



Plant virus metagenomics: what we know and why we need to know more

Anthony H. Stobbe¹ and Marilyn J. Roossinck^{1,2*}

¹ Department of Plant Pathology and Environmental Microbiology, Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA

² State Agricultural Biotechnology Centre, Murdoch University, Perth, WA, Australia

*Correspondence: mjr25@psu.edu

Edited by:

Jeff H. Chang, Oregon State University, USA

Reviewed by:

Ming-Bo Wang, Commonwealth Scientific and Industrial Research Organisation, Australia

Johan Burger, Stellenbosch University, South Africa

Keywords: ecogenomics, virus ecology, wild plant viruses, virus biodiversity, bioinformatics

INTRODUCTION

In the past decade the concept of plant viruses as strictly disease-causing entities has been challenged. While the most well-studied and obvious interactions between plants and viruses are related to disease, there are several examples of mutualistic relationships between plants and viruses, both indirect and direct. These mutualistic interactions have not been fully explored, and many questions remain unanswered. One problem is the lack of knowledge of plant viruses in nature. Metagenomic surveys have estimated that only a small fraction of virus species are known. Additionally, globalization has led to the increased movement of plant material and virus movement. As viruses move from one area to another, new potential hosts offer the possibility of new interactions, both negative and positive.

BENEFICIAL PLANT-VIRUS INTERACTIONS

Viruses have been associated with plant disease since they were first described in 1898 (Beijerinck, 1898), but in recent years viruses with positive impacts on the plant hosts they are associated with also have been described. Negative interactions are mostly studied as disease symptoms such as stunting or necrosis, and the vast majority of virus research has focused on the disease aspect of these interactions. Beneficial interactions involve environmental protection to the host plant, protection against other pathogens, or control of plant responses to nutritional needs (reviewed in Roossinck, 2011). Plant viruses confer drought and cold tolerance

to plants as conditional mutualists: the plant is harmed by the viruses under normal conditions, but benefited under extreme conditions. This was demonstrated for several different viruses and plant hosts (Xu et al., 2008). Mild strains of plant viruses protect plants from more severe isolates, a phenomenon known as cross-protection (Fraser, 1998) that led to the initial generation of virus-induced pathogen protection in transgenic plants. Endogenous pararetroviral elements in plants can confer resistance to exogenous viruses (Staginnus et al., 2007). The coat protein gene of a persistent virus in white clover affects the development of nodules under varying nitrogen levels, and this could be transferred to other legumes (Nakatsukasa-Akune et al., 2005). *Curvularia* thermal tolerance virus is a mycovirus that infects a plant fungal endophyte, *Curvularia protuberata*. When both virus and fungus are present in hot springs panic grass (*Dichanthelium lanuginosum*) the holobiont is able to grow in soil temperatures up to 65°C (Márquez et al., 2007). Many more examples of mutualistic viruses can be found in other hosts (Roossinck, 2011). In addition, viruses are important in population control of their hosts, and marine viruses are probably extremely important to the movement of carbon and trace elements in the microbiome of the oceans (Danovaro et al., 2011).

PLANT VIRUS ECOLOGY AND EVOLUTION

The existence of plant-virus mutualistic relationships should not be surprising

when one considers the numerous examples of mutualistic relationships between plants or animals and other microbes. Despite examples, there has been very little focus on exploring mutualistic relationships among plants and viruses. Viruses are also involved in the complex interactions between plants and insects, and can alter insect feeding behavior, fecundity, and ability to invade new territory (reviewed in Roossinck, 2013).

Further complicating our understanding of plant-virus interactions is the role globalization has on the relationships between viruses and their plant hosts. Viruses are not stationary, and their movement geographically and between host species can have drastic effects on the ecology of a given area. Climate change can alter the behavior of many virus vectors, promoting the spread of viral distribution across a larger geographic area (Lebouvier et al., 2011). A prime example of the impact of viruses on plant species balance is the well-studied beneficial effect the luteoviruses *Barley yellow dwarf virus* and *Cereal yellow dwarf virus* had on the invasive annual grasses in California grasslands (Malmstrom et al., 2005).

Using the estimation that over 20,000 microbes have invaded the United States (Pimentel et al., 2005) as a measure of virus movement worldwide, it is reasonable to suggest that this significant movement of viruses gives the opportunity of a virus jumping from one plant host species to another, which in turn leads to new plant-virus interactions. Metagenomic surveys can be useful for nations who are interested in

protecting their crops against invasive diseases (MacDiarmid et al., 2013). A working knowledge of the geographic location, host range, and potential effects of plant viruses can assist such nations in developing effective policies that distinguish those viruses that will have negative economic impacts from viruses which are benign or even beneficial.

Viruses impact the evolution of plants at many levels, and plants clearly affect the populations of viruses that infect them. There are examples of specific interactions between plants and viruses, such as the silencing suppression genes found in many RNA viruses. While not as prevalent as in animals or bacteria, there have been instances of horizontal gene exchange from viruses to plants. Repeat sequences of geminiviruses have been found in *Nicotiana spp.* (Bejarano et al., 1996; Ashby et al., 1997), and pararetroviruses are frequently found integrated into plants (Hohn et al., 2008). Sequences from cytoplasmic RNA viruses are found in plant genomes (Liu et al., 2010; Chiba et al., 2011). There is also evidence that Closterovirus genes have integrated into the mitochondria of grapevines, *Vitis vinifera* (Goremykin et al., 2009). Viruses have been said to be responsible for a large amount of genetic flow in several different systems (Bock, 2010; Liu et al., 2010; Wu and Zhang, 2011). This in turn would increase the genetic plasticity of the hosts offering the opportunity for novel interactions to take place.

WE DON'T KNOW WHAT IS OUT THERE

In the past decade there have been a few metagenomic type surveys exploring plant virus biodiversity in wild plants, insects, and a few other environments (Wren et al., 2006; Roossinck et al., 2010; Ng et al., 2011; Roossinck, 2012). Some of these studies have used a more ecological approach, “ecogenomics,” that looks at the viral populations in individuals rather than in the entire environment that is typical of metagenomic studies (Roossinck et al., 2010). The most surprising result is that we know very little of the size and diversity of plant virus families. These surveys have revealed that the true diversity of virus species is much larger than earlier estimates, with the discovery of new virus isolates, species, families, and

even higher level virus groups (Labonté and Suttle, 2013). An additional surprising result is that viruses in wild plants do not cause any visible symptoms. With the knowledge of how little we know of the biodiversity of viruses, new techniques, methods, and questions need to be developed in order to detect and identify these new viruses. In addition, the full extent of plant-virus interactions cannot be fully studied until we have a better understanding of the ecology of plant viruses. While the metagenomic surveys are a start, there are still many challenges ahead.

The viruses found using metagenomic sequencing data can be described in three different ways: (1) Known-knowns: virus species or isolates that are already known to be in the environment being surveyed; (2) Unknown-knowns: new virus species or isolates of a known family, or known viruses that have not been found previously in the surveyed environment and; (3) Unknown-unknowns: viruses that are completely novel and share little to no sequence similarity with other known viruses. Sequencing data for each instance can be analyzed differently based on the questions being addressed. The removal of non-viral sequences from the sample either before or after sequencing will, of course, increase the chances of identifying viruses within a metagenomic sequence dataset, so care should be taken in both sample preparation before sequencing and manipulation of sequence data after sequencing. Methods to enrich for plant viral-specific sequences include isolation of virus-like particles (Muthukumar et al., 2009), enrichment for double-stranded RNA (Roossinck et al., 2010), and the use of siRNAs (Kreuze et al., 2009). All of these methods have strengths and weaknesses, but the use of double-stranded RNA has given the deepest analyses so far. For known-knowns and unknown-knowns, screening of the sequence dataset for the presence of known viruses can drastically reduce the amount of time needed for analysis and as such detection and identification of viruses (Stobbe et al., 2013).

The large amount of sequencing data that shares little or no nucleotide similarity with known sequences in curated

databases such as GenBank suggests that there are still many unknown microbes that have yet to be described, with unknown-unknown viruses likely to be prevalent among them. These unknown sequences continue to be difficult to identify and will require new and novel methods to assign to a taxon. There have been significant efforts to describe these unknown-unknowns in different environments, with new analysis methods tailored for virus discovery either by sequence similarity or by clustering genes (Williamson et al., 2008; Kristensen et al., 2010; Wu et al., 2010; Ames et al., 2013; Labonté and Suttle, 2013). Protein sequences expressed by viral-specific genes, such as the RNA dependent RNA polymerase (RdRp), can be used to detect both unknown-knowns and unknown-unknowns (Kristensen et al., 2010).

Unknown-unknown viruses share little or no sequence similarity with known viruses, so quality *de novo* assembly is essential as genome mapping or BLAST assisted assembly is not an option. Assembly programs tailored for virus *de novo* assembly have been created, and modifications to assembly processes have been used to generate fully assembled genomes of extremely low titer microbes (Albertsen et al., 2013). Additionally, new exciting sequencing platforms, such as the Oxford Nanopore, offer the ability to sequence an entire viral genome in a single read (Schneider and Dekker, 2012). This ability removes the need for assembly altogether. However, even with a perfectly assembled genome, the need to identify the genome is still there. By clustering the unknown sequences, the biodiversity of a given environment can be estimated, and this has led to the discovery of new families of single stranded DNA viruses (Labonté and Suttle, 2013).

CONCLUSIONS

There is still much to be discovered on the topic of plant-microbe interactions, and of plant-virus interactions in particular. Metagenomics offers us a unique tool to elucidate the current state of viruses in plants and the role viruses play in these interactions. When these studies are done on individual plant samples rather than

pooled samples from a larger environment, a system known as “ecogenomics” (Roossinck et al., 2010), they provide meaningful data for deeper ecological analyses of the distribution of viruses and potential host- or environmental-specific interactions.

ACKNOWLEDGMENTS

The authors acknowledge the National Science Foundation grant numbers EF-0627108, EPS-0447262, IOS-0950579, and IOS-1157148 and The United States Department of Agriculture grant number OKLR-2007-01012 for previous and continuing research support, and the Pennsylvania State University.

REFERENCES

- Albertsen, M., Hugenholtz, P., Sharskewski, A., Nielsen, K. L., Tyson, G. W., and Nielsen, P. H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotech.* 31, 533–541. doi: 10.1038/nbt.2579
- 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
- Ashby, M. K., Warry, A., Bejarano, E. R., Khashoggi, A., Burrell, M., and Lichtenstein, C. P. (1997). Analysis of multiple copies of geminiviral DNA in the genome of four closely related *Nicotiana* species suggest a unique integration event. *Plant Mol. Biol.* 35, 313–321. doi: 10.1023/A:1005885200550
- Bejerinck, M. W. (1898). “Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves,” in *Phytopathological Classics*, No. 7, ed J. Johnson (St. Paul, MN: American Phytopathological Society), 33–52.
- Bejarano, E. R., Khashoggi, A., Witty, M., and Lichtenstein, C. (1996). Integration of multiple repeats of geminiviral DNA into nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci. U.S.A.* 93, 759–764. doi: 10.1073/pnas.93.2.759
- Bock, R. (2010). The give-and-take of DNA: horizontal gene transfer in plants. *Trends Plant Sci.* 15, 11–22. doi: 10.1016/j.tplants.2009.10.001
- Chiba, S., Kondo, H., Tani, A., Saisho, D., Sakamoto, W., Kanematsu, S., et al. (2011). Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS Pathog.* 7:e1002146. doi: 10.1371/journal.ppat.1002146
- Danovaro, R., Corinaldesi, C., Dell’anno, A., Fuhrman, J. A., Middelburg, J. J., Noble, R. T., et al. (2011). Marine viruses and global climate change. *FEMS Microbiol. Rev.* 35, 993–1034. doi: 10.1111/j.1574-6976.2010.00258.x
- Fraser, R. S. S. (1998). “Introduction to classical cross protection,” in *Plant Virology Protocols*, eds G. D. Foster, and S. C. Taylor (Totowa, NJ: Humana Press), 13–24. doi: 10.1385/0-89603-385-6:13
- Goremykin, V. V., Salamini, F., Velasco, R., and Viola, R. (2009). Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* 26, 99–110. doi: 10.1093/molbev/msn226
- Hohn, T., Richert-Pöggeler, K. R., Staginnus, C., Harper, G., Schwarzacher, T., Teo, C. H., et al. (2008). “Evolution of integrated plant viruses,” in *Plant Virus Evolution*, ed M. J. Roossinck (Heidelberg: Springer), 53–81.
- 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
- Kristensen, D. M., Mushegian, A. R., Dolja, V. V., and Koonin, E. V. (2010). New dimensions of the virus world discovered through metagenomics. *Trends Microbiol.* 18, 11–19. doi: 10.1016/j.tim.2009.11.003
- Labonté, J. M., and Suttle, C. A. (2013). Previously unknown and highly divergent viruses populate the oceans. *ISME J.* 7, 2169–2177. doi: 10.1038/ismej.2013.110
- Lebouvier, M., Laparie, M., Hullé, M., Marais, A., Cozic, Y., Lalouette, L., et al. (2011). The significance of the sub-Antarctic Kerguelen Islands for the assessment of the vulnerability of native communities to climate change, alien insect invasions and plant viruses. *Biol. Invasions* 13, 1195–1208. doi: 10.1007/s10530-011-9946-5
- Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., et al. (2010). Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. *J. Virol.* 84, 11879–11887. doi: 10.1128/JVI.00955-10
- 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
- Malmstrom, C. M., McCullough, A. J., Johnson, H. A., Newton, L. A., and Borer, E. T. (2005). Invasive annual grasses indirectly increase virus incidence in California native perennial bunchgrasses. *Oecologia* 145, 153–164. doi: 10.1007/s00442-005-0099-z
- Márquez, L. M., Redman, R. S., Rodriguez, R. J., and Roossinck, M. J. (2007). A virus in a fungus in a plant—three way symbiosis required for thermal tolerance. *Science* 315, 513–515. doi: 10.1126/science.1136237
- Muthukumar, V., Melcher, U., Pierce, M., Wiley, G. B., Roe, B. A., Palmer, M. W., et al. (2009). Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *J. Virol. Methods* 141, 169–173. doi: 10.1016/j.virusres.2008.06.016
- Nakatsukasa-Akune, M., Yamashita, K., Shimoda, Y., Uchiumi, T., Abe, M., Aoki, T., et al. (2005). Suppression of root nodule formation by artificial expression of the *TrEnodDRI* (coat protein of *White clover cryptic virus 2*) gene in *Lotus japonicus*. *Mol. Plant Microbe Interact.* 18, 1069–1080. doi: 10.1094/MPMI-18-1069
- Ng, T. F. F., Duffy, S., Polston, J. E., Bixby, E., Vallad, G. E., and Breitbart, M. (2011). Exploring the diversity of plant DNA viruses and their satellites using vector-enabled metagenomics on whiteflies. *PLoS ONE* 6:e19050. doi: 10.1371/journal.pone.0019050
- Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52, 273–288. doi: 10.1016/j.ecolecon.2004.10.002
- Roossinck, M. J. (2011). The good viruses: viral mutualistic symbioses. *Nat. Rev. Microbiol.* 9, 99–108. doi: 10.1038/nrmicro2491
- Roossinck, M. J. (2012). Plant virus metagenomics: biodiversity and ecology. *Annu. Rev. Genet.* 46, 357–367. doi: 10.1146/annurev-genet-110711-155600
- Roossinck, M. J. (2013). Plant virus ecology. *PLoS Pathog.* 9:e1003304. doi: 10.1371/journal.ppat.1003304
- Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., et al. (2010). Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol. Ecol.* 19, 81–88. doi: 10.1111/j.1365-294X.2009.04470.x
- Schneider, G. F., and Dekker, C. (2012). DNA sequencing with nanopores. *Nat. Biotech.* 30, 326–329. doi: 10.1038/nbt.2181
- Staginnus, C., Gregor, W., Mette, M. F., Teo, C. H., Borroto-Fernández, E. G., Machado, M. L. D. C., et al. (2007). Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. *BMC Plant Biol.* 7:24. doi: 10.1186/1471-2229-7-24
- Stobbe, A. H., Daniels, J., Espindola, A. S., 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
- Williamson, S. J., Rusch, D. B., Yooseph, S., Halpern, A. L., Heidelberg, K. B., Glass, J. I., et al. (2008). The Sorcerer II global ocean sampling expedition: metagenomic characterization of viruses within aquatic microbial samples. *PLoS ONE* 3:e1456. doi: 10.1371/journal.pone.0001456
- Wren, J. D., Roossinck, M. J., Nelson, R. S., Sheets, K., Palmer, M. W., and Melcher, U. (2006). Plant virus biodiversity and ecology. *PLoS Biol.* 4:e80. doi: 10.1371/journal.pbio.0040080
- Wu, D.-D., and Zhang, Y.-P. (2011). Eukaryotic origin of a metabolic pathway in virus by horizontal gene transfer. *Genomics* 98, 367–369. doi: 10.1016/j.ygeno.2011.08.006
- Wu, Q., Luo, Y., Lu, R., Lau, N., Lai, E. C., Li, W.-X., et al. (2010). Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1606–1611. doi: 10.1073/pnas.0911353107

Xu, P., Chen, F., Mannas, J. P., Feldman, T., Sumner, L. W., and Roossinck, M. J. (2008). Virus infection improves drought tolerance. *New Phytol.* 180, 911–921. doi: 10.1111/j.1469-8137.2008.02627.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 February 2014; paper pending published: 05 March 2014; accepted: 29 March 2014; published online: 16 April 2014.

Citation: Stobbe AH and Roossinck MJ (2014) Plant virus metagenomics: what we know and why we need to know more. Front. Plant Sci. 5:150. doi: 10.3389/fpls.2014.00150

This article was submitted to Plant Genetics and Genomics, a section of the journal Frontiers in Plant Science.

Copyright © 2014 Stobbe and Roossinck. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.