Govindappa, M. (2012). Mechanisms of biocontrol and plant growth-promoting activity in soil bacterial species of *Bacillus* and *Pseudomonas*: a review. *Biocontrol Sci. Technol.* 22, 855–872. doi: 10.1080/09583157.2012.694413

Schreiber, F., Gumrich, P., Daniel, R., and Meinicke, P. (2010). Treephyler: fast taxonomic profiling of metagenomes. *Bioinformatics* 26, 960–961. doi: 10.1093/bioinformatics/btq070

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9:811–814. doi: 10.1038/nmeth.2066

Smith, J. (2007). Microbial forensics and Ag biosecuirty: a national priority. *Vanguard* 2007, 10–13.

Smith, O., Clapham, A., Rose, P., Liu, Y., Wang, J., and Allaby, R. G. (2014). A complete ancient RNA genome: identification, reconstruction and evolutionary history of archaeological Barley Stripe Mosaic Virus. *Sci. Rep.* 4, 4003. doi: 10.1038/srep04003

Stobbe, A. (2013). *Virus Detection in a Metagenomic Sequence Dataset: Methods and Applications.* Ph.D. thesis, Oklahoma State University, Stillwater, OK.

Stobbe, A. H., Daniels, J., Espindola, A., Verma, R., Melcher, U., Ochoa-Corona, F., et al. (2013). E-probe Diagnostic Nucleic acid Analysis (EDNA): a theoretical approach for handling of next generation sequencing data for diagnostics. *J. Microbiol. Methods* 94, 356–366. doi: 10.1016/j.mimet.2013.07.002

Stobbe, A. H., Schneider, W. L., Hoyt, P. R., and Melcher, U. (2014). Screening metagenomic data for viruses using the e-probe diagnostic nucleic acid assay (EDNA). *Phytopathology* (in press).

Teeling, H., and Glockner, F. O. (2012). Current opportunities and challenges in microbial metagenome analysis-a bioinformatic perspective. *Brief. Bioinform.* 13, 728–742. doi: 10.1093/bib/bbs039

Umu, O. C. O., Oostindjer, M., Pope, P. B., Svihus, B., Egelandsdal, B., Nes, I. F., et al. (2013). Potential applications of gut microbiota to control human physiology. *Antonie Van Leeuwenhoek* 104, 609–618. doi: 10.1007/s10482-013-0008-0

Villamor, D. E. V., and Eastwell, K. C. (2013). Viruses associated with rusty mottle and twisted leaf diseases of sweet cherry are distinct species. *Phytopathology* 103, 1287–1295. doi: 10.1094/phyto-05-13-0140-r

Wang, J., McLenachan, P. A., Biggs, P. J., Winder, L. H., Schoenfeld, B. I. K., Narayan, V. V., et al. (2013). Environmental bio-monitoring with high-throughput sequencing. *Brief. Bioinform.* 14, 575–588. doi: 10.1093/bib/bbt032

Wood, D. E., and Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 15, R46. doi: 10.1186/gb-2014-15-3-r46

Zhao, Y., Tang, H., and Ye, Y. (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next generation sequencing data. *Bioinformatics* 28, 125–126. doi: 10.1093/bioinformatics/btr595