USING HIGH-THROUGHPUT TECHNOLOGIES TO ADVANCE OUR UNDERSTANDING ON POTATO VIRUS Y INFECTIONS

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Washington Luís da Silva

May 2018

USING HIGH-THROUGHPUT TECHNOLOGIES TO ADVANCE OUR UNDERSTANDING ON POTATO VIRUS Y INFECTIONS

Washington Luis da Silva, Ph.D.

Cornell University 2018

ABSTRACT

Potato virus Y (PVY) is a major virus pathogen of potato worldwide. Surveys indicate that recombinant strains of PVY have emerged in recent years to predominate in the U.S. potato crop and that the genetic diversity among and within PVY strains is prodigious. Vegetative propagation of potato via tubers allows PVY to survive year to year and to be transported over long distances. Whereas, aphids are primarily responsible for spread of PVY within the crop and over regional distances. Furthermore, the tuber necrotic strain (PVY^{NTN}) is associated with potato tuber necrotic ringspot disease (PTNRD), a tuber deformity that negatively impacts tuber quality, marketability, and poses a serious threat to seed and commercial potato production worldwide.

To map loci that influence tuber and foliar symptoms in potatoes infected with PVY^{NTN} and the length of tuber dormancy, two potato populations were genotyped with a potato SNP chip. QTL analyses revealed major-effect QTLs in a Waneta x Pike cross for mosaic on chromosomes 4 and 5, and for PTNRD and for foliar-necrosis symptoms on chromosomes 4 and 5, respectively. QTLs for dormancy were detected on chromosomes 2, 3, 5, 6, and 8, in a Waneta x Superior cross. Locating QTLs associated with PVY-related symptoms and tuber dormancy provides a

framework for breeders to develop varieties with resistance to multiple PVY-symptoms and to manipulate tuber dormancy length.

Illumina high-throughput sequencing was used to study the *quasispecies* diversity of 15 isolates representing seven different PVY strains and to investigate how transmission modes (insect and mechanical) are contributing to the evolution and diversification of PVY. Eight were PVY^N isolates and six of those came from the same geographic region in two different years. A consensus sequence, without indels or insertions, was successfully extracted from the sequenced reads of each isolate after being mapped to the strain reference genome. PVY^N had a higher population genetic diversity than any other strain evaluated and the population genetic diversity of the PVY^N isolates, differed between collection years and sites. We found that the population diversity of PVY varies by the virus strain but doesn't differ among transmission modes. Our data suggest that each transmission mode exerts unique selection pressures on the virus population and allows different mutations to accumulate and become fixed. Knowledge of how rapidly PVY can evolve and of the factors driving PVY *quasispecies* diversity could be used to enhance the efficiency of PVY management practices in potato fields.

BIOGRAPHICAL SKETCH

Washington Luís da Silva was born in the city of Contagem, in Minas Gerais state, Brazil. When he was five years old, he and his family moved to a farm in the small village of Divinolândia de Minas, in Minas Gerais state, where he attended primary and secondary school. In March 2003, he started his undergraduate major in Agronomy Engineering at Universidade Federal de Viçosa (UFV) and promptly started working with ecology and behavior of leaf-cutting ants. He was awarded a major scholarship from the Programa Institucional de Bolsas de Iniciação Científica (Institutional Program of Scientific Initiation, Minas Gerais State Science Scholarship) - FAPEMIG PROBIC/UFV. Soon thereafter, he was awarded another scholarship from the Programa Institucional de Bolsas de Iniciação Científica (Institutional Program of Scientific Initiation, National Brazilian Science Scholarship) - PIBIC/CNPq.

From March 2007 to September 2008 he participated in an exchange program in the United States, MAST international program at the University of Minnesota, and was stationed at the Mahoney's Garden Center in Winchester Massachusetts to learn about garden plant cultivation and commercialization in the United States. During this time, he also worked as a museum guide volunteer at The Harvard Museum of Natural History and he took classes on the taxonomy, anatomy, and evolution of *Cetacea*. In February of 2009, he resumed his classes at UFV and started working as a research assistant with fluorescence microscopy studying the effects of silicon on the defense system of wheat against infection by *Pyricularia grisea*. Washington was awarded another scholarship from the Programa Institucional de Bolsas de Iniciação Científica (Institutional Program of Scientific Initiation, National Brazilian Science Scholarship) - PIBIC/CNPq.

He completed his bachelor of science from UFV in July 2010 and in August of that year he joined the Department of Plant Pathology and Crop Physiology at Louisiana State University to pursue his Master's degree in Plant Health under the guidance of Professor Christopher A. Clark. His Master's project focused on studying the etiology and biology of end rot and soft rot disease complexes in sweetpotato storage roots. He served as the treasurer and also member of the executive committee of the Plant Pathology and Crop Physiology Graduate Student Association for two years. He was also a member of the Committee of Courses and Curriculum of the Plant Pathology and Crop Physiology Department from 2011 to 2012. During his Master's studies, he received four travel awards, two from The American Phytopathological Society (APS), one from the Department of Plant Pathology and Crop Physiology at LSU, and one from the LSU graduate school, to present his research results at APS professional meetings. Washington and his wife Rachel, whom he married in March 2009, had their first daughter, Aurelia, in January of 2013 while attending graduate school.

Washington started his Ph.D. studies at Cornell University in June 2013 in the Plant Pathology & Plant-Microbe Biology Section of the School of Integrative Plant Science. Under the guidance of Professor Stewart Gray, Washington migrated from fungi and bacteria study to the world of viruses and devoted his time as a Ph.D. student to investigate how transmission mechanisms (insect, vegetative propagation through tubers, and mechanical) contribute to the evolution and diversification of *Potato virus Y*. In a collaborative effort involving researchers from Cornell University, USDA-ARS, Michigan State University, Biomathematics and Statistics Scotland, and The James Hutton Institute, he also conducted studies looking for QTLs for potato tuber dormancy breakage and for tuber and foliar symptoms caused by PVY in autotetraploid potato crosses.

Washington was an active student leader serving as treasurer and vice president of the Plant Pathology Graduate Student Association for two terms each. He was always eager to help with departmental activities and was a great ambassador for the department in several of the new graduate student recruitment programs. While at Cornell, Washington won several awards and honors, including a Cornell University Doctoral Diversity Fellowship, a National Potato Council Scholarship, and multiple travel awards from the American Phytopathological Society and Cornell Graduate School. In 2015, Washington was nominated a Cornell Graduate Dean Scholar and was also given the Freedom of the City title from his hometown, Divinolândia de Minas in Brazil. Washington is very proud of everything he has accomplished in life and extremely grateful to all of the mentors who have guided him and for the recognition he received throughout his education.

For Rachel, Aurélia, and Camila

ACKNOWLEDGEMENTS

This thesis would not have been possible without the assistance and the help of those who in one way or another donated their valuable assistance in the conduct of this study, to only some of whom it is possible to give individual mention herein.

My highest possible gratitude goes to Dr. Stewart Gray, whose sincerity and encouragement I will never forget. Dr. Gray taught me the pleasure of doing quality work, he gave me the freedom to pursue my own research interest while professionally guiding my path along the way. I have become a better communicator and a better team player under his leadership. He was there for me, academically and as a friend, giving me the support and inspiration as I hurdled all the obstacles in the completion of this important phase of my life. It has been an honor to be his student and thank you very much for the importunity and for taking a chance on me.

I would like to thank my committee members Dr. Michelle Cilia, Dr. Walter De Jong, Dr. Zhangjun Fei for their time and helpful comments. I lost count of how many times I knocked on their doors asking for help – never once I heard the word "no". They made me feel as if I was a colleague rather than a student.

I am very grateful to have had the opportunity to be part of the Cornell University Plant Pathology & Plant-Microbe Biology team. It was a prolific and exciting experience that I will carry along forever – a dream that became reality - - GO BIG RED! I am indebted and grateful to the students, staff, and faculty members of this department for the precious help and support during my Ph.D. studies.

I would like to acknowledge the academic and technical support of Cornell University, which was first class. My special thanks to the United States Department of Agriculture (USDA) Specialty Crop Research Initiative (SCRI), the UK Biotechnology and Biological Sciences Research Council (BBSRC), the University of New York Diversity Fellowship program, and the National Potato Council for the financial support, which made my time at Cornell possible.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Potato (*Solanum tuberosum* L.) is ranked the third staple food in the world and *Potato virus Y* (PVY) is the most important disease problem in seed potato production. That is not due to crop loss or yield reduction, but because seed certification programs require that seed lots meet strict minimal PVY infection tolerances. This can jeopardize the livelihood of seed potato growers as rejected lots cannot be replanted for seed production and, in most US states, seed lots with virus incidence exceeding 5-10% can only be sold for consumption. Seed potatoes sold for commercial production or for food are much less profitable than if they are sold as seed. On the other hand, infections of PVY in commercial potato fields can affect both yield and tuber quality, resulting in yield losses of up to 80% (Quenouille et al. 2013). Therefore, it is of interest for all sectors of the potato industry to improve ways to manage PVY infection in potatoes.

PVY is the type species of the genus *Potyvirus* in the family *Potyviridae* and it was first associated with disease in potato in early 1930s (Smith 1931). Many other members of the *Solanaceae* family e.g. pepper, eggplant, tomato, and tobacco are also hosts of PVY. This virus can be transmitted by mechanical inoculation, through vegetatively propagated material (e.g. infected tubers), and by aphids in a non-persistent manner. PVY is a positive-sense, single-stranded RNA virus with a 9.7-kb genome, which expresses a single polyprotein that is subsequently cleaved by three virus-specific proteases into at least 10 mature proteins (Adams et al. 2011). A ribosomal frame shift in the *Potyvirus* genome was discovered recently that allows translation of a small ORF (PIPO), encoding an 11th known *Potyvirus* protein (P3N-PIPO) (Chung

et al. 2008). Several strains of PVY have been described worldwide, and many of them appear to be recombinants of the ordinary strain (PVY^O) with the necrotic strain (PVY^N) (Hu et al. 2009). While PVY^O and PVY^N are easily distinguished by commercially available ELISA kits, the recombinants and other parental strains require more sophisticated tests, such as multiplex RT-PCR, for their proper identification. Because PVY is transmitted in a non-persistent manner, control of aphid vectors is of limited use; thus, the main control strategies are the use of certified potato seed and breeding for resistance.

Foliar symptoms induced by PVY vary by host, environment, and by the strain of the virus. In potato, it is common to observe one or more of the following symptoms in plants infected with PVY^O; stunting, mosaic, leaf distortion, yellowing of the leaves, vein necrosis, or leaf drop. PVY infections can be asymptomatic in potato leaves infected late in the season; also, some recombinant PVY strains cause very mild to no symptoms, and some potato cultivars (e.g. Russet Norkotah, Shepody, and Silverton) lack symptom expression (Gray et al. 2010). All of these features make it difficult for inspectors to observe and estimate disease incidence. PVY is also the causal agent of the so-called potato tuber necrotic ringspot disease (PTNRD), a tuber deformity associated with infections by necrotic strains of the virus.

The seed potato industry has identified PVY as its most serious disease problem and is looking for ways to eliminate the strains of the virus that cause PTNRD. This disease poses a serious threat to the seed and commercial production industries because it negatively impacts tuber quality and reduces trade and marketing opportunities. Commercial growers are being impacted by a shortage of certified seed and a reduction in tuber quality resulting in an inability to meet contract obligations and in product rejection by processors and consumers. Research on

PVY incidence and distribution has documented the occurrence of the tuber necrotic strains of PVY in every seed producing state in US and the overall incidence of the necrotic strains has risen from 5% of the total PVY in 2006 to 18% in 2011. A similar emergence of necrotic strains of PVY was observed in Europe through the 1980s and 1990s and now those strains make up nearly 100% of the PVY affecting potato in Eastern and Western Europe. As consequence, they have lost many favored potato cultivars due to their susceptibility to PTNRD. Efforts are underway to develop appropriate management strategies to minimize PVY in seed stocks, but a long-term solution is the development of PTNRD resistant potatoes and ensuring that germplasm released from breeding programs is not susceptible to PTNRD.

In recent years there have been rapid shifts in the virus populations that have allowed several recombinant strains to become predominant in most major potato production areas (Karasev and Gray 2013). Various factors may be influencing the shift from the well-described ordinary strain, PVY^{O}, to the recombinant PVY strains. While this is a complex problem with many possible causes, some reasons are widely accepted by researchers and potato growers. It is believed that because some recombinant strains induce mild or no symptoms on foliage, some popular varieties do not show symptoms, and late season infections go undetected that these recombinant strains are able to pass unnoticed through visual disease inspection in the field, increasing their frequency. In contrast, strains that cause visible symptoms in potatoes are often detected and eliminated, which reduces their frequency. This is very true in seed potato fields, where symptomatic plants are rogued to reduce viral incidence to meet requirements for seed certification. In addition, recombinant strains have been reported to be trans-

mitted more efficiently by aphids than PVY^O (Mondal and Gray 2017). In the long run, recombinant strains are more likely to disseminate their populations than non-recombinant ones. Furthermore, current laboratory tests can be misleading as some new recombinants are not detected by the traditional tests, ELISA and multiplex RT-PCR. Consequently, these emerging strains are able to evade detection and get perpetuated.

Another potential factor for the emergence of recombinant strains is the mechanisms of virus transmission between plants. PVY is transmitted horizontally (plant to plant) by aphids and vertically (mother to daughter) by infected tubers (Barker 1994). These transmission modes can induce strong bottlenecks, which drastically reduce the number of viral genomes transferred to the new host (Gutiérrez et al. 2012). Indeed, it has been observed with zucchini yellow mosaic virus (ZYMV) that certain mutations in the viral population go to fixation in aphid-vectored plants but remain at low frequency in mechanically inoculated plants, suggesting a major fitness advantage of the aphid transmitted variants (Simmons et al. 2012). Likewise, studies of the genetic structure of PVY populations *in planta* could reveal if any PVY variants are preferentially selected in each transmission mode, possibly shedding light on the surge of PVY recombinant strains.

LITERATURE CITED

Adams, M.J., Zerbini, F.M., French, R., Rabenstein, F., Stenger, D.C., and Valkonen, J.P.T. 2011. Family *Potyviridae*. Pages 1069–1089 in: Virus taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses, A. King, M. Adams, E. Carstens, and E. Lefkowitz, eds. Elsevier, Oxford.

Ali, M.C., Maoka, T., Natsuaki, K.T., and Natsuaki, T. 2010. The simultaneous differentiation of *Potato virus Y* strains including the newly described strain PVY^{NTN-NW} by multiplex PCR assay. Journal of Virological Methods 165:15-20.

AlMaarri, K., Massa, R., and AlBiski, F. 2012. Evaluation of some therapies and meristem culture to eliminate Potato Y potyvirus from infected potato plants. Plant Biotechnology 29:237-243.

Andrews, S. 2010. FastQC: A Quality Control tool for High Throughput Sequence Data Availabel online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Aronesty, E. 2011. Command-line tools for processing biological sequencing data. ea-utils.

Barker, H. 1994. Incidence of potato virus Y infection in seed and ware tubers of the potato cv. Record. Ann. Appl. Biol. 124:179-183.

Betz-Stablein, B.D., Töpfer, A., Littlejohn, M., Yuen, L., Colledge, D., Sozzi, V., Angus, P., Thompson, A., Revill, P., Beerenwinkel, N., Warner, N., and Luciani, F. 2016. Single-molecule sequencing reveals complex genome variation of hepatitis B virus during 15 years of chronic infection following liver transplantation. Journal of Virology 90:7171-7183.

Carrasco, P., de la Iglesia, F., and Elena, S.F. 2007. Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in *Tobacco Etch Virus*. Journal of Virology 81:12979-12984.

Chung, B.Y.-W., Miller, W.A., Atkins, J.F., and Firth, A.E. 2008. An overlapping essential gene in the Potyviridae. Proc. Natl. Acad. Sci. 105:5897-5902.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. Fly 6:80-92.

Crosslin, J.M., Hamm, P.B., Eastwell, K.C., Thornton, R.E., Brown, C.R., Corsini, D., Shiel, P.J., and Berger, P.H. 2002. First report of the necrotic strain of *Potato virus Y* (PVY^N) on potatoes in the Northwestern United States. Plant Dis. 86:1177-1177.

Cuevas, J.M., Willemsen, A., Hillung, J., Zwart, M.P., and Elena, S.F. 2015. Temporal dynamics of intra-host molecular evolution for a plant RNA virus. Mol. Biol. Evo.

Cuevas, J.M., Delaunay, A., Visser, J.C., Bellstedt, D.U., Jacquot, E., and Elena, S.F. 2012. Phylogeography and molecular evolution of *Potato virus Y*. PLoS ONE 7:e37853.

Domingo, E., Sheldon, J., and Perales, C. 2012. Viral quasispecies evolution. Microbiology and Molecular Biology Reviews 76:159-216.

Eigen, M., and Biebricher, C. 1988. RNA Genetics, Vol. III, Variability of RNA Genomes. Pages 211-245 in: Eds E. Domingo, JJ Holland and P. Ahlquist. CRC Press Inc., Florida.

García-Arenal, F., Fraile, A., and Malpica, J.M. 2001. Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. 39:157-186.

Gray, S., De Boer, S., Lorenzen, J., Karasev, A., Whitworth, J., Nolte, P., Singh, R., Boucher, A., and Xu, H.M. 2010. *Potato virus Y:* An Evolving Concern for Potato Crops in the United States and Canada. Plant Dis. 94:1384-1397.

Green, K.J., Brown, C.J., Gray, S.M., and Karasev, A.V. 2017. Phylogenetic study of recombinant strains of Potato virus Y. Virology 507:40-52.

Gutiérrez, S., Michalakis, Y., and Blanc, S. 2012. Virus population bottlenecks during within-host progression and host-to-host transmission. Current Opinion in Virology 2:546-555.

Hu, X., Karasev, A., Brown, C., and Lorenzen, J. 2009. Sequence characteristics of *Potato virus Y* recombinants. Journal of General Virology 90:3033-3041.

Jakab, G., Droz, E., Brigneti, G., Baulcombe, D., and Malnoe, P. 1997. Infectious *in vivo* and *in vitro* transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. Journal of General Virology 78:3141-3145.

Karasev, A., and Gray, S. 2013. Continuous and Emerging Challenges of *Potato virus Y* in Potato. Annu. Rev. Phytopathol. 51:571-586.

Karasev, A.V., Nikolaeva, O.V., Hu, X., Sielaff, Z., Whitworth, J., Lorenzen, J.H., and Gray, S.M. 2010. Serological Properties of Ordinary and Necrotic Isolates of *Potato virus Y*: A Case Study of PVY^{N} Misidentification. Am. J. Potato Res. 87:1-9.

Karasev, A.V., Hu, X., Brown, C.J., Kerlan, C., Nikolaeva, O.V., Crosslin, J.M., and Gray, S.M. 2011. Genetic Diversity of the Ordinary Strain of Potato virus Y (PVY) and Origin of Recombinant PVY Strains. Phytopathology 101:778-785.

Kehoe, M.A., Coutts, B.A., Buirchell, B.J., and Jones, R.A.C. 2014. Plant virology and next generation sequencing: Experiences with a *Potyvirus*. PLoS ONE 9:e104580.

Kerlan, C., and Moury, B. 2008. *Potato virus Y*. Pages 287-296 in: Encyclopedia of Virology (Third Edition), B.W.J. Mahy and M.H.V.v. Regenmortel, eds. Academic Press, Oxford.

Krzywinski, M.I., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. 2009. Circos: An information aesthetic for comparative genomics. Genome Res.

Kutnjak, Denis, Rupar, M., Gutierrez-Aguirre, I., Curk, T., Kreuze, J.F., and Ravnikar, M. 2015. Deep sequencing of virus derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscape of a plant virus population. Journal of Virology.

Langmead B, S.S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-2079.

Lorenzen, J., Piche, L.M., Gudmestad, N.C., Meacham, T., and Shiel, P. 2006a. A Multiplex PCR Assay to Characterize *Potato virus Y* Isolates and Identify Strain Mixtures. Plant Dis. 90:935-940.

Lorenzen, J., Meacham, T., Berger, P.H., Shiel, P.J., Crosslin, J.M., Hamm, P.B., and Kopp, H. 2006b. Whole genome characterization of *Potato virus Y* isolates collected in the western USA and their comparison to isolates from Europe and Canada. Arch Virol 151:1055-1074.

Mahmoud, S.Y.M., Hosseny, M.H., and Abdel-Ghaffar, M.H. 2009. Evaluation of Some Therapies to Eliminate Potato Y Potyvirus from Potato Plants. International Journal of Virology 5:64-76.

Martínez, F., Lafforgue, G., Morelli, M.J., González-Candelas, F., Chua, N.-H., Daròs, J.-A., and Elena, S.F. 2012. Ultradeep sequencing analysis of population dynamics of virus escape mutants in RNAi-mediated resistant plants. Mol. Biol. Evo. 29:3297-3307.

Mondal, S., and Gray, S.M. 2017. Sequential acquisition of Potato virus Y strains by Myzus persicae favors the transmission of the emerging recombinant strains. Virus Research 241:116-124.

Moreno, I.M., Malpica, J.M., Díaz-Pendón, J.A., Moriones, E., Fraile, A., and García-Arenal, F. 2004. Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. Virology 318:451-460.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Meth 5:621-628.

Nicot, N., Hausman, J.-F., Hoffmann, L., and Evers, D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany 56:2907-2914.

Ogawa, T., Tomitaka, Y., Nakagawa, A., and Ohshima, K. 2008. Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. Virus Research 131:199-212.

Perales, C., Martín, V., Ruiz-Jarabo, C.M., and Domingo, E. 2005. Monitoring sequence space as a test for the target of selection in viruses. J. Mol. Biol. 345:451-459.

Quenouille, J., Vassilakos, N., and Moury, B. 2013. *Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. Molecular Plant Pathology 14:439-452.

Quinlan, A.R., and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

R Development Core Team, R. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-00, URL http://www.r-project.org/.

Sanjuán, R., Moya, A., and Elena, S.F. 2004. The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. 101:8396-8401.

Schubert, J., Fomitcheva, V., and Sztangret-Wiśniewska, J. 2007. Differentiation of *Potato virus Y* strains using improved sets of diagnostic PCR-primers. Journal of Virological Methods 140:66-74.

Simmons, H.E., Dunham, J.P., Stack, J.C., Dickins, B.J., Pagan, I., Holmes, E.C., and Stephenson, A.G. 2012. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. The Journal of general virology 93:1831-1840.

Sims, D., Sudbery, I., Ilott, N.E., Heger, A., and Ponting, C.P. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nature reviews. Genetics 15:121-132.

Smith, K.M. 1931. On the Composite Nature of Certain Potato Virus Diseases of the Mosaic Group as Revealed by the use of Plant Indicators and Selective Methods of Transmission. Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character 109:251-267.

Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings Bioinf. 14:178-192.

Urcuqui-Inchima, S., Haenni, A.-L., and Bernardi, F. 2001. Potyvirus proteins: a wealth of functions. Virus Research 74:157-175.

Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., and Andino, R. 2006. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439:344-348.

Wilm, A., Aw, P.P.K., Bertrand, D., Yeo, G.H.T., Ong, S.H., Wong, C.H., Khor, C.C., Petric, R., Hibberd, M.L., and Nagarajan, N. 2012. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 40:11189-11201.

Wysoker, A., Tibbetts, K., and Fennell, T. 2013. Picard tools version 1.90. http://picard.sourceforge.net/.

Zhong, S., Fei, Z., and Giovannoni, J.J. 2012. Modifications for performing multiplex strand-specific RNA sequencing using Illumina HiSeq. Cold Spring Harb Protoc.

Zhong, S., Joung, J.G., Zheng, Y., Chen, Y.R., Liu, B., Shao, Y., Xiang, J.Z., Fei, Z., and Giovannoni, J.J. 2011. High-throughput illumina strand-specific RNA sequencing library preparation. Cold Spring Harb Protoc:940-949.

Zwart, M.P., Willemsen, A., Daròs, J.-A., and Elena, S.F. 2014. Experimental Evolution of Pseudogenization and Gene Loss in a Plant RNA Virus. Mol. Biol. Evo. 31:121-134.

CHAPTER 2

Mapping Loci that Control Tuber and Foliar Symptoms Caused by PVY in Autotetraploid Potato (*Solanum tuberosum* L.)

Washington da Silva*, Jason Ingram*, Christine A. Hackett[†], Joseph J. Coombs[‡], David Douches[‡], Glenn Bryan[§], Walter De Jong[††,1], and Stewart Gray*[,**,2]

*School of Integrative Plant Science, Plant Pathology and Plant-Microbe Biology Section, Cornell University, Ithaca, NY 14853, U.S.A.

[†]Biomathematics and Statistics Scotland, Invergowrie, Dundee, DD2 5DA, UK

[‡]Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, Michigan 48824

[§]The James Hutton Institute, Invergowrie, Dundee, DD2 5DA, United Kingdom

[††]School of Integrative Plant Science, Plant Breeding and Genetics Section, Cornell University, Ithaca, NY 14853, U.S.A.

**USDA, ARS, Emerging Pests & Pathogens Research Unit, Ithaca, NY 14853, U.S.A.

Corresponding authors: E-mails: [1]Walter De Jong - wsd2@cornell.edu; [2]Stewart Gray - smg3@cornell.edu

TARGET JOURNAL: G3

KEYWORDS: genetic linkage map, QTL, autotetraploid potato, single-nucleotide polymorphism, *Potato Virus Y*, PTNRD

ABSTRACT

Potato tuber necrotic ringspot disease (PTNRD) is a tuber deformity associated with infection by the tuber necrotic strain of *Potato virus Y* (PVY^{NTN}). PTNRD negatively impacts tuber quality and marketability and poses a serious threat to seed and commercial potato production worldwide. PVY^{NTN} symptoms differ in the cultivars Waneta and Pike: Waneta expresses severe PTNRD and foliar mosaic with vein and leaf necrosis, whereas Pike does not express PTNRD and mosaic is the only foliar symptom. To map loci that influence tuber and foliar symptoms, 236 F_1 progeny of a cross between Waneta and Pike were inoculated with PVY^{NTN} isolate NY090029 and genotyped using 12,808 Potato SNPs. Foliar symptom type and severity were monitored for 10 weeks, while tubers were evaluated for PTNRD expression at harvest and again after 60 days in storage. Pairwise correlation analyses indicate a strong association between PTNRD and vein necrosis ($\tau = 0.4195$). QTL analyses revealed major-effect QTLs on chromosomes 4 and 5 for mosaic, 4 for PTNRD, and 5 for foliar-necrosis symptoms. Locating QTLs associated with PVY-related symptoms provides a foundation for breeders to develop markers that can be used to screen out potato clones with undesirable phenotypes, e.g., those likely to develop PTNRD or to be symptomless carriers of PVY.

INTRODUCTION

Potato virus Y (PVY) is the most common and most serious virus affecting US potato production, and resistant potato cultivars represent the most effective control option (Karasev and Gray 2013a; Fulladolsa et al. 2015). PVY exists as a myriad of strains, including: the ordinary strain PVY^{O}, the tobacco vein necrosis strain PVY^{N}, the stipple streak strain PVY^{C}, and the tuber necrosis strain PVY^{NTN} that elicits potato tuber necrotic ringspot disease (PTNRD) (Karasev and Gray 2013a; Schubert et al. 2007). Until recently, North American potato breeding programs have not prioritized PVY resistance during selection. A lack of resistance and the popularity of several widely-planted varieties that are symptomless carriers of PVY have facilitated an increase in PVY incidence and contributed to the emergence of new PVY strains that cause PTNRD (Gray et al. 2010; Karasev and Gray 2013b). PTNRD poses a serious threat to the seed and commercial production industries by contributing to the rejection of seed lots for exceeding virus tolerance, as well as negatively impacting tuber quality (Karasev and Gray 2013a; Kerlan and Moury 2008). Some potato cultivars widely grown in the US and Canada are highly susceptible to PTNRD, such as Yukon Gold, Yukon Gem, Red Norland, Highland Russet, Alturas, Blazer, and Ranger Russet (McDonald and Singh 1996; Singh et al. 1998).

Resistance genes effective against PVY have been identified in cultivated and wild potato species (Cockerham 1970; Jones 1990; Fulladolsa et al. 2015; Karasev and Gray 2013a) and have been classified into two types, hypersensitive resistance (HR) and extreme resistance (ER) (Gebhardt and Valkonen 2001). HR is associated with the development of visible necrotic lesions at the point of infection. In some varieties the response can be a systemic necrosis mani-

fested as vein necrosis, leaf necrosis or leaf drop. All of these responses can contribute to limiting virus replication and systemic spread, as well as reducing aphid transmission efficiency of the virus from these plants. HR is conferred by *N* genes (Solomon-Blackburn and Barker 2001). The major *N* genes, Ny_{tbr} and Nc_{spl} (Celebi-Toprak et al. 2002; Moury et al. 2011), *Ny-1* (Szajko et al. 2008), and *Ny-2* (Szajko et al. 2014) have been mapped to chromosomes 4, 9, and 11, respectively. ER is asymptomatic, results in no detectable virus multiplication in inoculated plants, and is conferred by *R* genes (Solomon-Blackburn and Barker 2001). Several molecular markers have been developed for potato *R* genes, including: RYSC3 for detection of Ry_{adg} from *S. tuberosum* ssp. *andigena*, on chromosome 11 (Sorri et al. 1999; Kasai et al. 2000); 38–530 and CT220 for Ry_{chc} from *S. chacoense*, on chromosome 9 (Hosaka et al. 2001; Sato et al. 2006); and GP122, STM003, and YES3-3B for Ry_{sto} from *S. stoloniferum*, on chromosome 12 (Song et al. 2005; Song and Schwarzfischer 2008; Valkonen et al. 2008). Many of those markers have been successfully incorporated in breeding programs to develop PVY–resistant cultivars (Fulladolsa et al. 2015; Ottoman et al. 2009; Watanabe 2015).

Marker-assisted selection (MAS) has proven to be a fast and efficient tool to select cultivars with desirable traits in plant breeding (Xu and Crouch 2008). Developing markers linked to important genes in cultivated potato (*Solanum tuberosum* ssp. *tuberosum*) is more challenging than in many other crops, primarily because conducting linkage analyses is more difficult in autotetraploids than in diploids. Nevertheless, with the sequencing of the potato genome (PGSC 2011), followed by the development, validation, and release of the Infinium Potato SNP Arrays (Hamilton et al. 2011; Felcher et al. 2012), improvements of statistical models for analyzing SNP dosage in tetraploids (Hackett et al. 2013; Hackett et al. 2014; Preedy and

Hackett 2016; Hackett et al. 2001), and the development of TetraploidSNPMap – user-friendly software specifically designed to analyze SNP markers in polyploid germplasm (Hackett et al. 2017) – QTL analyses in potato have recently become much more feasible.

Developing varieties that do not express PTNRD upon infection is potentially a useful complement or alternative to developing varieties resistant to PVY. Genetic markers that breeders could use to select for lack of PTNRD expression would facilitate the development of such varieties. The goal of this research was to map genes that mediate PTNRD and other types of foliar symptoms induced by PVY infection (mosaic, vein necrosis, and leaf necrosis).

MATERIALS AND METHODS

Plant Material

The H25 mapping population comprises 236 F_1 progeny of a cross between the cultivars Waneta (as female) and Pike (as male). These two cultivars express different symptoms when infected by PVY isolate NY090029 (a PVY^{NTN} strain). Waneta expresses severe PTNRD and foliar mosaic with vein and leaf necrosis. Pike does not exhibit PTNRD and mild mosaic is the only foliar symptom.

True potato seeds of H25 were germinated on a bed of Cornell potting mix (Boodley and Sheldrake 1982). After one month, 80 seedlings were individually transplanted to 15-cm clay pots. Each seedling was vegetatively propagated via cuttings to increase the number of plants per genotype. One tuber of each parent was individually planted in a 15-cm clay pot and cuttings were taken from the sprouts, also to increase the number of plants per genotype. All cuttings, from parents and progeny, were dipped in Hormex rooting hormone #1 (Brooker Chem.

Corp., Chatsworth, CA) and planted individually into 96 well trays containing Cornell soil mix for rooting and grown for one month. Additionally, 200 true potato seeds from the H25 population were sterilized and placed into tissue culture media by the following method. Seeds were soaked overnight in a 1500ppm gibberellic acid solution, then the solution was removed and a 10% bleach solution was added and incubated for 10 minutes with periodic inverting of the tube. The bleach solution was removed and sterile H_2O was added to wash the seeds, repeating the washing step four times. Seeds were plated onto a sterile autoclaved size 1 Whatman circle filter paper in a petri dish damped with H_2O. Petri dishes were sealed with parafilm and placed under growth lights (16hr light/day) until the seeds sprouted. The young sprouts were then transferred to Murashige and Skoog medium. One-hundred and fifty-six progeny were established in tissue culture and these plants, as well as the progeny sown directly into soil, were used for virus phenotyping and SNP genotyping ("the mapping population"). Six well-rooted plants from each clone of the mapping population and the parents were transplanted individually in four-liter plastic pots containing Cornell soil mix. During all steps of the experiment, including germination and sprouting, plants were maintained in an insect-free greenhouse under 16h days at 25±3^{o}C.

Phenotypic data

Two weeks after transplanting, five plants from each clone and the two parents were inoculated with PVY^{NTN} isolate NY090029. One plant from each clone and each parent was left uninoculated as a negative control. To prepare viral inoculum, the PVY^{NTN} isolate NY090029 (maintained in lyophilized tobacco tissue at -80^{o}C) was mechanically inoculated to individual tobacco (*Nicotiana tabacum*) plants at the three- to five-leaf stage. Lyophilized tissue (100 mg)

was homogenized in 500 µl of phosphate-buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, and pH adjusted to 7.4 with HCl) and rubbed onto carborundum-dusted (325 mesh) tobacco leaves. Inoculated plants were kept in a greenhouse for one month. Infected tobacco leaves were harvested, ground in PBS in a volume of 1:5 (1g of leaf to 5ml of PBS), and filtered through cheesecloth to produce the inoculum. Then, potato plants were inoculated by using a cotton swab to lightly rub the PVY inoculum using carborundum as an abrasive. The same plants were inoculated twice more, with one-week intervals between inoculations. Plant infection status was checked by ELISA using a 4C3 commercial kit (Agdia Inc., Elkart IN, USA), following the manufacturer's directions.

One and three weeks after the final round of inoculations, foliar symptoms were evaluated on each plant (Figure 1). Mosaic was scored on this scale: 0 = no symptoms, 1 = mild mosaic (mosaic pattern muted but present), 2 = typical mosaic (mosaic pattern evident and some leaf rugosity *"rough or wrinkled surface"* possible), and 3 = severe mosaic (mosaic pattern evident, plant stunting, rugosity and deformation on leaves). The grading for leaf and vein necrosis symptoms was binary: 0 = no symptoms, 1 = symptoms present. Four months after the potato cuttings were transplanted, the vines were removed, the pots were left to dry out for three weeks, and then tubers were harvested. At harvest and three months later, PTNRD severity was visually evaluated for each tuber as follows: 0= no PTNRD, 1= 1-10% PTNRD, 2= 11-25% PTNRD, 3= 26-50% PTNRD, 4= 51-75% PTNRD, and 5= 76-100% PTNRD (Figure 2). For subsequent analyses the highest disease value – the most severe symptoms observed among the five plants tested for each genotype – was used. In a pilot study, using the linkage maps from the

full population (236 clones), we ran QTL analyses on a subset of the population (85 clones) using the mean and the highest disease values, and found the same significant QTLs for both types of data. We elected to use the highest disease values in all subsequent analyses. Pairwise correlation analyses were performed on the phenotypic dataset with the non-parametric Kendall's tau rank correlation coefficient to measure the strength of the relationship between each type of symptom. All statistical analyses and plotting for data visualization were performed in R (R Core Team 2016) using the R packages Hmisc (version 4.0-0) (Harrell Jr. 2016) and corrplot (version 0.77) (Wei and Simko 2016).

SNP genotyping

DNA from 236 progeny clones and their parents was extracted from frozen leaf tissue using a QIAGEN DNeasy Plant Mini Kit (Qiagen, Valencia, CA-USA), following the manufacturer's directions. DNA was quantified with the Quant-it PicoGreen assay (Invitrogen, San Diego, CA-USA) and adjusted to a concentration of 50ng μL^{-1}. The population was genotyped with the Illumina Infinium V2 Potato SNP Array (12,808 SNPs: original SolCAP Infinium 8303 Potato SNP Array with 4,500 additional SNPs to increase coverage in candidate genes and R-gene hotspots) (Hamilton et al. 2011). Illumina GenomeStudio software (Illumina, Inc., San Diego, CA-USA) was used for initial sample quality assessment and generating marker theta values (which give dosage allelic information for parents and offspring). In an autotetraploid mapping population, five allele dosages (AAAA, AAAB, AABB, ABBB, and BBBB) are possible and are expected to consist of theta scores in five clusters, centering around 0.0, 0.25, 0.50, 0.75, and 1.0, respectively. Tetraploid (5-cluster) genotyping was based on theta value thresholds, using a custom script from

the SolCAP project (Hirsch et al. 2013). Using this script, 5-cluster calling and filtering were performed to remove low quality markers and markers with multiple hits to the potato genome sequence of *Solanum tuberosum* group Phureja DMI-3 516 R44 (Sharma et al. 2013). SNPs with >20% missing genotype calls in the population were excluded from the dataset.

Linkage map construction and QTL analysis

Construction of linkage maps and QTL analysis of each chromosome were performed as described in (Hackett et al. 2014; Hackett et al. 2013; Preedy and Hackett 2016; Hackett et al. 2017). All linkage and QTL analyses involving testing for distorted segregation, clustering analysis, calculation of recombination fractions and LOD (logarithm of the odds) scores, ordering of SNPs, and inference of parental phase, were performed in TetraploidSNPMap. Markers with significance of the χ^2 goodness-of-fit statistic less than 0.001 for simplex SNPs and 0.01 for duplex or greater dosage SNPs were flagged as distorted. To detect and remove problematic markers and for ordering of SNPs, the following analyses were performed: hierarchical clustering analyses using average linkage clustering of SNPs with expected ratios, 2-point analyses to calculate the recombination frequency and LOD score for the SNPs pairs in each possible phase, and multidimensional scaling analysis (MDS) to calculate the best order for the SNPs in the linkage group (Preedy and Hackett 2016; Hackett et al. 2017). Finally, the phases of the ordered SNPs were inferred as far as possible by the automated phase analysis in TetraploidSNPMap and completed manually prior to carrying out QTL analysis.

QTL analysis was run for each linkage group separately using three input files: the linkage map, the SNP data for the linkage group, and the phenotypic trait dataset. For each trait, interval mapping displayed the LOD profile on the chromosome, giving the LOD score statistics,

percentage variation explained, and QTL effect for each homologous chromosome. Ninety percent and 95% LOD thresholds were obtained to establish the statistical significance of each QTL position using permutation tests (Churchill and Doerge 1994) with 300 permutations (Hackett et al. 2014). Simple models for the genotype means estimated at the most probable QTL position were calculated using the Schwarz Information Criterion (SIC) (Schwarz 1978), models with the lowest value for SIC are considered the best models (Hackett et al. 2014). Linkage maps and QTL positions were generated in MapChart 2.30 (Voorrips 2002).

Concordance between the linkage maps generated in this study and the potato reference genome (PGSC Version 4.03 Pseudomolecules) was evaluated in MareyMap R package version 1.3.1 (Rezvoy et al. 2007). Plots of the genetic position (cM) with the physical position (Mb) of each SNP marker in each chromosome were generated using the graphical interface MareyMapGUI, the interpolation method "cubic splines" was used to calculate the curve slope.

Data availability

All the raw data from this study was compiled in .txt tables and are available in the supplemental files: Table S1 and Table S2. Complementary information for the Results and Discussion section are provided in Support Information: Figures S1, S2, and S3, and Tables S3 and S4.

RESULTS AND DISCUSSION

Genotyping and preliminary SNP marker processing

The 12808 SNPs from the new Illumina Potato V2 SNP Array (12K) were used to genotype the parents and 236 offspring in this study. After a pre-filtering step to remove SNPs with

missing theta values, low quality, and those with multiple hits to the potato reference genome PGSC Version 4.03 Pseudomolecules, 4,859 SNPs were selected for downstream analyses. Of these, 1063 SNPs had missing data in >20% of the population, and were also excluded from the dataset. The remaining 3796 SNPs were loaded into TetraploidSNPMap and 1258 distorted SNPs with chi-square statistics having a significance less than 0.001 were removed. Hierarchical clustering analyses easily grouped the remaining 2538 markers into 12 linkage groups (Supporting Information, Table S1). A total of 95 SNPs was flagged as duplicated and 17 were excluded as outliers after clustering, 2-point, and MDS analyses.

Approximately, 65% (1583) of the markers followed the parental genotype configurations of simplex (AAAA X AAAB, AAAB X AAAA), duplex (AAAA X AABB, AABB X AAAA), and double-simplex (AAAB X AAAB, ABBB X ABBB), while ~ 35% (843) were between simplex-duplex (AAAB X AABB) and double-duplex (AABB X AABB) configurations (Table S3). The large number and diversity of configurations of SNPs in our dataset allowed for the construction of high-density linkage maps, which significantly increased the chances for the detection of significant QTLs for the traits studied (Massa et al. 2015; Hackett et al. 2013; Li et al. 2014; Hackett et al. 2014).

Linkage map construction and QTL analysis

The 2426 SNPs were mapped to the 12 potato chromosomes with chromosomes 1 and 12 having the highest and the lowest number of mapped SNPs (281 and 138), respectively (Table 1). Overall, 1809 SNPs segregated in “Waneta”, 1962 segregated in “Pike”, and 1345 SNPs segregated in both parents. The total genetic distance for each of the parental maps was 1052.6 cM (for Waneta) and 1097.1 cM (for Pike), with the map lengths of individual chromo-

somes ranging from 72.1 to 120.6 cM. There was an average of 157 SNP markers per chromosome and a marker density of ~1.75 SNPs per cM. The genetic maps of both parents, covered, on average, 98% of the PGSC v4.03 Pseudomolecules (Table 1, and Figure 3).

Vein Necrosis positively correlated with PTNRD

Non-parametric Kendall's tau rank correlation analyses indicated a weak correlation among mosaic and other symptom types (PTNRD, foliar necrosis, and vein necrosis). In contrast, vein necrosis exhibited the highest correlation with other symptom types especially PTNRD (Figure 4) – an indication that when vein necrosis is observed, there is a high chance of PTNRD development in tubers. The evaluation of PTNRD requires a lot of time as tubers need to be stored for at least two months after harvest for full expression of the symptoms. Knowing that vein necrosis is correlated with PTNRD may benefit potato growers and researchers alike.

Significant QTLs were identified on chromosomes 4 and 5 for mosaic and leaf necrosis

Mosaic symptoms were frequent in the population, with 219 of the 236 offspring expressing symptoms (Figure 5). This was not surprising, as we had found in preliminary studies that PVY isolate NY090029 is highly virulent and elicited severe mosaic in most inoculated plants including both parents. In contrast, only 31 and 172 clones developed leaf necrosis and vein necrosis, respectively (Figure 5). QTL analyses revealed significant QTLs on chromosomes 4 and 5 for mosaic and leaf necrosis (Table 2 and Figures 6, 7, S1, S2). No significant QTLs were detected for vein necrosis in the population.

On chromosome 4, the QTLs had maximum LOD scores of 5.20 and 4.44 explaining 7.2% and 5.6% of the trait variances for mosaic and leaf necrosis, respectively. These LOD scores

were above the upper 95% LOD permutation thresholds of 3.95 and 3.81 and the QTL peaks were located at positions 51 cM and 46 cM for mosaic and leaf necrosis, respectively. Analyses of different simple genetic models were performed with TetraploidSNPMap to determine the best simple fitting model for each trait. For mosaic, the best model was a simplex allele (AAAB) on homologous chromosome 4 (H4, Fig. 6) of Waneta, with the B allele associated with a decrease in symptom expression. This model had the lowest SIC, -73.94, in comparison with the full model (SIC = -55.94). For leaf necrosis, the best model was a simplex allele (BAAA) on homologous chromosome 5 of Pike (H5, Fig. 7), with the B allele associated with a decrease in symptom expression. This model had SIC = -151.98, while the SIC for the full model was -145.80.

On chromosome 5, the maximum LOD scores were 7.34 and 5.20 and those QTLs explained 10.9% and 6.6% of the phenotypic variance for mosaic and leaf necrosis, respectively (Table 2). The LOD peaks were located at positions 43 cM and 31 cM and their scores were above the upper 95% LOD permutation thresholds of 3.65 and 3.76 for mosaic and leaf necrosis, respectively. The simpler models analyses estimated a double-simplex and a duplex genotype for mosaic and leaf necrosis, respectively. For mosaic, the best model was an ABAA X AABA configuration on homologous chromosomes 2 and 7 (H2+H7, Figures S1 and S2) with the B allele associated with a decrease in symptom expression, and both parents contributing the B allele to their offspring. This model had the lowest SIC, -70.68, in comparison with the full model (SIC = -66.20). For leaf necrosis, the best model was an ABBA configuration on homolo-

gous chromosomes 2 and 3 of Waneta (H2+H3, Figure S1) with the B allele associated with a decrease in symptom expression. The SIC for this model was -141.62, the full model had SIC = -136.41.

Analyses of the concordance between the linkage maps and the potato reference genome (PGSC Version 4.03 Pseudomolecules) for chromosomes 4 and 5 generated graphs that were consistent with published chromosome structures (Figure S3) (Massa et al. 2015; Felcher et al. 2012; Sharma et al. 2013).

A major-effect QTL for PTNRD expression was detected on chromosome 4

One hundred and forty-five clones produced tubers that expressed some degree of PTNRD. Of the 89 remaining clones, 11 clones did not produce tubers and 78 produced tubers with no PTNRD. A PTNRD QTL was detected on chromosome 4 that had a LOD score of 5.82, explained 8.6% of the trait variance (Table 2), and was above the 95% LOD permutation upper threshold of 3.92. The QTL peak was located at 46 cM and analyses of different genetic models indicated that an allele from Pike explains the trait variance. The QTL is linked to a simplex SNP (GGGG x TGGG), with the T allele associated with a decrease in disease on homologous chromosome 5 (H5, Figure 7). This model had the lowest SIC of – 28.32 compared to the full model with SIC = - 23.61. The closest SNP with this configuration is the SNP solcap_snp_c2_39848 at genetic position 47.09 cM and physical position 35.68 Mb. This QTL was located in the central region of chromosome 4, the same region where QTLs for mosaic and leaf necrosis were detected. The center of chromosome 4 harbors two known genes, Ny_{tbr} and Nc_{spl}, that cause HR in potatoes when infected with PVY^{O} and PVY^{C}, respectively (Celebi-Toprak et al. 2002; Moury

et al. 2011). It is possible that alleles of these genes influence PTNRD, mosaic, and/or leaf necrosis symptoms. R genes frequently occur in tightly linked clusters (Michelmore and Meyers 1998) and the distribution of such genes and QTLs is not random in the potato genome (Gebhardt and Valkonen 2001). Even though PTNRD was strongly correlated with vein necrosis in the dataset, we failed to identify any statistical significant QTL for vein necrosis in our analyses. We did detect a peak for vein necrosis at position 38 cM on chromosome 4; however, this peak was below the 90% LOD permutation upper threshold of 3.5. As the trait variance explained by this peak is below 5%, a larger population size (> 400) may be needed to detect this QTL (Hackett et al. 2014).

Conclusion

The detection of major QTLs for different PVY symptom types in close proximity to each other on chromosome 4 suggests that markers diagnostic for specific haplotypes of this region may prove useful for breeders who want to select genes that confer resistance to infection and/or multiple PVY-related symptoms. Finally, it is important to point out that QTL analysis is approximate, as the disease traits evaluated in this study are ordinal or binary scores and so definitely not normal. However, basing significance on permutation of this data helps, in part, to address this problem.

ACKNOWLEDGEMENTS

The authors would like to thank Kelly Zarka and Natalie Kirkwyland for helping with tissue culture and Daniel Zarka for helping genotyping the H25 population. This study was

funded in part by USDA-SCRI (2009-51181-05894 and 2014-51181-22373), the UK Biotechnology and Biological Sciences Research Council grant # BB/L011840/1 as part of the joint USDA-NSF-NIH-BBSRC Ecology and Evolution of Infectious Diseases program. The work of Glenn Bryan, Christine Hackett, and the TetraploidSNPMap software development were funded by the Rural & Environment Science & Analytical Services Division of the Scottish Government. Washington da Silva was partially supported by a SUNY Diversity Fellowship and by the 2014/2015 National Potato Council (NPC) scholarship.

LITERATURE CITED

Boodley, J.W., and R. Sheldrake, 1982 Cornell peat-like mixes for commercial plant growing. *Cornell Inf. Bull. 43.*

Celebi-Toprak, F., S.A. Slack, and M.M. Jahn, 2002 A new gene, Ny_{tbr} , for hypersensitivity to *Potato virus Y* from Solanum tuberosum Maps to Chromosome IV. *Theoretical and Applied Genetics* 104 (4):669-674.

Churchill, G.A., and R.W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* 138 (3):963.

Cockerham, G., 1970 Genetical studies on resistance to potato viruses X and Y. *Heredity* 25 (3):309-348.

Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton *et al.*, 2012 Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. *PLoS ONE* 7 (4):e36347.

Fulladolsa, A.C., F.M. Navarro, R. Kota, K. Severson, J.P. Palta *et al.*, 2015 Application of Marker Assisted Selection for *Potato virus Y* Resistance in the University of Wisconsin Potato Breeding Program. *American Journal of Potato Research* 92 (3):444-450.

Gebhardt, C., and J.P.T. Valkonen, 2001 Organization of genes controlling disease resistance in the potato genome. *Annual Review of Phytopathology* 39 (1):79-102.

Gray, S., S. De Boer, J. Lorenzen, A. Karasev, J. Whitworth *et al.*, 2010 *Potato virus Y:* An Evolving Concern for Potato Crops in the United States and Canada. *Plant Disease* 94 (12):1384-1397.

Hackett, C.A., B. Boskamp, A. Vogogias, K.F. Preedy, and I. Milne, 2017 TetraploidSNPMap: Software for Linkage Analysis and QTL Mapping in Autotetraploid Populations Using SNP Dosage Data. *J Hered* 00 (00):1-5.

Hackett, C.A., J.E. Bradshaw, and G.J. Bryan, 2014 QTL mapping in autotetraploids using SNP dosage information. *Theoretical and Applied Genetics* 127 (9):1885-1904.

Hackett, C.A., J.E. Bradshaw, and J.W. McNicol, 2001 Interval Mapping of Quantitative Trait Loci in Autotetraploid Species. *Genetics* 159 (4):1819-1832.

Hackett, C.A., K. McLean, and G.J. Bryan, 2013 Linkage Analysis and QTL Mapping Using SNP Dosage Data in a Tetraploid Potato Mapping Population. *PLoS ONE* 8 (5):e63939.

Hamilton, J., C. Hansey, B. Whitty, K. Stoffel, A. Massa *et al.*, 2011 Single nucleotide polymorphism discovery in elite north american potato germplasm. *BMC Genomics* 12 (1):302.

Harrell Jr., F.E., 2016 Hmisc: Harrell Miscellaneous. R package version 4.0-0. https://cran.r-project.org/package=Hmisc.

Hirsch, C.N., C.D. Hirsch, K. Felcher, J. Coombs, D. Zarka *et al.*, 2013 Retrospective View of North American Potato (*Solanum tuberosum* L.) Breeding in the 20th and 21st Centuries. *G3: Genes|Genomes|Genetics* 3 (6):1003-1013.

Hosaka, K., Y. Hosaka, M. Mori, T. Maida, and H. Matsunaga, 2001 Detection of a simplex RAPD marker linked to resistance to *Potato virus Y* in a tetraploid potato. *American Journal of Potato Research* 78 (3):191-196.

Jones, R.A.C., 1990 Strain group specific and virus specific hypersensitive reactions to infection with potyviruses in potato cultivars. *Annals of Applied Biology* 117 (1):93-105.

Karasev, A., and S. Gray, 2013a Continuous and Emerging Challenges of *Potato virus Y* in Potato. *Annual Review of Phytopathology* 51:571-586.

Karasev, A.V., and S.M. Gray, 2013b Genetic Diversity of *Potato virus Y* Complex. *American Journal of Potato Research* 90 (1):7-13.

Kasai, K., Y. Morikawa, V.A. Sorri, J.P. Valkonen, C. Gebhardt *et al.*, 2000 Development of SCAR markers to the PVY resistance gene Ry_{adg} based on a common feature of plant disease resistance genes. *Genome* 43 (1):1-8.

Kerlan, C., and B. Moury, 2008 *Potato virus Y*, pp. 287-296 in *Encyclopedia of Virology (Third Edition)*, edited by B.W.J. Mahy and M.H.V.v. Regenmortel. Academic Press, Oxford.

Li, X., Y. Wei, A. Acharya, Q. Jiang, J. Kang *et al.*, 2014 A Saturated Genetic Linkage Map of Autotetraploid Alfalfa (*Medicago sativa* L.) Developed Using Genotyping-by-Sequencing Is Highly Syntenous with the Medicago truncatula Genome. *G3: Genes|Genomes|Genetics* 4 (10):1971-1979.

Massa, A.N., N.C. Manrique-Carpintero, J.J. Coombs, D.G. Zarka, A.E. Boone *et al.*, 2015 Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (*Solanum tuberosum* L.). *G3: Genes|Genomes|Genetics* 5 (11):2357-2364.

McDonald, J.G., and R.P. Singh, 1996 Response of potato cultivars to North American isolates of PVY^{NTN}. *American Potato Journal* 73 (7):317-323.

Michelmore, R.W., and B.C. Meyers, 1998 Clusters of Resistance Genes in Plants Evolve by Divergent Selection and a Birth-and-Death Process. *Genome Research* 8 (11):1113-1130.

Moury, B., B. Caromel, E. Johansen, V. Simon, L. Chauvin *et al.*, 2011 The Helper Component Proteinase Cistron of *Potato virus Y* Induces Hypersensitivity and Resistance in Potato Genotypes Carrying Dominant Resistance Genes on Chromosome IV. *Molecular Plant-Microbe Interactions* 24 (7):787-797.

Ottoman, R.J., D.C. Hane, C.R. Brown, S. Yilma, S.R. James *et al.*, 2009 Validation and Implementation of Marker-Assisted Selection (MAS) for PVY Resistance (Ry_{adg} gene) in a Tetraploid Potato Breeding Program. *American Journal of Potato Research* 86 (4):304-314.

PGSC, 2011 The Potato Genome Sequencing Consortium. Genome sequence and analysis of the tuber crop potato. *Nature* 475 (7355):189-195.

Preedy, K.F., and C.A. Hackett, 2016 A rapid marker ordering approach for high-density genetic linkage maps in experimental autotetraploid populations using multidimensional scaling. *Theoretical and Applied Genetics* 129 (11):2117-2132.

R Core Team, 2016 A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rezvoy, C., D. Charif, L. Guéguen, and G.A.B. Marais, 2007 MareyMap: an R-based tool with graphical interface for estimating recombination rates. *Bioinformatics* 23 (16):2188-2189.

Sato, M., K. Nishikawa, K. Komura, and K. Hosaka, 2006 *Potato virus Y* Resistance Gene, Ry_{chc}, Mapped to the Distal End of Potato Chromosome 9. *Euphytica* 149 (3):367-372.

Schubert, J., V. Fomitcheva, and J. Sztangret-Wiśniewska, 2007 Differentiation of *Potato virus Y* strains using improved sets of diagnostic PCR-primers. *Journal of Virological Methods* 140 (1–2):66-74.

Schwarz, G., 1978 Estimating the Dimension of a Model. *The Annals of Statistics* 6 (2):461-464.

Sharma, S.K., D. Bolser, J. de Boer, M. Sønderkær, W. Amoros *et al.*, 2013 Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. *G3: Genes|Genomes|Genetics* 3 (11):2031-2047.

Singh, R.P., M. Singh, and J.G. McDonald, 1998 Screening by a 3-primer PCR of North American PVY^{N} isolates for European-type members of the tuber necrosis-inducing PVY^{NTN} subgroup. *Canadian Journal of Plant Pathology* 20 (3):227-233.

Solomon-Blackburn, R.M., and H. Barker, 2001 A review of host major-gene resistance to potato viruses X, Y, A and V in potato: genes, genetics and mapped locations. *Heredity* 86 (1):8-16.

Song, Y.-S., L. Hepting, G. Schweizer, L. Hartl, G. Wenzel *et al.*, 2005 Mapping of extreme resistance to PVY (Ry_{sto}) on chromosome XII using anther-culture-derived primary dihaploid potato lines. *Theoretical and Applied Genetics* 111 (5):879-887.

Song, Y.-S., and A. Schwarzfischer, 2008 Development of STS Markers for Selection of Extreme Resistance (Ry_{sto}) to PVY and Maternal Pedigree Analysis of Extremely Resistant Cultivars. *American Journal of Potato Research* 85 (2):159-170.

Sorri, V.A., K.N. Watanabe, and J.P.T. Valkonen, 1999 Predicted kinase-3a motif of a resistance gene analogue as a unique marker for virus resistance. *Theoretical and Applied Genetics* 99 (1):164-170.

Szajko, K., M. Chrzanowska, K. Witek, D. Strzelczyk-Żyta, H. Zagórska *et al.*, 2008 The novel gene *Ny-1* on potato chromosome IX confers hypersensitive resistance to *Potato virus Y* and is an alternative to *Ry* genes in potato breeding for PVY resistance. *Theoretical and Applied Genetics* 116 (2):297-303.

Szajko, K., D. Strzelczyk-Żyta, and W. Marczewski, 2014 *Ny-1* and *Ny-2* genes conferring hypersensitive response to *Potato virus Y* (PVY) in cultivated potatoes: mapping and marker-assisted selection validation for PVY resistance in potato breeding. *Molecular Breeding* 34 (1):267-271.

Valkonen, J., K. Wiegmann, J. Hämäläinen, W. Marczewski, and K. Watanabe, 2008 Evidence for utility of the same PCR-based markers for selection of extreme resistance to *Potato virus Y* controlled by Ry_{sto} of Solanum stoloniferum derived from different sources. *Annals of Applied Biology* 152 (1):121-130.

Voorrips, R.E., 2002 MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity* 93 (1):77-78.

Watanabe, K., 2015 Potato genetics, genomics, and applications. *Breeding Science* 65 (1):53-68.

Wei, T., and V. Simko, 2016 Corrplot: Visualization of a Correlation Matrix. R package version 0.77. https://cran.r-project.org/package=corrplot.

Xu, Y., and J.H. Crouch, 2008 Marker-Assisted Selection in Plant Breeding: From Publications to Practice. *Crop Science* 48 (2):391-407.

Tables:

Table 1 – Summary of the parental linkage maps, Waneta (Wan) and Pike.

	No. Mapped SNPs			Map Length, (cM)		Map Length, (Mb[a])		PGSC v4.03 PM, (Mb[a])	Map Coverage[a]		Average Interloci Distance, (cM)	
Chr	Total	Wan	Pike	Wan	Pike	Wan	Pike	DM	Wan	Pike	Wan	Pike
1	281	208	255	114.1	120.6	88.1	88.3	88.7	0.99	1.00	0.55	0.47
2	200	157	167	79.6	85.6	42.8	42.5	48.6	0.88	0.87	0.51	0.51
3	223	195	112	88.5	80.6	61.5	58.2	62.3	0.99	0.93	0.45	0.72
4	227	168	205	91.2	97.3	71.8	71.8	72.2	0.99	0.99	0.54	0.47
5	168	129	135	72.1	84.3	50.5	51.7	52.1	0.97	0.99	0.56	0.62
6	261	188	215	72.9	74.7	59.2	58.6	59.5	0.99	0.99	0.39	0.35
7	247	199	202	78.8	80.8	55.5	55.5	56.8	0.98	0.98	0.40	0.40
8	178	130	127	86.4	86.8	56.5	56.5	56.9	0.99	0.99	0.66	0.68
9	195	157	153	107.1	107.1	60.4	61.3	61.5	0.98	1.00	0.68	0.70
10	141	92	125	99.2	99.2	59.5	59.3	59.8	1.00	0.99	1.08	0.79
11	167	93	153	77.8	74.2	43.9	44.6	45.5	0.96	0.98	0.84	0.48
12	138	93	113	84.9	105.9	59.8	61.0	61.2	0.98	1.00	0.91	0.94
Total	**2426**	**1809**	**1962**	**1052.6**	**1097.1**	**709.5**	**709.3**	**725.1**	**0.98**	**0.98**	**0.58**	**0.56**

[a]Map length (Mb) and map coverage values are based on the PGSC Version 4.03 Pseudomolecules of the potato reference genome *Solanum tuberosum* group Phureja DM1-3 516 R44 (DM).

Table 2 – QTL information for the traits analyzed in the H25 population.

Trait	Chr	QTL position (cM)	LOD	Variance Explained (%)	SIC	Homologous Chr
Mosaic	4	51	5.2	7.2	-73.9	h4
Mosaic	5	43	7.3	10.9	-70.7	h2+h7
Vein Necrosis	-	-	-	-	-	-
Leaf Necrosis	4	46	4.4	5.6	-152.0	h5
Leaf Necrosis	5	31	5.2	6.6	-141.6	h2+h3
PTNRD	4	46	5.8	8.4	-28.3	h5

QTL – quantitative trait loci, Chr – chromosome, LOD – logarithm of the odds, CI – confidence interval, SIC - Schwarz Information Criterion.

Figures:

Figure 1 – Foliar symptom severity ratings for 236 clones from the H25 population. Ratings were scored on a 0 to 3 scale with 0 = no disease and 3 = most severe symptoms for mosaic and a 0 to 1 scale for leaf and vein necrosis with 0 = no disease and 1 = disease. Black arrows highlight each PYV symptoms.

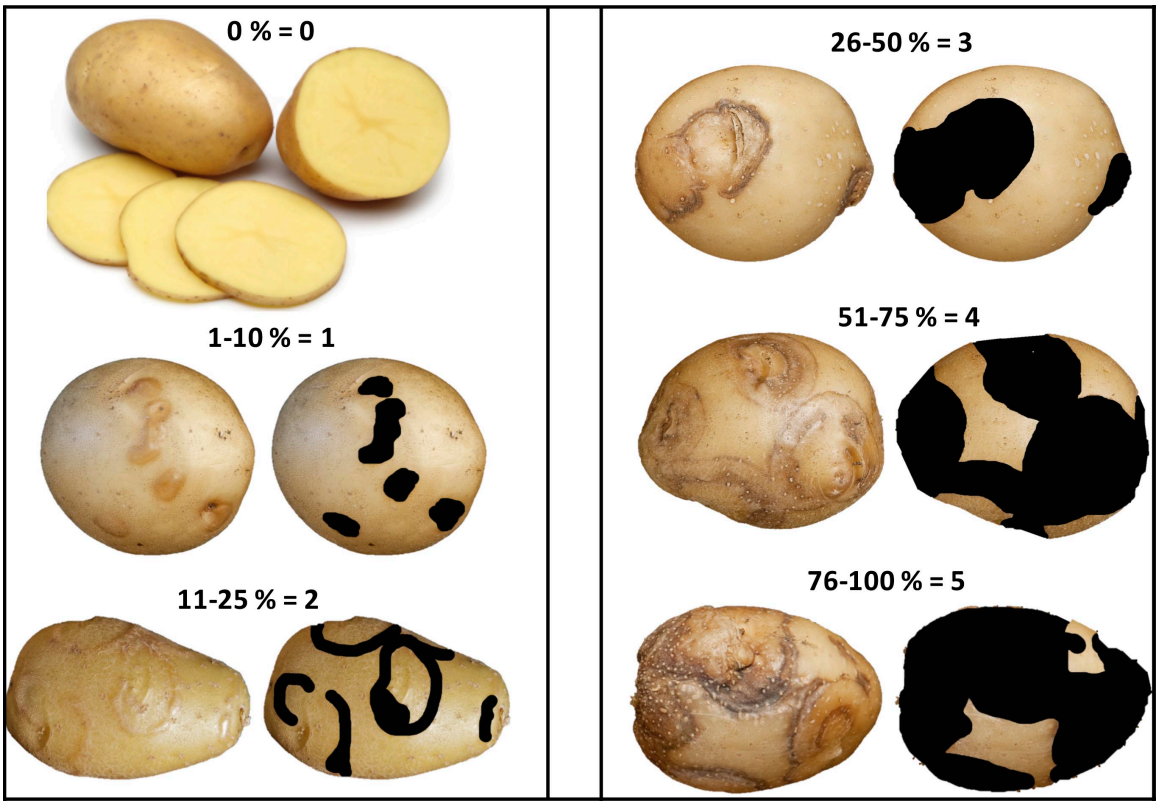


Figure 2 – PTNRD severity rating of the tubers from 236 clones of the H25 population. Ratings were based on a 0 to 5 scale with 0 = no disease and 5 = most severe symptoms.

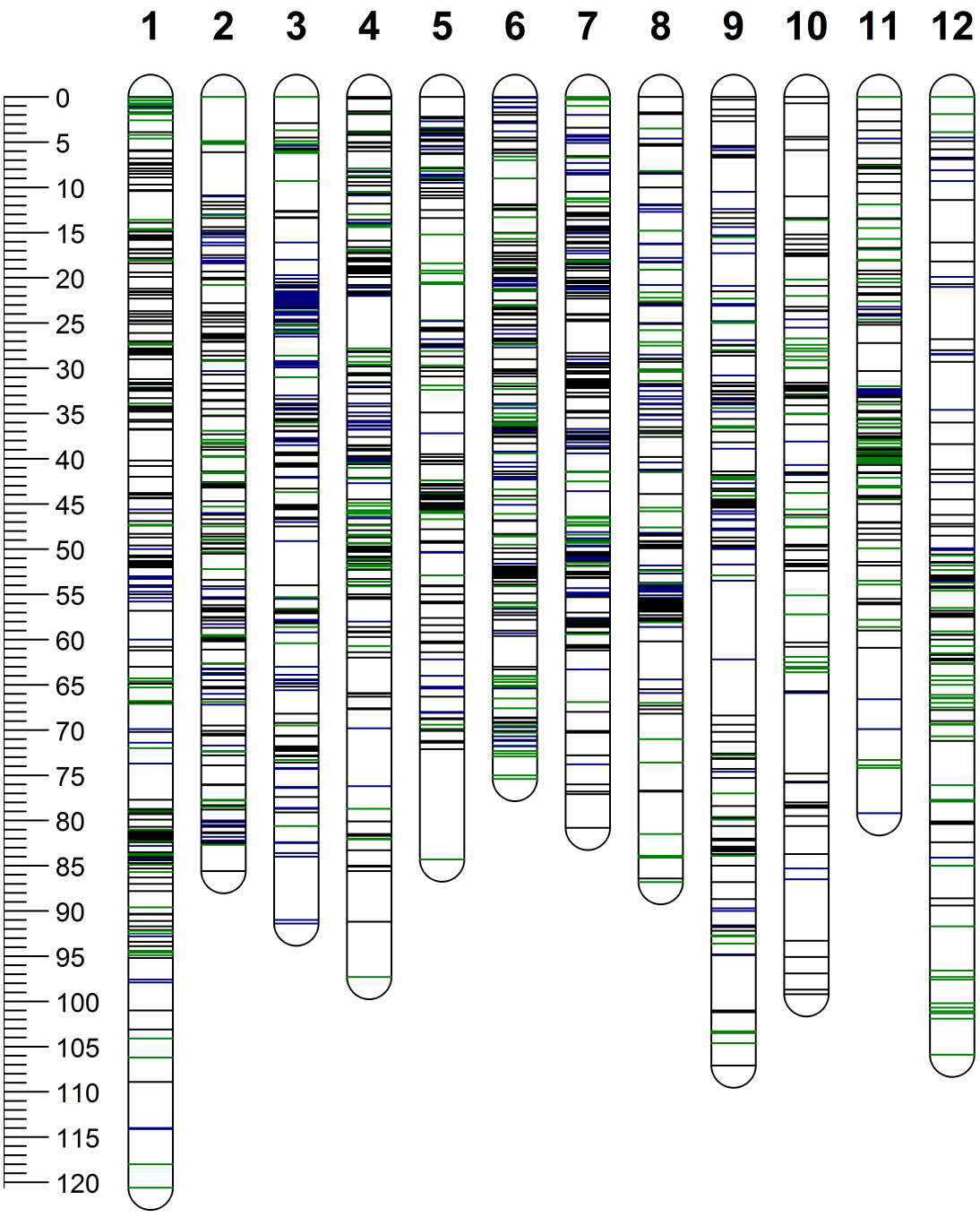


Figure 3 – Distribution of single-nucleotide polymorphism (SNP) markers on 12 chromosomes (1-12) of the parents (Waneta and Pike). The scale bar shows the genetic distance in cM. SNPs positions are represented by green lines (Waneta), blue lines (Pike), and black lines (both parents) across each chromosome.

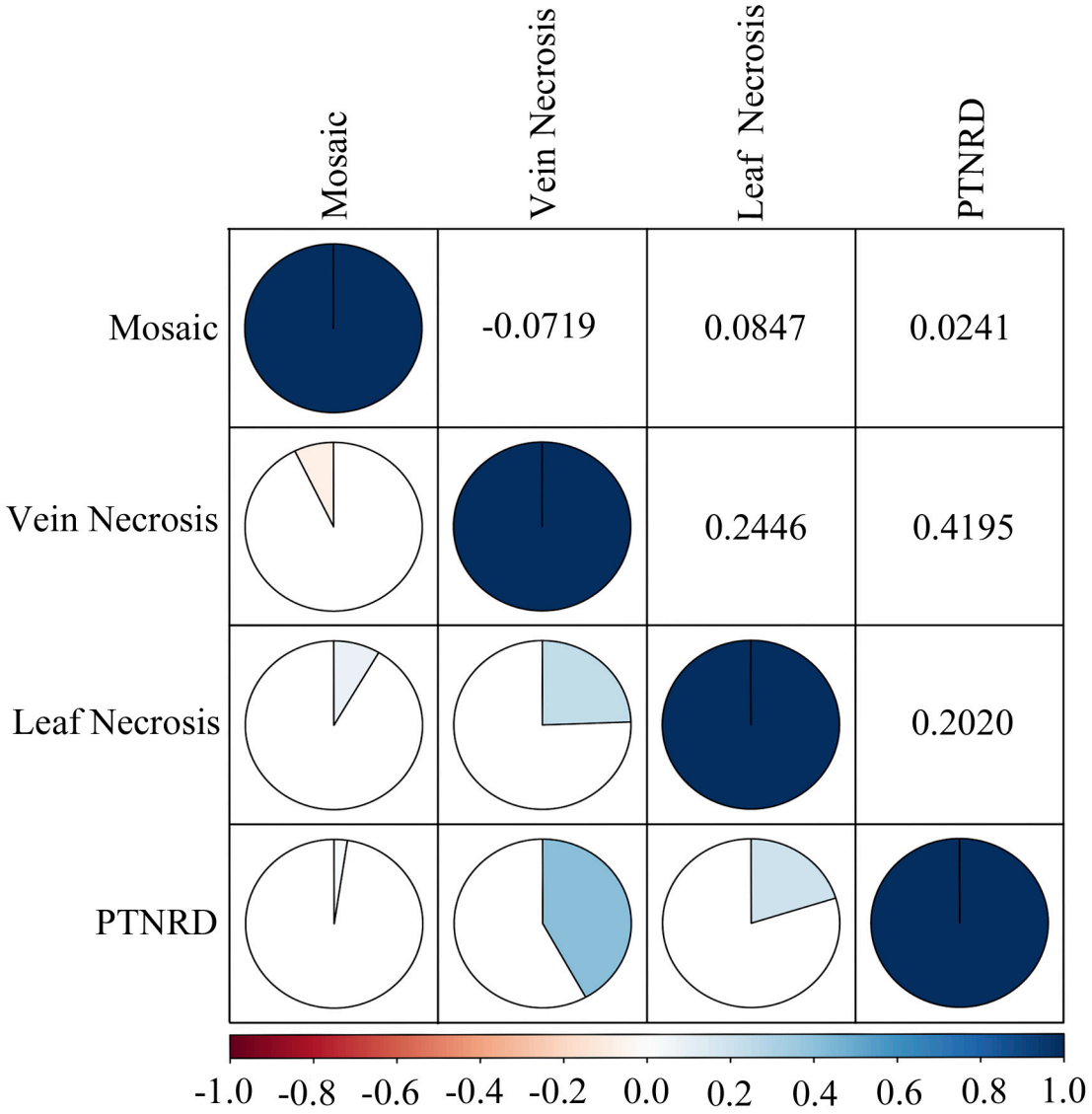


Figure 4 – Pairwise correlation analyses using the non-parametric Kendall's tau rank correlation coefficient to measure the strength of the relationship between each type of symptom expression. Positive correlations are displayed in blue and negative correlations in red. Color intensity is proportional to the correlation coefficients.

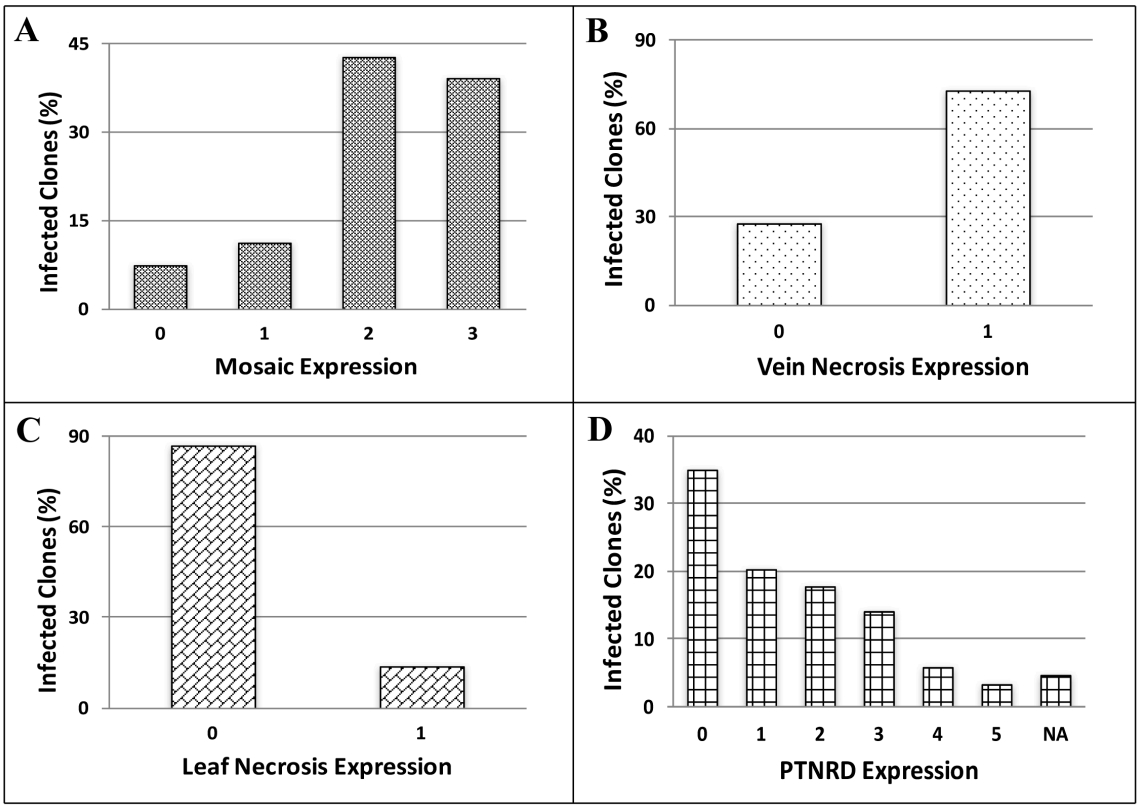


Figure 5 – Percentage of Waneta x Pike offspring expressing varying degrees of foliar (A) Mosaic, (B) Vein Necrosis, (C) Leaf Necrosis, and tuber symptoms (D) PTNRD. See Figures 1 and 2 for symptom scales. NA = number of clones that did not produce tubers; PTNRD was not evaluated with them.

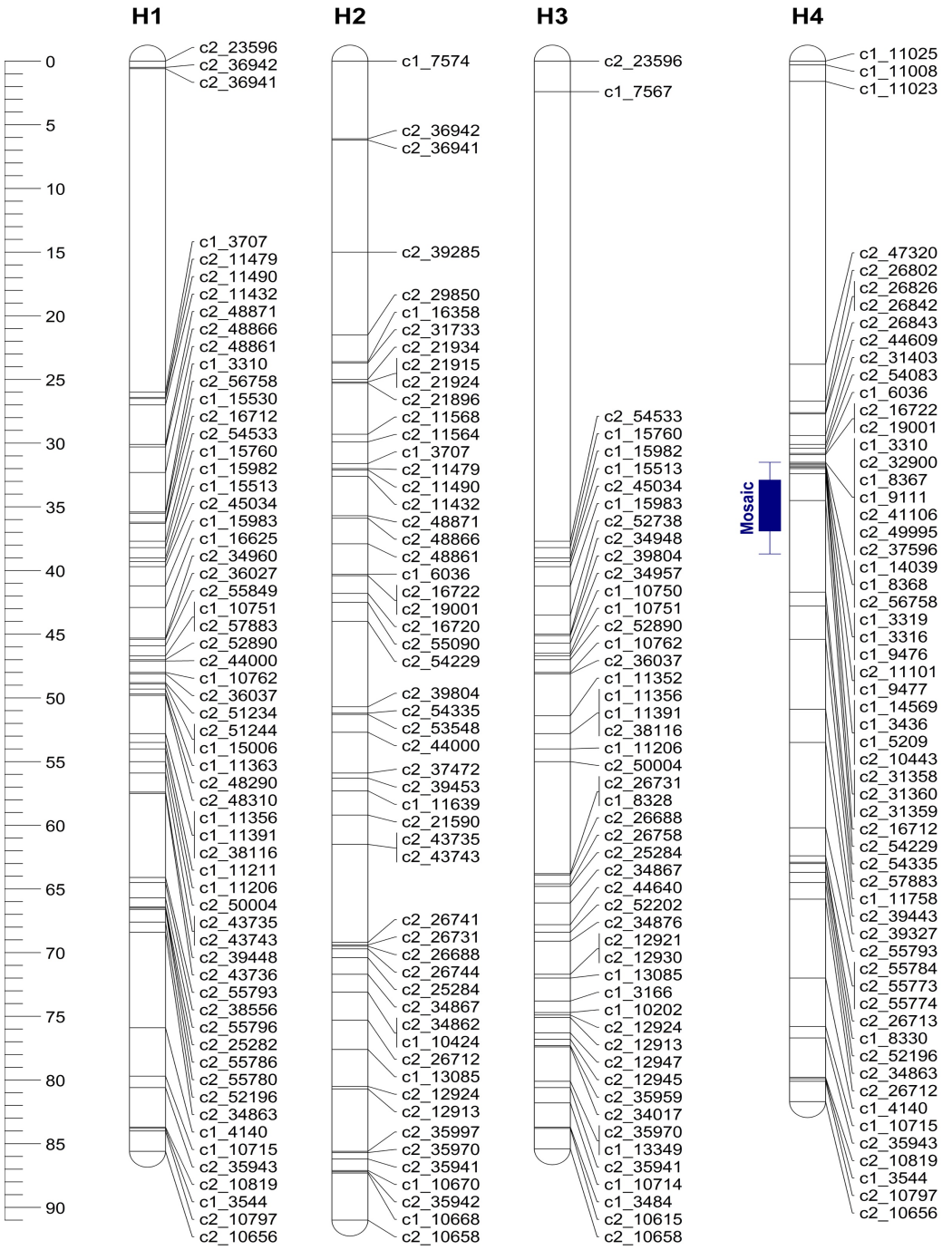


Figure 6 – Linkage map of Waneta chromosome 4 (H1-H4 = homologous maps). The blue bar corresponds to the 95% support LOD interval for the QTLs locations for leaf necrosis and mosaic, respectively. Whiskers represent the two LOD support interval and the solid box represents the one LOD support interval for the QTL location.

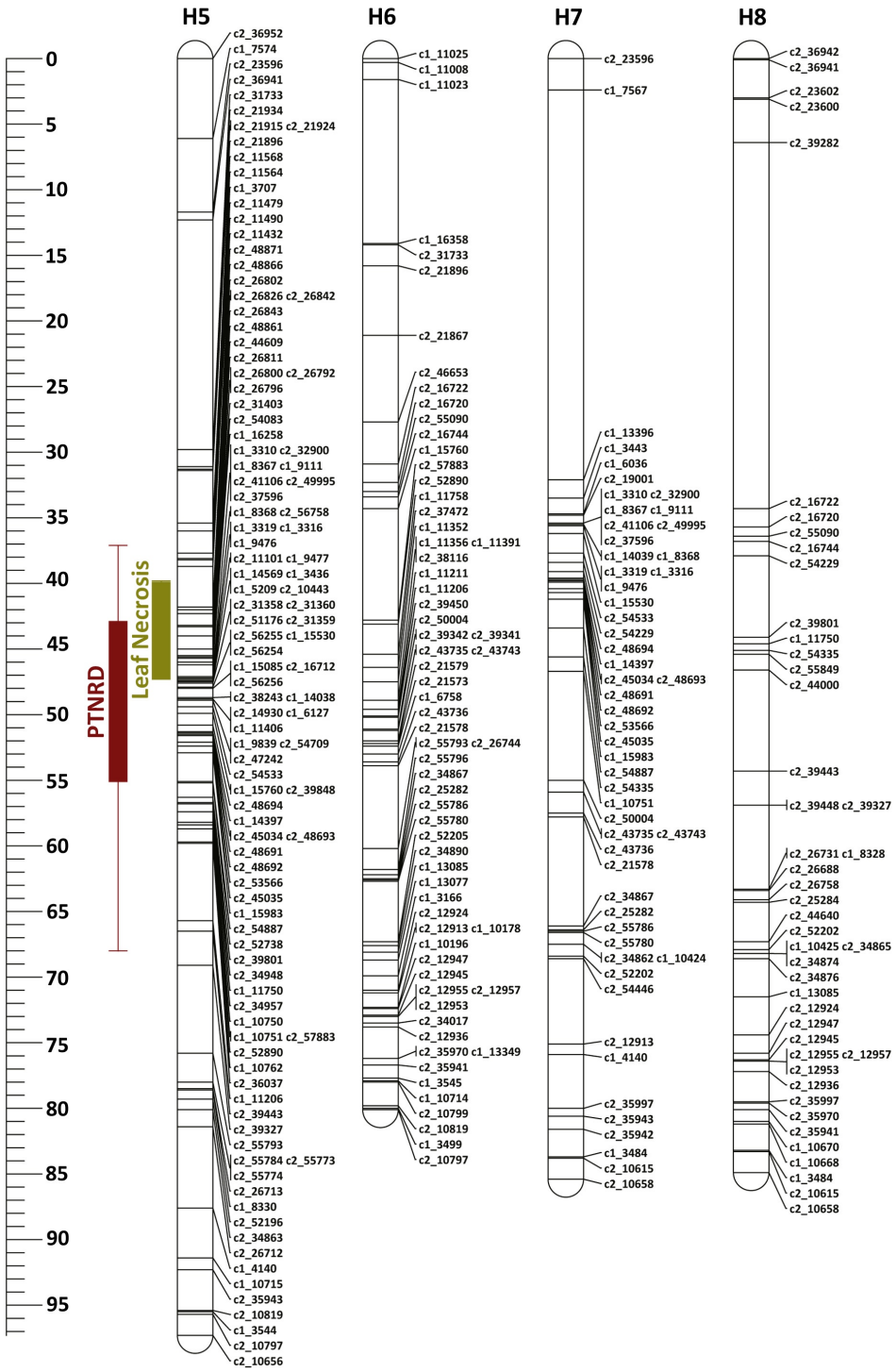


Figure 7 – Linkage map of Pike chromosome 4 (H5-H8 = homologous maps). The brown and olive green bars correspond to the support LOD intervals for the QTL locations for PTNRD and Leaf Necrosis, respectively. For each bar, whiskers represent the two LOD support interval and the solid box represents the one LOD support interval for the QTL location.

CHAPTER 3

QTL Analysis of Tuber Dormancy in Autotetraploid Potato (*Solanum tuberosum* L.)

Abstract

Potato tuber dormancy is an economically important trait. It plays a large role in determining the storability of tubers for the processing industry and fresh market. Therefore, there is a great need to understand mechanisms and processes underlying tuber dormancy in potato. To facilitate the mapping of loci that control dormancy in elite potato germplasm, a cross was made between the cultivars Waneta (long tuber dormancy) and Superior (intermediate tuber dormancy). Both parents and 197 F_1 progeny were genotyped with 12,808 Potato SNP markers. The time until dormancy break, measured as weeks-to-sprouting after harvest, ranged from five to 15 weeks in the population. Waneta and Superior broke dormancy at 15 and seven weeks, respectively. QTL analyses revealed major-effect QTLs for dormancy on chromosomes 2, 3, 5, 6, and 8. The chromosome 8 QTL explained 34.1% of the trait variance. The position of the chromosome 8 locus encompasses a gene, 1-deoxy-D-xylulose 5-phosphate synthase (DXS), that has previously been implicated in dormancy as overexpression of bacterial DXS results in early tuber sprouting. Mapping QTLs associated with potato tuber dormancy may make it possible to manipulate tuber dormancy length as needed, e.g., breeding short dormancy varieties for tropical countries and long dormancy varieties for temperate regions.

Introduction

Potato tuber dormancy plays a vital role in the ability of most potato varieties to overwinter, reducing the chances of the tubers being killed by unfavorable winter conditions. Moreover, dormant tubers are highly resistant to pathogen attacks (Aksenova et al. 2013). It is also important economically as it dictates how long varieties can be stored before processing (Gebhardt et al. 2014). The length of tuber dormancy differs among potato varieties (Bogucki and Nelson 1980; Bamberg 2010; Van Ittersum 1992). The potato processing industry currently controls sprouting by using chemical sprout suppressants and/or by storing tubers at low temperature. However, concerns have been raised about both the toxicity and the persistence of chemical sprout inhibitors; while low temperature storage results in an undesirable accumulation of reducing sugars, which in turn increase the incidence of brown color on processed products (Paul et al. 2016).

Abscisic acid, several different physiological processes (e.g., phytohormones interactions) and epistatic interactions have all been implicated in the regulation of potato tuber dormancy (Aksenova et al. 2013; Suttle 2007). Most studies on the genetics of tuber dormancy to date have been conducted at the diploid level (van den Berg et al. 1996; Freyre et al. 1994; Simko et al. 1997). Although this simplifies genetic analyses, the relevance to commercial potato breeding is less clear, as the alleles segregating in diploid potatoes are often not the same as those that segregate in elite tetraploid germplasm. One prior study in tetraploid potato, by Bradshaw et al. 2008, used AFLP and SSR markers to identify QTLs on six different potato chromosomes that explain variation in the length of tuber dormancy.

The development of new high-throughput genotyping technologies, such as the Illumina Infinium 8300 potato SNP Array (Felcher et al. 2012), combined with recent improvements in mapping software (Hackett et al. 2017), now make it possible to construct high density linkage maps and conduct routine QTL mapping in autotetraploid potato. Here, using the robust 12,808 Potato SNP array and the TetraploidSNPMap software suite, we built a high-density linkage map and located QTLs for tuber dormancy in a segregating autotetraploid potato population.

Materials and Methods

Plant Material

The mapping population NYQ1 consist of 197 F_1 plants resulting from a cross between the cultivars Waneta (as female) and Superior (as male). Tubers of Superior typically break dormancy seven weeks after harvest, whereas tubers of Waneta can take up to 15 weeks for dormancy to break (Fig. 1). True potato seeds of NYQ1 were spread on a bed containing Cornell potting mix (Boodley and Sheldrake 1982). After one month, seedlings were individually transplanted to 15 cm clay pots. Each seedling was vegetatively propagated via cuttings to increase the number of plants per genotype. One tuber of each parent was individually planted in a 15 cm clay pot and cuttings were taken from the sprouts, also to increase the number of plants per genotype. All cuttings, from parents and progeny, were dipped in Hormex rooting hormone #1 Brooker Chem. Corp. (Chatsworth, CA) and planted individually in single wells in trays of 96 wells containing Cornell soil mix for rooting. One month later, four well-rooted cuttings from each clone

as well as the parents were transplanted individually into four-liter plastic pots containing Cornell soil mix. Plants were grown to maturity, harvested, and tubers kept in cold storage room for six months. From germination to harvest, plants were maintained in an insect-free greenhouse under 16:8h light:dark conditions at 25±3^{o}C.

Measuring dormancy

Tubers of both parents and all progeny were planted as 4-hill plots in the field in 2015. At harvest, 10 tubers of 100-200g from each genotype were washed, placed in individual trays, stored in the dark at room temperature, and then evaluated weekly for presence of sprouts. Dormancy was considered broken when five of the 10 tubers from each genotype had developed 6 mm or longer sprouts. A histogram of dormancy length for all genotypes was plotted in R (R Core Team 2016).

SNP genotyping

DNA from progeny and parents was extracted from frozen leaf tissue using a QIAGEN DNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA), following the manufacturer's directions. DNA was quantified with the Quant-it PicoGreen assay (Invitrogen, San Diego, CA-USA) and adjusted to a concentration of 50ng μL^{-1}. The population and parents were genotyped with the Illumina Infinium V2 Potato SNP Array (12,808 SNPs: original SolCAP Infinium 8303 Potato SNP Array with 4,500 additional SNPs to increase coverage in candidate genes and R-gene hotspots) (Hamilton et al. 2011). Illumina GenomeStudio software (Illumina, Inc., San Diego, CA, USA) was used for

initial sample quality assessment and generating marker theta values (which provide allelic dosage information for parents and offspring). In an autotetraploid mapping population, five allele dosages (AAAA, AAAB, AABB, ABBB, and BBBB) are possible and are expected to consist of theta scores in five clusters, centering around 0.0, 0.25, 0.50, 0.75, and 1.0, respectively. Tetraploid (5-cluster) genotyping was based on theta value thresholds, using a custom script from the SolCAP project (Hirsch et al. 2013). Using this script, 5-cluster calling and filtering were performed to remove low quality markers and markers with multiple hits to the potato genome sequence of *Solanum tuberosum* group Phureja DMI-3 516 R44 (Sharma et al. 2013). SNPs with >20% missing genotype calls in the population were excluded from the dataset.

Linkage map construction and QTL analysis were performed in TetraploidSNPMap (Hackett et al. 2017) as detailed in da Silva et al. 2017. For each trait, interval mapping displayed the LOD profile on the chromosome, giving the LOD score statistics, percentage variation explained, and QTL effect for each homologous chromosome. 90%, 95%, and 99% LOD thresholds were obtained to establish the statistical significance of each QTL position using genome-wide permutation test (Churchill and Doerge 1994) with 992 permutations (Hackett et al. 2014). Simple models for the genotype means estimated at the most probable QTL position were calculated using the Schwarz Information Criterion (SIC) (Schwarz 1978); models with the lowest value for SIC are considered the best models (Hackett et al. 2014). Linkage maps and QTL positions were plotted in MapChart 2.30 (Voorrips 2002). Concordance between the linkage maps generated in this study and the potato reference genome (PGSC Version 4.03 Pseudomolecules) was evaluated in MareyMap R package version 1.3.1 (Rezvoy et al. 2007). Plots of

the genetic position (cM) with the physical position (Mb) of each SNP marker in each chromosome were generated using the graphical interface MareyMapGUI, the interpolation method "cubic splines" was used to calculate the curve slope.

Data availability

All the raw data from this study are available in supplemental .txt files: Table S1 and Table S2. Complementary information for the Results and Discussion section are provided in Support Information: Figures S1, S2, and S3, and Tables S3 and S4.

RESULTS AND DISCUSSION

Phenotypic evaluations

Tubers from the female (Waneta) and male (Superior) parents broke dormancy at 15 and seven weeks after harvest, respectively. The tuber dormancy of Superior is typical (plus or minus a few weeks) of most potato cultivars, while the dormancy of Waneta is unusually long. Dormancy in their offspring ranged from five to 15 weeks with a mean of nine weeks (Figure 1). The continuous, normal distribution-like nature of tuber dormancy in the offspring suggests that this trait is controlled by multiple genes.

Data processing pre-genotyping

After a pre-filtering step to remove SNPs with missing theta values, SNPs of low quality, and SNPs with multiple hits to the potato reference genome PGSC Version 4.03 Pseudomolecules

(Table S2), 4364 of the 12808 SNPs on the new Illumina Potato V2 SNP Array (12K) were selected for downstream analyses. Of these, 847 SNPs had missing data in >20% of the population, and these were also excluded. The remaining 3517 SNPs were loaded into TetraploidSNPMap. Subsequently, 845 SNPs with unexpected segregation ratios (ie with chi-square significance statistics for goodness of fit less than 0.001) were removed. Hierarchical clustering grouped the remaining 2672 markers into 12 linkage groups. A total of 156 SNPs were flagged as duplicated and a further 203 were excluded as outliers after clustering, 2-point, and multidimensional scaling (MDS) analyses (Table S3). Four SNPs previously annotated as located on chromosome 4 were misplaced on chromosome 7 by TetraploidSNPMap. However, they were identified as outliers by MDS analysis and were manually returned to chromosome 4. After running a new MDS analysis, these SNPs were confirmed to be on chromosome 4, exhibiting LOD score relationships of more than 20 with other chromosome 4 SNP markers.

A total of 62% (1436) of the markers followed simplex (AAAA X AAAB, AAAB X AAAA), duplex (AAAA X AABB, AABB X AAAA), and double-simplex (AAAB X AAAB, ABBB X ABBB) configurations, while 38% (877) exhibited triplex (AAAA X ABBB, AAAB X BBBB) and other complex configurations less ideal for linkage and QTL mapping (Table S4). The large number of markers and the wide range of parental genotype configurations in this dataset increase the likelihood of detecting significant QTLs for the trait studied (da Silva et al. 2017).

Linkage map construction

Chromosomes 1 and 12 had the highest and the lowest number of mapped SNPs (250 and 104), respectively, out of the 2313 SNPs mapped to the 12 potato chromosomes (Table 1). A total of

1729 SNPs segregated in “Waneta”, 1737 segregated in “Superior”, and 794 SNPs segregated in both parents (Table 1). The total genetic distance for each parental map was 1061.9 cM (for Waneta) and 1052 cM (for Superior), with the map lengths of individual chromosomes ranging from 67.7 to 143.9 cM. The genetic maps of both parents covered, on average, 99% of the PGSC v4.03 Pseudomolecules. There was an average of 144 SNP markers per chromosome and a marker density of ~1.64 SNPs per cM (Table 1 and Figure 2). Analyses of the concordance between genetic and physical maps for chromosomes 8 (Figure S1) and all other chromosomes (data not shown) generated graphs with expected shapes, as has also been observed in other studies (Massa et al. 2015; Felcher et al. 2012; Sharma et al. 2013; da Silva et al. 2017).

Major-effect QTLs for dormancy breakage detected on multiple chromosomes

Significant QTLs for length of tuber dormancy were identified on chromosomes 2, 3, 5, 6, and 8 (Figure 2 and Table 2). The chromosome 5 and 8 QTL were detected on single homologs of the male parent (Superior) while the QTL on chromosomes 2, 3, and 6 were detected on two homologs each (Table 2). The 90%, 95%, and 99% LOD genome-wide permutation thresholds were 4.9, 5.5, and 6.3, respectively, after running 992 permutations.

On chromosome 2, the peak of the QTL was located at 75 cM and explained 9.3% of the trait variance. This QTL’s LOD score of 5.9 was above the upper 95% LOD permutation threshold of 5.5. Analyses of different simple genetic models were performed with TetraploidSNPMap to determine the best simple model that fits the trait. The best model was a duplex (AAAA x ABAB) configuration on homologous chromosomes H6 and H8 of Superior, with the B allele associated with a decrease of tuber dormancy length (Figure 2, S2 and Table 2). This model had

the lowest SIC, -5, in comparison with the full model (SIC = 9). Solcap_snp_c2_4505, with configuration (TTTT X TCTC), was the closest SNP to the QTL peak position (Table 3). This SNP is in a cytochrome P450 (CYP) gene. It has been reported that CYPs are upregulated during the transition from the dormant to the non-dormant state of potato tubers (Campbell et al. 2008), an indication that an oxidative regime is activated in meristems after dormancy breakage. However, the role of such genes in controlling dormancy length in potato tubers has not been fully elucidated (Aksenova et al. 2013).

On chromosome 3, the maximum LOD score was 5.1 and the QTL explained 7.3% of the phenotypic variance. The LOD peak was located at 76 cM and was above the upper 90% LOD permutation threshold of 4.9. The simple models analyses estimated a double-simplex genotype (AABA X AABA) for the trait on homologous chromosomes H4 and H7. Both parents contributed the B allele, which is associated with a decrease of tuber dormancy length. The SIC for this model was -7, while the full model had SIC = 3. The closest SNP to the QTL peak is solcap_snp_c1_7132 (superscaffold PGSC0003DMT400046769) configuration (TTCT X TTCT), which is located in an uncharacterized CASP membrane protein (Table 3). The QTL found on chromosome 5 of Superior was a simplex allele (AAAA x AAAB) on homologous chromosome H8, with the B allele associated with a decrease in tuber dormancy length and the SIC for this model was 9 whereas the full model SIC was 19. This QTL was located at 22 cM and explained 5.7% of the trait variance. The LOD score was 5.0, which was above the upper 90% LOD permutation threshold of 4.9. The closest SNP to the QTL peak is solcap_snp_c2_11829 configuration (TTTT X TTTC).

The QTLs located on chromosomes 6 and 8 explained the most trait variance, 10.9% and 34.1%, respectively. The LOD peaks were at 13 cM and 91 cM, with maximum LOD scores of 6.8 and 19.9, above the upper 99% LOD permutation thresholds of 6.3 for chromosomes 6 and 8, respectively. The simple models analyses estimated a double-simplex configuration for chromosome 6 and a simplex configuration for chromosome 8. The best models were BAAA X ABAA on homologous chromosomes H1 and H6 (Figures S6 and S7) for chromosome 6 and AAAA X BAAA configuration on homolog H5 for chromosome 8. Those models had SIC values of -4 and -11 in comparison with the full model SIC values of 12 and -1 for chromosomes 6 and 8, respectively. On chromosome 6 allele B is associated with an increase of tuber dormancy, while allele B is linked to a decrease of dormancy on chromosome 8.

The SNP located closest to the QTL peak on chromosome 6 was solcap_snp_c2_8822 configuration (ACCC X CACC), which is in a gene for 60S ribosomal protein L44. A simplex SNP [solcap_snp_c2_51367 - configuration (TTTT X ATTT)] was located in close proximity to the QTL peak on chromosome 8. This SNP mapped to the distal end of chromosome 8, and is located in a 1-deoxy-D-xylulose 5-phosphate synthase (DXS) gene (Table 3). DXS is the precursor of thiamine and pyridoxol in one of the isoprenoid biosynthetic pathways (Walter et al. 2002). Interestingly, overexpressing a bacterial DXS in potato triggers disturbances in the isoprenoid metabolic network and the end result is an early tuber sprouting phenotype (Morris et al. 2006). Marker-assisted selection with solcap_snp_c2_51367 may prove useful in selecting for large differences in tuber dormancy. Further testing of the role of potato DXS gene to investigate its possible association with tuber sprouting is also warranted.

CONCLUSION

In tropical climates where at least two potato crops are grown a year, varieties with short (or no) dormancy are highly desired. On the other hand, temperate climates need longer dormancy varieties for the extended storage period imposed by long-lasting winters. Similarly, the potato processing industry favors varieties that can be stored for a long time to provide raw material year-round. Identifying the loci that influence dormancy, as well as the magnitude and direction of each allele's effect, will be useful for marker-assisted selection (MAS) programs that aim to develop genotypes adapted for a particular climate or end use. This study has reported QTL for dormancy on multiple chromosomes of both Waneta and Superior, two of which (on chromosomes 6 and 8), are of relatively large effect. The SNPs closest to each QTL peak can serve as tools for breeders who wish to more efficiently develop new potato varieties with short, long, or intermediate dormancy.

ACKNOWLEDGEMENTS

The authors would like to thank Daniel Zarka for valuable help with genotyping and Matthew Falise for assistance in phenotyping the NYQ1 population. This study was funded in part by USDA-SCRI (2009-51181-05894 and 2014-51181-22373) and by a UK Biotechnology and Biological Sciences Research Council grant # BB/L011840/1 as part of the joint USDA-NSF-NIH-BBSRC Ecology and Evolution of Infectious Diseases program. The work of Glenn Bryan, Christine Hackett, and the TetraploidSNPMap software development were funded by the Rural & Environment Science & Analytical Services Division of the Scottish Government. Washington da

Silva was partially supported by a SUNY Diversity Fellowship and by a 2014/2015 National Potato Council (NPC) scholarship.

LITERATURE CITED

Aksenova, N.P., L.I. Sergeeva, T.N. Konstantinova, S.A. Golyanovskaya, O.O. Kolachevskaya *et al.*, 2013 Regulation of potato tuber dormancy and sprouting. *Russian Journal of Plant Physiology* 60 (3):301-312.

Bamberg, J.B., 2010 Tuber Dormancy Lasting Eight Years in the Wild Potato *Solanum jamesii*. *American Journal of Potato Research* 87 (2):226-228.

Bogucki, S., and D.C. Nelson, 1980 Length of dormancy and sprouting characteristics of ten potato cultivars. *American Potato Journal* 57 (4):151-157.

Boodley, J.W., and R. Sheldrake, 1982 Cornell peat-like mixes for commercial plant growing. *Cornell Inf. Bull. 43.*

Campbell, M., E. Segear, L. Beers, D. Knauber, and J. Suttle, 2008 Dormancy in potato tuber meristems: chemically induced cessation in dormancy matches the natural process based on transcript profiles. *Functional & Integrative Genomics* 8 (4):317-328.

Churchill, G.A., and R.W. Doerge, 1994 Empirical threshold values for quantitative trait mapping. *Genetics* 138 (3):963.

da Silva, W., J. Ingram, C.A. Hackett, J.J. Coombs, D. Douches *et al.*, 2017 Mapping Loci that Control Tuber and Foliar Symptoms Caused by PVY in Autotetraploid Potato (*Solanum tuberosum* L.). *G3: Genes|Genomes|Genetics Accepted*.

Felcher, K.J., J.J. Coombs, A.N. Massa, C.N. Hansey, J.P. Hamilton *et al.*, 2012 Integration of Two Diploid Potato Linkage Maps with the Potato Genome Sequence. *PLoS ONE* 7 (4):e36347.

Freyre, R., S. Warnke, B. Sosinski, and D.S. Douches, 1994 Quantitative trait locus analysis of tuber dormancy in diploid potato (Solanum spp.). *Theoretical and Applied Genetics* 89 (4):474-480.

Gebhardt, C., C. Urbany, and B. Stich, 2014 Dissection of Potato Complex Traits by Linkage and Association Genetics as Basis for Developing Molecular Diagnostics in Breeding Programs, pp. 47-85 in *Genomics of Plant Genetic Resources: Volume 2. Crop productivity, food security and nutritional quality*, edited by R. Tuberosa, A. Graner and E. Frison. Springer Netherlands, Dordrecht.

Hackett, C.A., B. Boskamp, A. Vogogias, K.F. Preedy, and I. Milne, 2017 TetraploidSNPMap: Software for Linkage Analysis and QTL Mapping in Autotetraploid Populations Using SNP Dosage Data. *J Hered* 00 (00):1-5.

Hackett, C.A., J.E. Bradshaw, and G.J. Bryan, 2014 QTL mapping in autotetraploids using SNP dosage information. *Theoretical and Applied Genetics* 127 (9):1885-1904.

Hamilton, J., C. Hansey, B. Whitty, K. Stoffel, A. Massa *et al.*, 2011 Single nucleotide polymorphism discovery in elite north american potato germplasm. *BMC Genomics* 12 (1):302.

Hirsch, C.N., C.D. Hirsch, K. Felcher, J. Coombs, D. Zarka *et al.*, 2013 Retrospective View of North American Potato (*Solanum tuberosum* L.) Breeding in the 20th and 21st Centuries. *G3: Genes|Genomes|Genetics* 3 (6):1003-1013.

Massa, A.N., N.C. Manrique-Carpintero, J.J. Coombs, D.G. Zarka, A.E. Boone *et al.*, 2015 Genetic Linkage Mapping of Economically Important Traits in Cultivated Tetraploid Potato (*Solanum tuberosum* L.). *G3: Genes|Genomes|Genetics* 5 (11):2357-2364.

Morris, W.L., L.J.M. Ducreux, P. Hedden, S. Millam, and M.A. Taylor, 2006 Overexpression of a bacterial 1-deoxy-D-xylulose 5-phosphate synthase gene in potato tubers perturbs the isoprenoid metabolic network: implications for the control of the tuber life cycle. *Journal of Experimental Botany* 57 (12):3007-3018.

Paul, V., R. Ezekiel, and R. Pandey, 2016 Sprout suppression on potato: need to look beyond CIPC for more effective and safer alternatives. *Journal of Food Science and Technology* 53 (1):1-18.

R Core Team, 2016 A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rezvoy, C., D. Charif, L. Guéguen, and G.A.B. Marais, 2007 MareyMap: an R-based tool with graphical interface for estimating recombination rates. *Bioinformatics* 23 (16):2188-2189.

Schwarz, G., 1978 Estimating the Dimension of a Model. *The Annals of Statistics* 6 (2):461-464.

Sharma, S.K., D. Bolser, J. de Boer, M. Sønderkær, W. Amoros *et al.*, 2013 Construction of Reference Chromosome-Scale Pseudomolecules for Potato: Integrating the Potato Genome with Genetic and Physical Maps. *G3: Genes|Genomes|Genetics* 3 (11):2031-2047.

Simko, I., S. McMurry, H.M. Yang, A. Manschot, P.J. Davies *et al.*, 1997 Evidence from Polygene Mapping for a Causal Relationship between Potato Tuber Dormancy and Abscisic Acid Content. *Plant Physiology* 115 (4):1453-1459.

Suttle, J.C., 2007 Chapter 14 - Dormancy and Sprouting, pp. 287-309 in *Potato Biology and Biotechnology*. Elsevier Science B.V., Amsterdam.

van den Berg, J.H., E.E. Ewing, R.L. Plaisted, S. McMurry, and M.W. Bonierbale, 1996 QTL analysis of potato tuber dormancy. *Theoretical and Applied Genetics* (93):317-324.

Van Ittersum, M.K., 1992 Variation in the duration of tuber dormancy within a seed potato lot. *Potato Research* 35 (3):261-269.

Voorrips, R.E., 2002 MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity* 93 (1):77-78.

Walter, M.H., J. Hans, and D. Strack, 2002 Two distantly related genes encoding 1-deoxy-d-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *The Plant Journal* 31 (3):243-254.

Tables:

Table 1 – Summary of the parental linkage maps, Waneta (Wan) and Superior (Sup).

	No. Mapped SNPs			Map Length, (cM)		Map Length, (Mb[a])		PGSC v4.03 PM, (Mb[a])	Map Coverage[a]		Average Interloci Distance, (cM)	
Chr	Total	Wan	Sup	Wan	Sup	Wan	Sup	DM	Wan	Sup	Wan	Sup
1	250	194	194	114.6	102.6	88.5	88.5	88.7	1.00	1.00	0.59	0.53
2	212	157	157	81.1	79.5	47.7	47.6	48.6	0.98	0.98	0.52	0.51
3	231	196	161	94.3	94.3	61.9	61.9	62.3	0.99	0.99	0.48	0.59
4	205	134	154	87.4	83.1	71.9	72	72.2	1.00	1.00	0.65	0.54
5	154	109	98	70.9	77.2	50.9	52	52.1	0.98	1.00	0.65	0.79
6	228	189	160	79.6	78.4	59.1	58.8	59.5	0.99	0.99	0.42	0.49
7	202	162	168	81.5	79.7	56	56	56.8	0.99	0.99	0.50	0.47
8	224	174	191	139.4	143.9	56.8	56.8	56.9	1.00	1.00	0.80	0.75
9	212	156	163	92.1	88.6	61.5	61.5	61.5	1.00	1.00	0.59	0.54
10	123	89	93	74.4	74.4	59.7	59.7	59.8	1.00	1.00	0.84	0.80
11	168	92	127	67.7	74.5	45.1	45.2	45.5	0.99	0.99	0.74	0.59
12	104	77	71	78.9	75.8	59.6	61.1	61.2	0.97	1.00	1.02	1.07
Total	**2313**	**1729**	**1737**	**1061.9**	**1052**	**718.7**	**721.1**	**725.1**	**0.99**	**0.99**	**0.61**	**0.61**

[a]Map length (Mb) and map coverage values are based on the PGSC Version 4.03 Pseudomolecules of the potato reference genome *Solanum tuberosum* group Phureja DM1-3 516 R44 (DM).

Table 2 – QTL for dormancy in the NYQ1 population.

Chr	QTL position (cM)	LOD	Variance Explained (%)	Model Homologous Chr
2	75	5.9	9.3	h6+h8 (-)
3	76	5.1	7.3	h4+h7 (-)
5	22	5.0	5.7	h8 (-)
6	13	6.8	10.9	h1+h6 (+)
8	91	19.9	34.1	h5(-)

QTL – quantitative trait loci, Chr – chromosome, LOD – logarithm of the odds, SIC - Schwarz Information Criterion.

Table 3 – SNP markers most closely linked to QTL for tuber dormancy

Chr	SNP Name*	Physical Position (Mb)	Superscaffold	Protein Match
2	solcap_snp_c2_4505	8.2	PGSC0003DMT400027430	Cytochrome P450
3	solcap_snp_c1_7132	54.3	PGSC0003DMT400046769	Membrane Protein
5	solcap_snp_c2_11829	5.7	PGSC0003DMT400065791	Conserved gene of unknown function
6	solcap_snp_c2_8822	52.4	PGSC0003DMT400069671	60S ribosomal protein L44
8	solcap_snp_c2_51367	39.9	PGSC0003DMT400074817	1-deoxyxylulose-5-phosphate synthase

*All SNP information (physical position of the SNPs, gene superscaffold, and protein match) are based on the current annotation of the PGSC Version 4.03 Pseudomolecules of the potato reference genome *Solanum tuberosum* group Phureja DM1-3 516 R44 (DM), Chr – chromosome.

Figures:

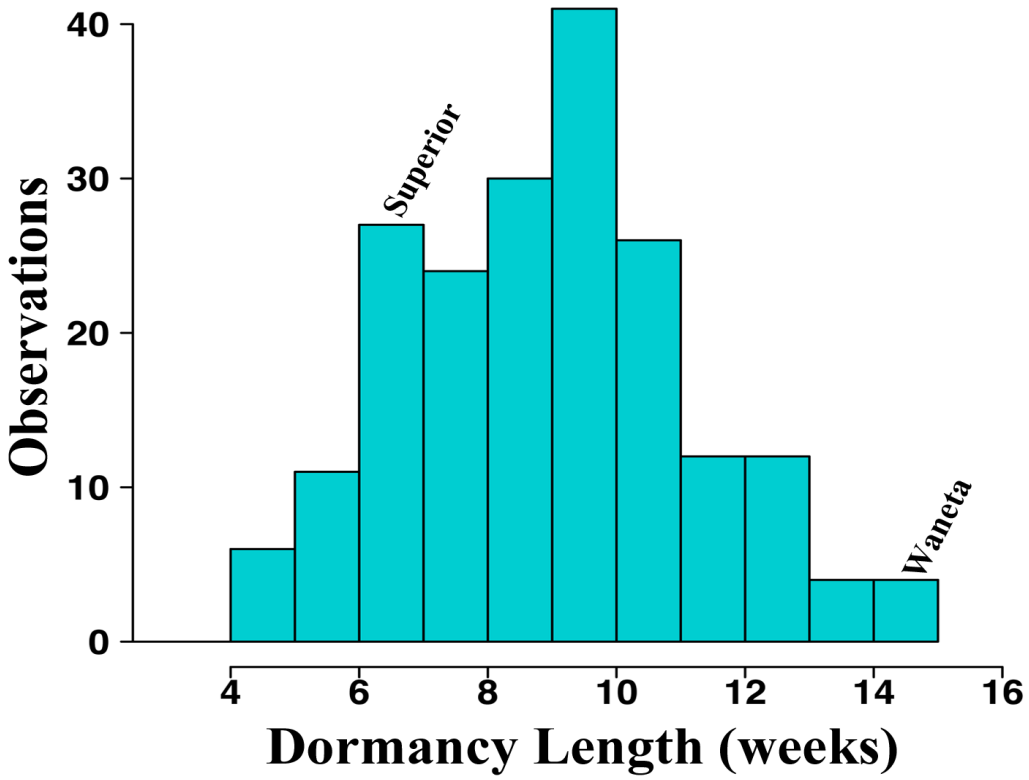


Figure 1 – Frequency distribution of dormancy length of tubers from NYQ1 progeny and parents. Waneta and Superior are the female and male parents, respectively. Length of dormancy was determined when five tubers, from a subsample of 10 tubers from each genotype, had 6-mm-long sprouts. Tubers were checked weekly.

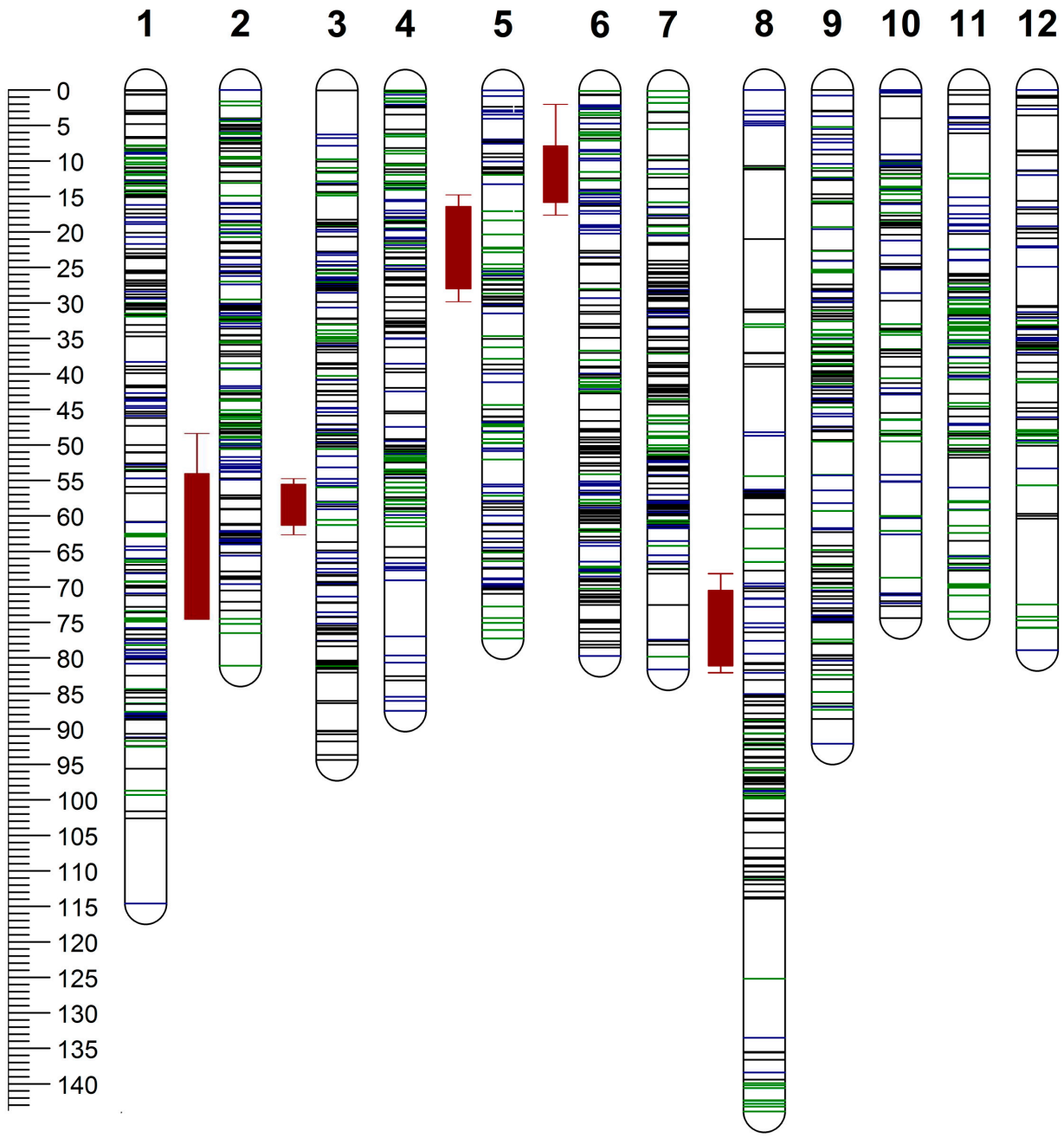


Figure 2 – PTNRD severity rating of the tubers from 236 clones of the H25 population. Ratings were based on a 0 to 5 scale with 0 = no disease and 5 = most severe symptoms.

CHAPTER 4

Illumina deep sequencing provides insights into the population structure of *Potato virus Y* strains

ABSTRACT

Genetic variation among strains and isolates of *Potato virus Y* (PVY) has been intensively studied and found to be extensive; however, little is known about the genetic diversity within single isolate quasispecies. The population structure of 15 PVY isolates were characterized using Illumina high-throughput sequencing coupled with novel and improved bioinformatics tools to identify and compare single nucleotide variants (SNVs) within and among PVY isolates. Seven of the isolates represented five strains of PVY found routinely in most US potato production areas. No SNVs were detected in three of the isolates, while three isolates contained one SNV and three SNVs were identified in the seventh isolate. Eight isolates belonged to the PVY^N strain. Two were collected in Europe and six collected in Montana; one collected in the 1990s, the remaining five in 2011 and 2013. In contrast with to the other strains, the number of SNVs identified in the PVY^N isolates ranged from nine to 56 and were non-uniformly distributed across the genome. The highest number of SNVs were in the CP cistron; no SNVs were identified in the P3, 6K1, or 6K2 cistrons. In national surveys of PVY in potato from 2004-2016, PVY^N isolates have only been recovered from two seed farms in Montana despite the distribution of potatoes from those farms to other production regions. The genetic diversity within the PVY^N isolate quasispecies doesn't seem to convey enhanced fitness, but rather appears to be leading to a reduction of this isolate's incidence in potato fields.

Introduction

Potato virus Y (PVY), the type species of the genus *Potyvirus* in the family *Potyviridae*, is the causal agent of potato mosaic, the most economically important virus disease affecting potato (*Solanum tuberosum* L) (Kerlan and Moury 2008), the third most important staple food crop in the world. Numerous strains of PVY have been described worldwide, and many of them are recombinants of the ordinary strain (PVY^{O}) and the tobacco necrotic strain (PVY^{N}) (Quenouille et al. 2013). Surveys of the U.S. seed potato crop have documented a rapid shift in the PVY strains that predominate in all major seed potato production areas in the U.S. (Karasev and Gray 2013). PVY^{O}, once the predominant strain throughout the U.S., has decreased in incidence and is being replaced by recombinant strains, e.g. PVY^{NWi} and PVY^{NTN}. PVY^{N}, until recently a quarantine pathogen in the U.S., has only been identified from two seed potato farms in Montana. In addition to strain diversification, there is also considerable variability of phenotypes and genotypes among isolates within a strain (Karasev and Gray 2013) Green et al. , but there has been limited investigations of genetic diversity within isolates of PVY.

The large RNA virus population size, high replication rate, and poor proofreading ability of the RNA-dependent-RNA-polymerase are all expected to lead to an increased level of genetic diversity in the viral population or "quasispecies", a concept that viruses are not just a collection of random mutants, but an interactive group of variants (Eigen and Biebricher 1988). Presumably, the diversity within the quasispecies allows the virus to quickly evolve and adapt to new and/or changing environments (García-Arenal et al. 2001; Vignuzzi et al. 2006). Adaptation may lead to a selection or predominance of certain variants within the quasispecies or selection

may be at the scale of the quasispecies (Domingo et al. 2012). A study on antibody target selection in foot-and-mouth disease virus (FMDV) showed that a quasispecies sharing a marked trait, and not a single individual, became dominant in the population (Perales et al. 2005).

Several recent studies have demonstrated the usefulness of high-throughput sequencing(HTS) technology to characterize the genetic structure of potyvirus populations (Kehoe et al. 2014; Zwart et al. 2014; Simmons et al. 2012; Kutnjak et al. 2015; Cuevas et al. 2015; Martínez et al. 2012; Betz-Stablein et al. 2016). In potyviruses, studies report distinct regions in the genome where variations generated by mutation and/or recombinations are tolerated, whereas genotypes with changes outside of these regions are rapidly removed from the quasispecies by selection pressures (Moreno et al. 2004). Other studies report the existence of gene-specific selective pressures acting in potyvirus evolution (Ogawa et al. 2008; Cuevas et al. 2012), with the embedded P3N-PIPO cistron being one of the main targets of selection. Another study reported strong evidence that purifying selection, i.e., selection against nonsynonymous substitutions, is the main evolutionary force affecting PVY (García-Arenal et al. 2001). Genetic diversity and selection pressures on potyvirus quasispecies are apparently diverse and the robustness of HTS makes this technology an ideal tool to investigate these evolutionary forces. Here we report the use of Illumina HTS and subsequent sequence analyses to examine the genetic diversity within PVY populations and compare the quasispecies structure of isolates within and among PVY strains.

1. Materials and Methods

1.1. Isolates and inoculation assay

An ongoing national survey of PVY isolates affecting the US seed potato crop identified six PVY^N isolates collected from potatoes grown in Montana (Montana PVY^N) in 2010 and 2013 (2010 – MT100006 and MT100017; 2013 – MT130014, MT130017, MT130039, and MT130054). The two isolates collected in 2010 originated from sister seed lots of 'Ranger Russet' grown on the same farm. Three additional isolates (MT130014, MT130039, MT130054) were collected from this same farm in 2013 from three different Ranger Russet seed lots. The MT130017 isolate was collected in 2013 from an unrelated Ranger Russet seed lot grown on a different seed farm. PVY^N isolates are rare in the U.S. and only have been collected from these two farms in Montana since survey efforts began in 2004. Two additional PVY^N isolates from Europe were included in the study, N605 (syn. CH-605) (GenBank: X97895.1) (Jakab et al. 1997), and NZ (GenBank: AM268435.1) (Schubert et al. 2007). Both N605 and NZ were originally obtained from Dr. Emmanule Jacquot (INRA, Rennes, France).

Seven additional isolates representing five additional PVY strains collected from potato were also included in this study. These included three non-recombinant isolates of the PVY^O strain, NY090031 and WI3 from New York and Wisconsin, respectively and a PVY^{O5} isolate from Montana, MT100010. PVY^{O5} is a serological and genetic variant within the PVY^O strain (Karasev et al. 2010). Four recombinant isolates included OR16, a PVY^{NO} from Oregon, WA316, a PVY^{NE11} from Washington, ME4, a PVY^{NTN} from Maine, and MN21 a PVY^{NWi} from Minnesota. The strain identity of all the isolates used in these studies was confirmed using two strain-specific RT-PCR multiplex diagnostic assays (Lorenzen et al. 2006a; Ali et al. 2010).

The isolates were maintained long term in lyophilized tobacco tissue stored at -80°C. To generate the virus infected material used in this study, lyophilized tissue was homogenized in

10 volumes of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, and pH adjusted to 7.4 with HCl) and used to mechanically inoculate a fully expanded leaf of a potato seedling, cv 'Goldrush', at the five-to-six-leaf stage, previously tested free from PVY infection by ELISA. Single plants were inoculated with each of the 15 isolates described above and maintained in an insect-free greenhouse under 16:8h light:dark conditions at 25 ±3°C. Three weeks post-inoculation, a systemically infected leaflet (~ 100 mg) was collected from the youngest fully expanded leaf of each inoculated plant, ground in liquid nitrogen, and stored at - 80°C for future RNA extraction.

1.2. RNA extraction, RT-PCR, and qRT-PCR

Total RNA was extracted from frozen tissue using the Qiagen RNeasy Plant Mini Kit (Qiagen, Valencia, CA-USA) following the manufacturer's directions. Total RNA concentration and quality was checked on a NanoDrop 200 Spectrophotometer (Thermo Fisher Scientific, West Palm Beach, FL) and by agarose gel electrophoresis prior to RNA-Seq library construction. Viral RNA in the samples was checked by a multiplexed RT-PCR protocol (Lorenzen et al. 2006b) to determine its strain purity.

The viral transcript level for each Montana PVY^N isolate RNA sample was also determined by a two-step qRT-PCR and normalized to the *EF1α* potato gene (Nicot et al. 2005) using $2^{-\Delta CT}$ values. A specific primer pair designed for the PVY^N genome (qPVY-fwd 5'-ATG GGC TTA TGG TTT GGT GC – 3'; qPVY-rev 5'-TCA ACG ATT GGT TTC AGT GGG-3') and the *EF1α* potato gene primer pair (Nicot et al. 2005) were used in all qRT-PCR reactions for data normalization. Two µg of total nucleic acid from each sample was treated with DNase I to remove any contaminating genomic DNA, and reverse transcribed using AMV First Strand cDNA Synthesis Kit (New

England Biolabs, Ipswich, MA-USA) following the manufacturer's protocol. All qPCR reactions were performed using the Power SYBR Green Master Mix (Life Technologies inc., Carlsband, CA-USA) in a Bio-Rad iCycler thermocycler (Bio-Rad inc., Hercules, CA-USA) using the following amplification program: 95 °C for 3 min, 40 cycles (95 °C for 20 s and 60 °C for 30 s, followed by 72 °C for 15 s). The subsequent program was used for the melt curve data collection and analysis: 95 °C for 1 min, 55 °C for 1 min, and 80 cycles 55 °C for 10 s, the setpoint temperature was increased by 0.5 °C from the beginning of the program. Virus transcript level was not quantified by qRT-PCR for the other nine isolates prior to library construction.

1.3. Library construction and sequencing

Strand-specific RNA-Seq libraries were constructed from 2 µg of total DNase I treated nucleic acid from each sample following a protocol for multiplex strand-specific RNA sequencing (Zhong et al. 2011), with modifications. Barcode-free universal adapters were added during the final PCR enrichment step to increase multiplexing scalability (Zhong et al. 2012). Briefly, oligo(dT)25 Dynabeads (Invitrogen, Carlsbad, CA-USA) was used to isolate poly(A) tailed RNA. First-strand cDNA was synthesized using random hexamer primers and the second-strand cDNA synthesized using dUTP primers. The reads were end-repaired, dA-tailed, and the desired barcode adapters added during the Y-shape adapter ligation step. Triple-SPRI purification and size selection was done using AMPure XP beads (NEB, Ipswich, MA-USA). Finally, PCR enrichment and barcode-free universal adapters were added. The 15 barcoded libraries were pooled in equal molarity and quality checked using a 2100 Bioanalyzer system (Agilent, Santa Clara, CA-USA), and sequenced on an Illumina HiSeq 2000/2500 system with the high-output mode; generating 100-

bp single-end reads, at the Center for Advanced Technology (CAT) Institute of Biotechnology, Cornell University, Ithaca, NY-USA.

1.4. HTS consensus genome reconstruction and phylogenetic analyses

Raw Illumina RNA-Seq reads were first demultiplexed using Illumina's CASAVA pipeline v1.8.2, and then checked for quality with FastQC (Andrews 2010). Filtering and trimming of adaptors and barcode sequences were done using fastq-mcf (Aronesty 2011). Trimmed reads containing any nucleotide with a PHRED quality score <30 and/or reads <20nt long were discarded.

Reference-based mapping for each of the Montana PVY^{N} and PVY^{N} isolates from Europe was performed by aligning trimmed reads against the PVY^{N} Mont reference genome (GenBank: AY884983.1) using Bowtie 2 (Langmead B 2012) with the default settings except "--local" set to "on". PVY^{N} Mont, originally named RRI, was collected in 2001 from Ranger Russet potato grown in the Pacific Northwest of the US (Crosslin et al. 2002). It was sequenced in 2006 and renamed to Mont (Lorenzen et al. 2006a). The following reference PVY strain genomes were used for the reference-based mapping analyses of the other isolates; PVY^{O} O_Z (GenBank: EF026074.1), PVY^{O5} ID-253 (GenBank: HQ912880.1), PVY^{NO} OR-1 (GenBank: DQ157179.1), PVY^{NE11} ID-20 (GenBank: DQ157180.1), PVY^{NTN} PB312 (GenBank: EF026075.1), PVY^{NWi} N1 (GenBank: HQ912863.1). A HTS consensus genome sequence for each individual isolate was derived from the reference-based mapping of the read alignments by using Samtools/BCFtools pipeline (Li et al. 2009). Each HTS consensus sequence was then checked manually for possible indels and nucleotide ambiguity codes. Sanger sequences of the two PVY^{N} isolates from Europe, N605 (GenBank: X97895.1) and NZ (GenBank: AM268435.1), were retrieved from the NCBI database. The complete genome

sequences for isolates NY090031, WI3, MT100010, OR16, WA316, ME4, and MN21 were determined by Sanger sequencing in a separate study (Green et al. 2017).

2.5 Analysis of the 15 PVY isolate populations

Reads from each Montana PVY^{N} isolate were aligned against the PVY^{N} Mont reference genome (Strain alignment) and against the corresponding isolate HTS consensus genome sequence (HTS alignment) using bowtie 2 as described above. For the other isolates, the reads were aligned against the strain reference genome (Strain alignment), the corresponding isolate HTS consensus genome sequence (HTS alignment), and against the Sanger sequence of each isolate (Isolate alignment). The coverage of each genome position for each isolate was derived, based on the alignments of unique RNA-seq reads to each isolate specific HTS consensus sequence, using Bedtools (Quinlan and Hall 2010).

To estimate the relatedness of the Montana PVY^{N} isolates, a phylogenetic analysis was performed on the HTS consensus sequences including those from the PVY^{N} isolates from Europe (N605 and NZ). To strengthen this analysis we added the outgroups, PVY$^{NE\text{-}11}$, isolate ID20, a recombinant PVY^{N} serotype (Karasev et al. 2011) (GenBank: HQ912867.1) and a non-PVY potyvirus, *Papaya leaf distortion mosaic virus* (PLDMV) (GenBank: AB088221.1). The phylogenetic analysis was performed using the CLC Viewer software (CLC Bio, Aarhus, Denmark) and the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and the nucleotide distance measure, Jukes-Cantor; maximum likelihood trees were constructed with 1000 bootstraps.

Single Nucleotide variations (SNVs) in each of the 15 isolate populations were identified using a stringent bioinformatics pipeline (Fig. 1). Sequences potentially derived from the same PCR product were collapsed into one unique read using Picard (Wysoker et al. 2013) to identify

possible sequencing errors introduced during library preparation (reverse transcription (RT) and/or amplification). If the RT and/or PCR steps introduce nucleotide changes in the reads, the Picard algorithm eliminates the potential replication of those changes. Thus, any pseudo-nucleotide change is kept in low frequency and can be removed from the dataset by the SNV identification software threshold. SNV identification was performed using LoFreq (Wilm et al. 2012) with the configuration set to eliminate SNVs with a frequency of <1%, PHRED >30, and identified in <10 unique reads. LoFreq uses Poisson-binomial distribution, taking into account the PHRED quality scores and the sequencing depth at each nucleotide position, to accurately distinguish true SNVs from sequencing artifacts.

SNVs identified in each Montana PVY^{N} isolate were annotated for possible amino acid changes (synonymous/nonsynonymous, start codon gains or losses, and stop codon gains or losses) using SnpEff (Cingolani et al. 2012) and the Mont reference genome sequence for annotations. A principle component analysis was used to estimate Montana PVY^{N} isolate clustering based on SNV distributions, positions, and frequencies using an in-house R script (R Development Core Team 2008). Also in R, pairwise correlation analyses of SNV positions and frequencies between isolates were calculated and plotted for each alignment.

2.6 Visualization of the results

The alignment maps were first visualized in IGV (Thorvaldsdóttir et al. 2013) for overall alignment check (data not shown). All the SNV mapping and graphs from different analyses were created in R using in-house scripts. For the Montana PVY^{N} isolates only, SNV positions and frequencies for each isolate and the average genome coverage for each mapping data were compiled and plotted in circular plots using Circos v 0.67-7 (Krzywinski et al. 2009).

2. Results and discussion

A specific bioinformatics pipeline was developed, *a priori*, and tested for the purpose of this experiment (Fig. 1). Over 197 M 100bp RNA-Seq raw reads were recovered. After the de-multiplexing and filtering processes, about 0.8 % of the reads mapped to the PVY genomes, ranging from 4229 to 266,380 PVY reads per samples (Table 1). All the sequencing metadata were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession numbers SRP059797 and PRJNA287818 for the SRA and BioProjet, respectively.

3.1 Poly(a) tail enrichment for HTS library preparation, a genome-unbiased procedure to deep sequence potyviruses

The presence of a poly(A) tail at the 3' of the PVY genome enabled the use of a strand-specific RNA-Seq library preparation protocol (Zhong et al. 2011) making it possible to multiplex cDNA libraries (Zhong et al. 2012) in Illumina HTS sequencers to reduce sequencing costs. This approach reduces the potential problem of primer/probe binding affinities bias to viral sequences during library preparation and eliminates problems of viral genome specific probes not hybridizing to viral haplotypes that have changes in primer binding sites. Those haplotypes would not be amplified and included in the library, which would interfere with the viral population analyses. Furthermore, because the host transcriptome is also sequenced in the samples, this approach can be particularly useful to analyze both the virus and the host response to the viral infection. Analyses of viral- and host-reads can be performed from the same sequencing data;

thus, potential RNA-seq library preparation errors that can occur during preparation of host and viral specific libraries from the same sample would be minimized.

3.2 RNA-seq reads varied among virus isolates and mapped unevenly throughout the PVY genome

The number of reads obtained from each PVY isolate ranged from 266,380 to 4229 with isolate WA316 having the highest number and isolate NZ having the fewest (Table 1). Although the number of reads from each PVY isolate did not distribute uniformly across the virus genome, there were no sequence gaps in the coding region (Figs 2 and 3) and a complete PVY HTS consensus sequence was successfully compiled for each isolate using the corresponding strain reference genome. On average, coverage was highest within the CP, VPg, and NIb cistrons and lowest in the 6K1 and P3 cistrons (Supplemental table 1).

This uneven genome coverage observed in the PVY isolates genomes may be due to secondary structure present in the virus RNA. This phenomenon is commonly seen in RNA-Seq experiments and it was first mentioned by (Mortazavi et al. 2008). This can result in problems with transcriptome experiments where high genome depth and coverage is necessary for differential gene expression analyses (Sims et al. 2014). Another possible explanation for the uneven distribution is that incomplete viral genomes were sequenced. Although, the RNA-seq library preparation protocol would enrich for RNAs containing a poly(A) tail if there were significant amounts of incomplete viral genomes in the RNA preparations, that would biased the number of reads toward the 3' end of the PVY genome, which it was not observed in our dataset.

3.3 Phylogenetic analyses clustered the Montana PVY^{N} isolates in a single group

Using the HTS consensus sequences; a phylogenetic analysis was performed to estimate the relatedness of the PVY^{N} isolates. The PVY^{N} isolates clustered separately from the $PVY^{NE\text{-}11}$ ID20 isolate and were divided into two distinct clades. The first clade comprised all the Montana PVY^{N} isolates plus the PVY^{N} Mont isolate. The second clade contained both the PVY^{N} isolates from Europe (N605 and NZ) (Fig. 4). This is an indication that PVY^{N} isolates from Europe and North America can be distinguished by using phylogenetic analysis on the consensus sequences generated by HTS data, even though they are all classified as belonging to the same group, European N strain, and considered to be tightly related when phylogenetic analyses are run on sequences of specific regions of the PVY genome (Lorenzen et al. 2006a; Cuevas et al. 2012). Isolate MT130017 was most distantly related to the other Montana PVY^{N} isolates, whereas isolates MT100006 and MT130014 were the most closely related isolates and likely split from a common ancestor. The non-PVY outlier virus sequence (PLDMV), clustered outside of the PVY group.

3.4 PVY^{N} quasispecies were more diverse than isolates from other strains

Using the HTS alignments, the greatest numbers of SNVs were identified in the PVY^{N} isolates (Table 1 and Figs 2 and 3) indicating the PVY^{N} isolates have the most diverse quasispecies among the PVY isolates examined. This plasticity of PVY^{N} genomes may help the virus to adapt to different selection pressures (Vignuzzi et al. 2006), however high diversity could drive the virus toward extinction. The virus could undergo a replicative collapse due to increases in mutation rate that collectively would be detrimental to the virus (Domingo et al. 2012).

In contrast to the PVY^N isolates, the number of SNVs identified in the HTS alignments for all the other isolates was <4 (Table 1 and Figs. 2 and 3), a surprisingly low number given the error rate described for potyviruses (Domingo et al. 2012). It is possible that our stringency for identifying SNVs, i.e. a frequency of > 1%, PHRED >30, and identified in <10 unique reads, underestimated the number of SNVs in regions of the genome that did not have high sequence coverage depth. Nonetheless, the quasispecies characterized for all PVY isolates outside of the PVY^N strain had limited diversity in this study. These isolates do represent different strains and geographic regions, but additional studies on additional isolates from more strains, geographic regions and perhaps host species should be conducted before any general conclusions can be made on the uniformity of PVY quasispecies.

In the Strain alignments, the highest number of SNVs was detected in the WI3 (PVY^O) isolate followed by PVY^N isolates from Europe (NZ and N605) and isolate NY090031 (PVY^O) (Table 1). This is an indication that the strain reference sequences used in the alignment of the reads from each of these isolates were perhaps not the best choices. However, the primary goal of the Strain alignment was to extract each isolate consensus sequence and that was achieved successfully. Moreover, it was noteworthy to discover many more SNVs in the PVY^N isolates from Europe (NZ and N605) than the Montana PVY^N isolates genomes when the reads were aligned to the Mont reference genome. Montana PVY^N and PVY^N from Europe are believed to have common origins (Lorenzen et al. 2006a) and, yet, our results show they have a very different population structure.

3.5. HTS consensus sequences represented virus populations better than Sanger sequences

A higher number of SNVs were identified using the Strain alignment than the HTS alignment for all isolates (Table 1), which was an anticipated result. The HTS consensus sequence was extracted from the sample reads and it likely to be a better representation of each virus population than the strain sequence available in the NCBI database. Similar conclusions can be drawn from the detection of more SNVs in the Isolate alignment than from HTS alignment for each isolate. The isolate sequence was acquired by Sanger sequencing approach – only one sequence was generated per sample. The HTS consensus sequences were acquired from an average of 75,484 80nt-long-reads per sample, coverage of approximately 622X per nucleotide position. Therefore, the HTS consensus sequence was expected to be more representative of the viral population - that is, having less SNVs when reads are aligned to them than when reads are aligned to Sanger sequences (either the strain or the specific isolate sequence).

Unexpectedly, more SNVs were detected in the Isolate than Strain alignment for the recombinant isolates: WA316 (PVY^{NE11}), ME4 (PVY^{NTN}), and MN21 (PVY^{NWi}) (Table 1). The Isolate alignment was acquired by aligning the reads to the isolate specific Sanger sequence. The strain reference sequence alignment is when reads were aligned to the strain Sanger genome sequence acquired from the NCBI database. Because the NCBI strain reference sequence was from the strain but not from the specific isolate of PVY used in this study, more SNVs were expected to be detected from the strain than from the Isolate alignment. Perhaps the two types of sequences, NCBI and isolate, were sequenced from different hosts, which could interfere with the population structure of the virus, resulting in differences in the Sanger sequences. The isolate sequences were obtained from PVY-infected tobacco(Green et al. 2017), but the host from which the NCBI reference sequences were obtained was unclear – no such information

could be retrieved from the NCBI database. Studies designed to investigate the interference of the host on the PVY consensus sequence acquired by Sanger approach should be conducted to clarify our speculation.

3.6. SNVs did not disrupt the cistron functions of Montana PVY^N isolates

A greater number of synonymous than nonsynonymous SNVs were detected in both the Strain and HTS alignments of the Montana PVY^N isolates with the exception of the MT100017 isolate HTS alignment, which had more nonsynonymous than synonymous SNVs (Fig. 5A and B). The SnpEff algorithm, used for the annotation analyses of the Montana PVY^N isolates, did not detect any start or stop codon gains or losses, indicating that the cistrons sequenced were functional. Furthermore, the well characterized motifs in all potyviruses, e.g. KITC and PTK in the HC-PRO coding region for aphid transmission, AVGSGKST in the CI for helicase function, GDD in NIb for replicase activity, and DAG in the CP coding region for aphid transmission (Urcuqui-Inchima et al. 2001), were not altered by any of the SNVs detected (Supplemental figure 1).

3.7. Montana PVY^N isolates have diverged from the ancestral Mont

More SNVs were detected in the Strain alignments than in the HTS alignments throughout the genome of each Montana PVY^N isolate (Fig. 6A and B). Most of the SNVs detected in the Strain alignment had 100% frequency (100% of reads aligned to that site had that specific SNV), indicating that these SNVs have become fixed or are close to fixation in the population (Fig. 6A). A majority of the SNVs were detected in the P1, CI, VPg, NIa, and NIb cistrons (Fig. 6A). Only two of the 100% frequency SNVs, one located in the VPg and the other in the NIb cistron, were found in all isolates. This is evidence that the Montana PVY^N has changed considerably over the

years in comparison to the Mont reference isolate. Isolated in 2001, Mont was the first PVY^{N} isolate reported from potatoes in the US (Crosslin et al. 2002), and we believe it is reasonable to consider Mont to be the ancestor of the recently collected Montana PVY^{N} isolates. We considered the Strain alignment of the Mont genome to be the ground zero for understanding the dynamics of Montana PVY^{N} populations. The Strain alignment was a crucial tool for comparisons of population structures among Montana PVY^{N} isolates.

3.8. A large number of SNVs were identified in the CP cistron in the quasispecies of Montana PVY^{N} isolates

All SNVs detected in HTS alignments had a frequency below 50% and were also found in Strain alignments (Fig. 6A and B). Overall, across the HTS alignments of Montana PVY^{N} isolates, the highest concentration of SNVs was detected in the CP cistron (Figs. 2 and 5B), which suggests that these regions are more tolerant to changes. However, this tolerance seems to be mainly for synonymous rather than nonsynonymous substitutions as more synonymous than nonsynonymous substitutions were observed throughout the PVY genome of all Montana PVY^{N} isolates with the exception of MT1000017, which had more nonsynonymous than synonymous substitutions (Fig. 5B). Similar results were obtained by (Moreno et al. 2004). When studying genetics of a population of *Watermelon mosaic virus* (WMV), those authors found CP and CI to be more constrained evolutionary than P3 and therefore more prone to be shaped by negative selection. On the other hand, the cistrons, P3, 6K1, and 6K2 seem to be less tolerant to changes as little or no SNVs were detected in these regions in the Montana PVY^{N} isolate genomes in both the Strain alignment (Fig. 6A) and HTS alignment (Figs. 2 and 6B).

The Montana PVY^{N} isolates, MT100006 and MT130017, had the highest number of SNVs in the CP cistron in HTS alignment (Figs. 2 and 6B) and lowest number of reads recovered (Table 1). This suggests that, possibly, those changes in the CP cistron interfere with replication resulting in low genome copies. This hypothesis is supported by qRT-PCR results that revealed isolates MT100006 and MT130017 had lower sequence copies in comparison with the other Montana PVY^{N} isolates (Table 1). CP is involved with aphid transmission efficiency and movement from cell-to-cell in the plant (Quenouille et al. 2013) and despite most of the changes detected being synonymous; still, they have a fitness cost to viruses (Carrasco et al. 2007; Sanjuán et al. 2004).

3.9. SNVs may be linked to viral transmission mode

The fact that only a few SNVs, from both alignments (Strain and HTS), were common to all Montana PVY^{N} isolates and that some SNVs were common among some isolates but not to others (Fig. 6A and B); suggest that these isolates may have gone through different selection pressures. We speculate that transmission mode plays an important role in this non-uniform distribution of SNVs between isolates. Perhaps, the genetic diversity structure of PVY populations changes according to the transmission mode. If this is the case, isolates that are transmitted by aphids may share SNVs while those isolates transmitted through infected tubers share other unique SNVs. This is a real scenario in potato fields where PVY infection starts primarily through infected tubers and it is spread in the field via aphid vectors.

3.10. MT130017, the most unrelated isolate among Montana PVY^{N} isolates

Principle component analysis (PCA) of the Montana PVY^N isolates compared the distribution, position, and frequency of SNVs identified in each isolate population using the HTS and strain alignment for each isolate (PVY^N Mont) (Fig. 7A and B). Using the SNVs identified from the strain alignments (here, because reads from each isolate were aligned to the same reference genome, the population structure among isolates could be compared), six PVY^N isolates were separated into three distinct groups. Isolate MT130017 stands alone, isolates MT130039 and MT100006 grouped together, as did isolates MT100017, MT130014, and MT130054 (Fig. 7A). This gives some insights on the closeness of the isolates within each group; it may be that the isolates within each group went through similar selection pressures. On the other hand, PCA of SNVs identified in HTS alignments did not resolve any groupings of the six Montana PVY^N isolates (Fig. 7B), which is reasonable. The HTS alignment was designed to identify SNVs particular to each isolate population by aligning reads to each isolate-specific consensus sequence. Therefore, unless SNVs are produced in a non-random manner, we would not expect to find the same SNVs in each isolate population in this type of alignment.

The pairwise correlation analyses of SNV positions and frequencies among the Montana PVY^N isolates reveal a weak correlation between isolate MT130017 and the other Montana PVY^N isolates using both Strain and HTS alignments (Fig. 8A and B). This agrees with the results of the PCA and phylogenetic analyses of HTS consensus sequences of the Montana PVY^N isolates; isolate MT130017 is the most unrelated isolate among all Montana PVY^N isolates, an indication that this isolate shares the least amount of SNVs among all Montana PVY^N isolates. This is supported by the fact that of all the SNVs that have become fixed (100% frequency) in the majority of the six Montana PVY^N isolates, only three SNVs are shared with MT130017 (Fig. 6A).

3. Conclusions

PVY^N has almost completely disappeared from US potato fields, and it has been found occasionally only on a limited number of farms in Montana. HTS sequence data from all strains indicated a higher level of genetic diversity in populations of PVY^N than other, more common strains. This suggests that the high genetic diversity found in the N isolate populations, very evident in the MT130017 isolate, might cause a fitness disadvantage. This could help to explain the shift in prevalence of PVY strains in potato fields, where recombinant strains are becoming the predominant strains in many potato production areas worldwide (Karasev and Gray 2013). PVY^N has lost the battle for dominance against those recombinant PVY strains in US potato fields.

Altogether, the results of this study are important to understand the evolutionary dynamics of PVY, and what happens to the virus population structure over years in a single geographical location. The discovery that PVY^N has a more diverse quasispecies than any other PVY strain sequenced in this study sheds lights on of a peculiar phenomenon, the shift in prevalence of PVY strains in US potato fields. Furthermore, the recognition that the NIb, CP, VPg, and CI proteins are under tight functional constraint has the potential to assist the development of new strategies for controlling virus infection in potato plants. Mutagenic compounds analogous to ribavirin, may be developed that target constrained proteins cistrons and increase the mutation rate in the virus population. This could initiate a replicative collapse and restrict PVY infection. In fact, Ribavirin combined with electrotherapy (Mahmoud et al. 2009) and/or meristem culture (AlMaarri et al. 2012) are effective means to efficiently generate PVY-free potato plantlets. These approaches are particularly beneficial for the seed potato industry as methods

to produce PVY-free potato nuclear seed stocks that meet strict PVY infection tolerances imposed by seed certification programs.

ACKNOWLEDGEMENTS

The authors would like to thank: Jason Ingram for helping with inoculations and isolates caracterization procedures; Jim Giovannoni for helping with the RNA-seq library preparation process; and Zhangjun Fei, Suzy Strickler, Noe Fernandez, and Denis Kutnjak for assisting with the bioinformatics analyses. This study was funded in part by USDA-SCRI (2009-51181-05894 and 2014-51181-22373), the UK Biotechnology and Biological Sciences Research Council grant # BB/L011840/1 as part of the joint USDA-NSF-NIH-BBSRC Ecology and Evolution of Infectious Diseases program. WD was partially supported by a SUNY Diversity Fellowship.

LITERATURE CITED

Adams, M.J., F.M. Zerbini, R. French, F. Rabenstein, D.C. Stenger *et al.*, 2011 Family *Potyviridae*, pp. 1069–1089 in *Virus taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses*, edited by A. King, M. Adams, E. Carstens and E. Lefkowitz. Elsevier, Oxford.

Ali, M.C., T. Maoka, K.T. Natsuaki, and T. Natsuaki, 2010 The simultaneous differentiation of *Potato virus Y* strains including the newly described strain PVY^{NTN-NW} by multiplex PCR assay. *Journal of Virological Methods* 165 (1):15-20.

AlMaarri, K., R. Massa, and F. AlBiski, 2012 Evaluation of some therapies and meristem culture to eliminate Potato Y potyvirus from infected potato plants. *Plant Biotechnology* 29 (3):237-243.

Andrews, S., 2010 FastQC: A Quality Control tool for High Throughput Sequence Data *Availabel online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*.

Aronesty, E., 2011 Command-line tools for processing biological sequencing data. *ea-utils*.

Barker, H., 1994 Incidence of potato virus Y infection in seed and ware tubers of the potato cv. Record. *Annals of Applied Biology* 124 (1):179-183.

Betz-Stablein, B.D., A. Töpfer, M. Littlejohn, L. Yuen, D. Colledge *et al.*, 2016 Single-molecule sequencing reveals complex genome variation of hepatitis B virus during 15 years of chronic infection following liver transplantation. *Journal of Virology* 90 (16):7171-7183.

Carrasco, P., F. de la Iglesia, and S.F. Elena, 2007 Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in *Tobacco Etch Virus*. *Journal of Virology* 81 (23):12979-12984.

Chung, B.Y.-W., W.A. Miller, J.F. Atkins, and A.E. Firth, 2008 An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences* 105 (15):5897-5902.

Cingolani, P., A. Platts, L.L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3. *Fly* 6 (2):80-92.

Crosslin, J.M., P.B. Hamm, K.C. Eastwell, R.E. Thornton, C.R. Brown *et al.*, 2002 First report of the necrotic strain of *Potato virus Y* (PVY^N) on potatoes in the Northwestern United States. *Plant Disease* 86 (10):1177-1177.

Cuevas, J.M., A. Delaunay, J.C. Visser, D.U. Bellstedt, E. Jacquot *et al.*, 2012 Phylogeography and molecular evolution of *Potato virus Y*. *PLoS ONE* 7 (5):e37853.

Cuevas, J.M., A. Willemsen, J. Hillung, M.P. Zwart, and S.F. Elena, 2015 Temporal dynamics of intra-host molecular evolution for a plant RNA virus. *Molecular Biology and Evolution*.

Domingo, E., J. Sheldon, and C. Perales, 2012 Viral quasispecies evolution. *Microbiology and Molecular Biology Reviews* 76 (2):159-216.

Eigen, M., and C. Biebricher, 1988 RNA Genetics, Vol. III, Variability of RNA Genomes, pp. 211-245 in *Eds E. Domingo, JJ Holland and P. Ahlquist. CRC Press Inc., Florida*.

García-Arenal, F., A. Fraile, and J.M. Malpica, 2001 Variability and genetic structure of plant virus populations. *Annual Review of Phytopathology* 39 (1):157-186.

Gray, S., S. De Boer, J. Lorenzen, A. Karasev, J. Whitworth *et al.*, 2010 *Potato virus Y:* An Evolving Concern for Potato Crops in the United States and Canada. *Plant Disease* 94 (12):1384-1397.

Green, K.J., C.J. Brown, S.M. Gray, and A.V. Karasev, 2017 Phylogenetic study of recombinant strains of Potato virus Y. *Virology* 507:40-52.

Gutiérrez, S., Y. Michalakis, and S. Blanc, 2012 Virus population bottlenecks during within-host progression and host-to-host transmission. *Current Opinion in Virology* 2 (5):546-555.

Hu, X., A. Karasev, C. Brown, and J. Lorenzen, 2009 Sequence characteristics of *Potato virus Y* recombinants. *Journal of General Virology* 90:3033-3041.

Jakab, G., E. Droz, G. Brigneti, D. Baulcombe, and P. Malnoe, 1997 Infectious *in vivo* and *in vitro* transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. *Journal of General Virology* 78 (12):3141-3145.

Karasev, A., and S. Gray, 2013 Continuous and Emerging Challenges of *Potato virus Y* in Potato. *Annual Review of Phytopathology* 51:571-586.

Karasev, A.V., X. Hu, C.J. Brown, C. Kerlan, O.V. Nikolaeva *et al.*, 2011 Genetic Diversity of the Ordinary Strain of Potato virus Y (PVY) and Origin of Recombinant PVY Strains. *Phytopathology* 101 (7):778-785.

Karasev, A.V., O.V. Nikolaeva, X. Hu, Z. Sielaff, J. Whitworth *et al.*, 2010 Serological Properties of Ordinary and Necrotic Isolates of *Potato virus Y*: A Case Study of PVY^{N} Misidentification. *American Journal of Potato Research* 87 (1):1-9.

Kehoe, M.A., B.A. Coutts, B.J. Buirchell, and R.A.C. Jones, 2014 Plant virology and next generation sequencing: Experiences with a *Potyvirus*. *PLoS ONE* 9 (8):e104580.

Kerlan, C., and B. Moury, 2008 *Potato virus Y*, pp. 287-296 in *Encyclopedia of Virology (Third Edition)*, edited by B.W.J. Mahy and M.H.V.v. Regenmortel. Academic Press, Oxford.

Krzywinski, M.I., J.E. Schein, I. Birol, J. Connors, R. Gascoyne *et al.*, 2009 Circos: An information aesthetic for comparative genomics. *Genome Research*.

Kutnjak, Denis, M. Rupar, I. Gutierrez-Aguirre, T. Curk, J.F. Kreuze *et al.*, 2015 Deep sequencing of virus derived small interfering RNAs and RNA from viral particles shows highly similar mutational landscape of a plant virus population. *Journal of Virology*.

Langmead B, S.S.L., 2012 Fast gapped-read alignment with Bowtie 2. *Nature methods* 9 (4):357-359.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25 (16):2078-2079.

Lorenzen, J., T. Meacham, P.H. Berger, P.J. Shiel, J.M. Crosslin *et al.*, 2006a Whole genome characterization of *Potato virus Y* isolates collected in the western USA and their comparison to isolates from Europe and Canada. *Archives of Virology* 151 (6):1055-1074.

Lorenzen, J., L.M. Piche, N.C. Gudmestad, T. Meacham, and P. Shiel, 2006b A Multiplex PCR Assay to Characterize *Potato virus Y* Isolates and Identify Strain Mixtures. *Plant Disease* 90 (7):935-940.

Mahmoud, S.Y.M., M.H. Hosseny, and M.H. Abdel-Ghaffar, 2009 Evaluation of Some Therapies to Eliminate Potato Y Potyvirus from Potato Plants. *International Journal of Virology* 5 ((2)):64-76.

Martínez, F., G. Lafforgue, M.J. Morelli, F. González-Candelas, N.-H. Chua *et al.*, 2012 Ultradeep sequencing analysis of population dynamics of virus escape mutants in RNAi-mediated resistant plants. *Molecular Biology and Evolution* 29 (11):3297-3307.

Mondal, S., and S.M. Gray, 2017 Sequential acquisition of Potato virus Y strains by Myzus persicae favors the transmission of the emerging recombinant strains. *Virus Research* 241 (Supplement C):116-124.

Moreno, I.M., J.M. Malpica, J.A. Díaz-Pendón, E. Moriones, A. Fraile *et al.*, 2004 Variability and genetic structure of the population of watermelon mosaic virus infecting melon in Spain. *Virology* 318 (1):451-460.

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold, 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* 5 (7):621-628.

Nicot, N., J.-F. Hausman, L. Hoffmann, and D. Evers, 2005 Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany* 56 (421):2907-2914.

Ogawa, T., Y. Tomitaka, A. Nakagawa, and K. Ohshima, 2008 Genetic structure of a population of *Potato virus Y* inducing potato tuber necrotic ringspot disease in Japan; comparison with North American and European populations. *Virus Research* 131 (2):199-212.

Perales, C., V. Martín, C.M. Ruiz-Jarabo, and E. Domingo, 2005 Monitoring sequence space as a test for the target of selection in viruses. *Journal of Molecular Biology* 345 (3):451-459.

Quenouille, J., N. Vassilakos, and B. Moury, 2013 *Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity. *Molecular Plant Pathology* 14 (5):439-452.

Quinlan, A.R., and I.M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26 (6):841-842.

R Development Core Team, R., 2008 R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-00, URL http://www.R-project.org*.

Sanjuán, R., A. Moya, and S.F. Elena, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proceedings of the National Academy of Sciences of the United States of America* 101 (22):8396-8401.

Schubert, J., V. Fomitcheva, and J. Sztangret-Wiśniewska, 2007 Differentiation of *Potato virus Y* strains using improved sets of diagnostic PCR-primers. *Journal of Virological Methods* 140 (1–2):66-74.

Simmons, H.E., J.P. Dunham, J.C. Stack, B.J. Dickins, I. Pagan *et al.*, 2012 Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. *J Gen Virol* 93:1831-1840.

Sims, D., I. Sudbery, N.E. Ilott, A. Heger, and C.P. Ponting, 2014 Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 15 (2):121-132.

Smith, K.M., 1931 On the Composite Nature of Certain Potato Virus Diseases of the Mosaic Group as Revealed by the use of Plant Indicators and Selective Methods of Transmission. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* 109 (762):251-267.

Thorvaldsdóttir, H., J.T. Robinson, and J.P. Mesirov, 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* 14 (2):178-192.

Urcuqui-Inchima, S., A.-L. Haenni, and F. Bernardi, 2001 Potyvirus proteins: a wealth of functions. *Virus Research* 74 (1–2):157-175.

Vignuzzi, M., J.K. Stone, J.J. Arnold, C.E. Cameron, and R. Andino, 2006 Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439 (7074):344-348.

Wilm, A., P.P.K. Aw, D. Bertrand, G.H.T. Yeo, S.H. Ong *et al.*, 2012 LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Research* 40 (22):11189-11201.

Wysoker, A., K. Tibbetts, and T. Fennell, 2013 Picard tools version 1.90. *http://picard.sourceforge.net*.

Zhong, S., Z. Fei, and J.J. Giovannoni, 2012 Modifications for performing multiplex strand-specific RNA sequencing using Illumina HiSeq. *Cold Spring Harb Protoc*.

Zhong, S., J.G. Joung, Y. Zheng, Y.R. Chen, B. Liu *et al.*, 2011 High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* (8):940-949.

Zwart, M.P., A. Willemsen, J.-A. Daròs, and S.F. Elena, 2014 Experimental Evolution of Pseudogenization and Gene Loss in a Plant RNA Virus. *Molecular Biology and Evolution* 31 (1):121-134.

Figure Legends

Figure 1: Scheme of the bioinformatics pipeline developed for this study.

Figure 2: Distribution and frequency of reads and SNVs throughout the genome of different PVY^{N} isolates Illumina deep-sequenced in this study. Reads were aligned against each corresponding isolate HTS consensus sequence. SNVs are represented by the red histogram bars with frequency scale 0 (none of the reads aligned to that region had the SNV) to 100 (100% of reads aligned to that region had the SNV). The grey curve represents the genome coverage.

Figure 3: Distribution and frequency of reads and SNVs throughout the genome of different PVY strains Illumina deep-sequenced in this study. Reads were aligned against each corresponding isolate HTS consensus sequence. SNVs are represented by the red histogram bars with frequency scale 0 (none of the reads aligned to that region had the SNV) to 100 (100% of reads aligned to that region had the SNV). The grey curve represents the genome coverage.

Figure 4: Phylogenetic analysis performed on the HTS consensus sequence from each PVY^{N} isolate, the Mont reference genome (GenBank: AY884983.1), and two genome sequences outliers - PVY^{NE-11} strain isolate ID20 (GenBank: HQ912867.1) and the papaya leaf distortion mosaic virus (PLDMV) (GenBank: AB088221.1), using the CLC Viewer software. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and maximum likelihood trees were constructed with 1000 bootstraps.

Figure 5: SNVs annotation for amino acid changes (synonymous/nonsynonymous) for each Montana isolate from reads aligned to the Mont reference genome (GenBank: AY884983.1) (A) and to the isolate specific HTS consensus sequence (B) as calculated by SnpEff algorithm using Mont reference genome (GenBank: AY884983.1) for based annotation reference.

Figure 6: Circular representation of PVY^N genome with plotted frequencies of SNVs from the different Montana isolates. Reads were aligned against the Mont reference genome (Strain alignment) (GenBank: AY884983.1) (A) and against the corresponding isolate HTS consensus genome sequence (HTS alignment) (B). The SNVs are represented by the red histogram bars with frequency scale 0 (none of the reads aligned to that region had the SNV) to 1 (100% of reads aligned to that region had the SNV). The grey area in the center represents the average genome coverage of all samples.

Figure 7: PCA of SNV distributions, positions, and frequencies between Montana isolates from reads aligned to the Mont reference genome (GenBank: AY884983.1) (A) and to the isolate specific HTS consensus sequence (B).

Figure 8: Pairwise correlation analyses of SNVs frequencies between Montana isolates from reads aligned to the Mont reference genome (GenBank: AY884983.1) (A) and to the isolate specific consensus sequence (B). Darker color indicates higher correlation of isolate pairs.

Supplemental figure 1: Snapshot of some of the translational changes induced by the SNVs found in each Montana isolate. In highlight, the well characterized motifs in the potyvirus genome (KITC, PTK, AVGSGKST, GDD, and DAG) were not altered by the detected SNVs. The Mont reference genome (GenBank: AY884983.1) was used for annotation reference.

Tables

Table 1: Number of reads aligned and SNVs detected in each PVY^N Montana isolate genome after adapters removal and filtering process. Strain align = reads aligned to the strain reference genome sequence. Isolate align = reads aligned to the isolate specific Sanger genome sequence. HTS cns align = reads aligned to the isolate specific HTS consensus sequence derived from this deep sequencing study. The qRT-PCR results, which estimate the virus titer, show the Ct value and the fold change for each isolate in the sample after normalization with the values of the potato *EF1α* reference gene used in the reactions. The fold change was calculated using the formula $2^{-\Delta\Delta Ct}$. - = data not collected.

Isolate	Strain	Reads Count	SNV Count				qRT-PCR Results	
			Strain align	High frequency*	Isolate align	HTS cns align	Delta Delta Ct	Fold Change
MT100006	N	5594	39	13	-	26	14.1	1.7
MT100017	N	136274	33	22	-	11	18.2	29.4
MT130014	N	34494	32	24	-	8	17	13.41
MT130017	N	5479	55	6	-	49	15.6	5.1
MT130039	N	79743	26	17	-	9	18.4	34.6
MT130054	N	234186	33	24	-	9	18.5	36.2
N605	N	90239	121	99	63	22	-	-
NZ	N	4229	131	75	123	56	-	-
NY090031	O	72658	106	106	2	0	-	-
WI3	O	132854	155	154	5	1	-	-
MT100010	O^5	52829	16	16	7	0	-	-
OR16	NO	101470	4	2	4	1	-	-
WA316	NE11	266380	9	8	12	1	-	-
ME4	NTN	11355	19	16	30	3	-	-
MN21	NWii	102882	6	6	20	0	-	-

* SNVs detected from the strain alignment that were more the 50% frequent - more than 50% of reads aligned to that region had the SNV.

Supplemental table 1: Average of reads aligned to each cistron of the PVY isolates genomes deep sequenced in this study.

Isolate	Strain	Cistrons and UTRs											
		5' UTR	**P1**	**HC-PRO**	**P3**	**6K1**	**CI**	**6K2**	**Vpg**	**Nia**	**Nib**	**CP**	**3' UTR**
MT100006	N	0	73	67	32	11	43	47	82	31	71	100	32
MT100017	N	78	1431	1432	736	380	1090	1511	2106	887	1886	2144	925
MT130014	N	7	397	359	198	86	233	436	427	223	441	702	312
MT130017	N	13	25	22	47	8	37	30	48	56	41	68	456
MT130039	N	38	809	1074	461	249	649	1130	998	583	995	1089	474
MT130054	N	58	2792	2744	1471	718	1738	2753	3566	1660	3041	3662	1176
N605	N	40	901	1108	493	158	755	654	1282	540	1234	1228	751
NZ	N	1	40	57	20	6	35	24	53	26	54	66	39
NY090031	O	12	873	446	550	126	511	477	1038	770	985	1730	311
WI3	O	20	1309	896	910	333	979	722	1812	1558	1847	2938	441
MT100010	O^5	22	586	365	378	143	427	279	743	595	726	1048	302
OR16	NO	19	1029	956	782	507	729	733	1279	962	1162	2525	408
WA316	NE11	260	3261	2727	1324	1231	2227	2604	3356	2140	3282	5421	1781
ME4	NTN	2	150	102	83	21	97	97	145	78	147	226	46
MN21	NWii	22	720	1183	606	440	624	751	1501	817	1240	2827	671

All Isolates Mean	40	960	903	539	294	678	817	1229	728	1143	1718	542
Montana N Mean	32	921	950	491	242	632	985	1204	573	1079	1294	562
All N Isolates Mean	29	809	858	432	202	572	823	1070	501	970	1133	521

Figures

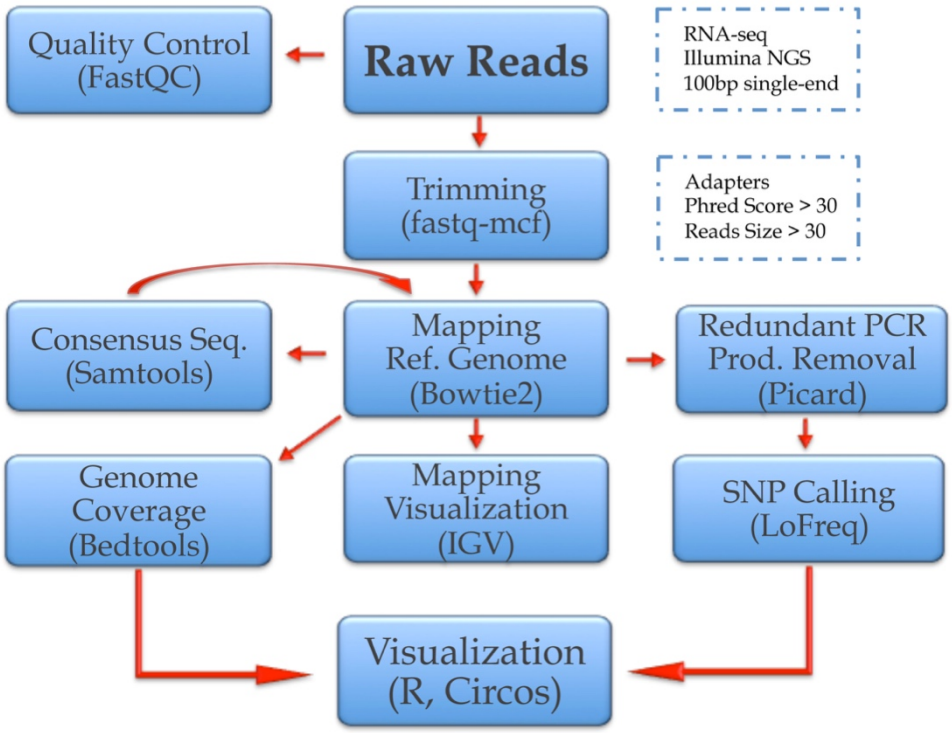


Figure 1

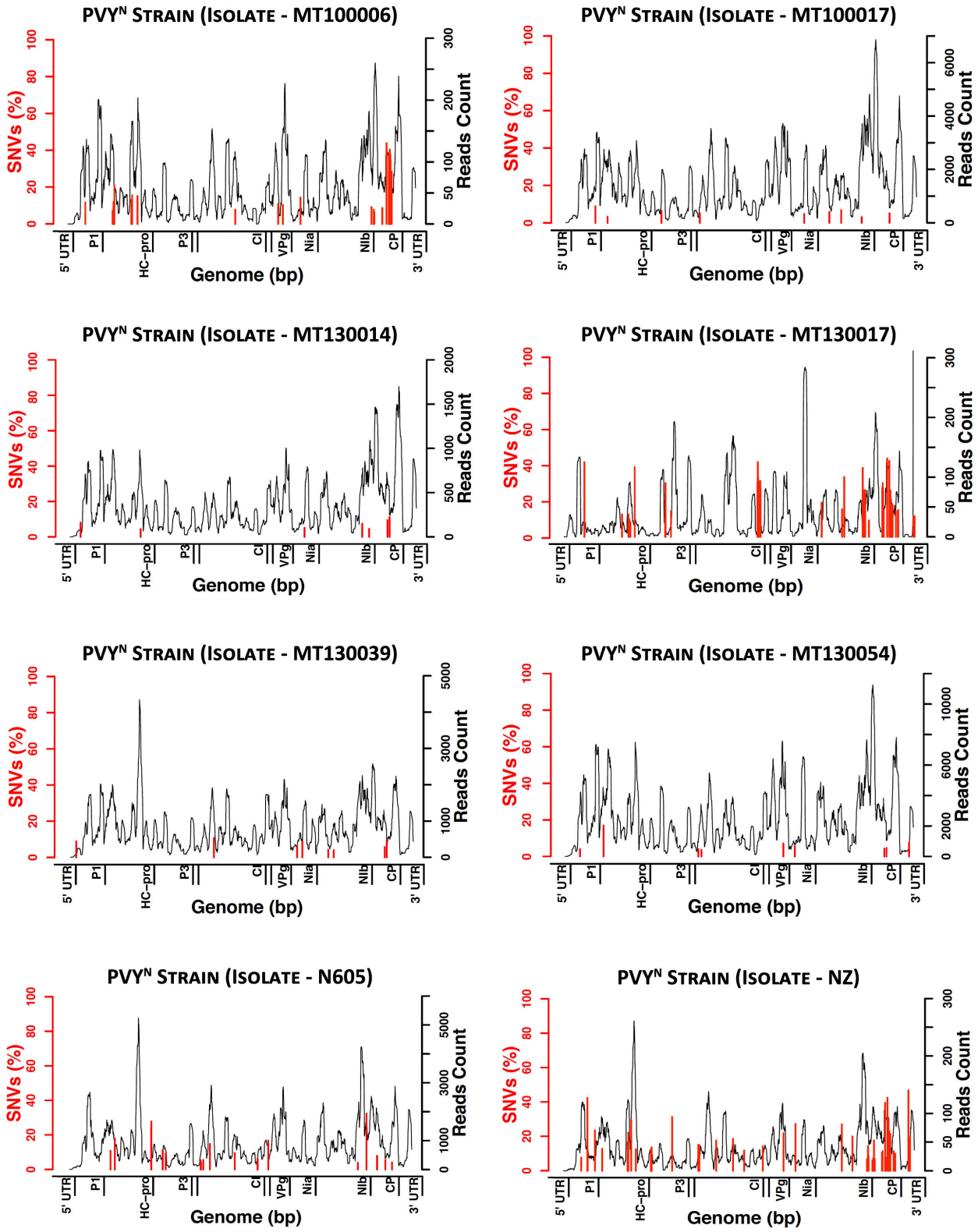

PVY^N STRAIN (ISOLATE - MT100006)
PVY^N STRAIN (ISOLATE - MT100017)
PVY^N STRAIN (ISOLATE - MT130014)
PVY^N STRAIN (ISOLATE - MT130017)
PVY^N STRAIN (ISOLATE - MT130039)
PVY^N STRAIN (ISOLATE - MT130054)
PVY^N STRAIN (ISOLATE - N605)
PVY^N STRAIN (ISOLATE - NZ)
SNVs (%)
Reads Count
Genome (bp)
5' UTR
P1
HC-pro
P3
CI
VPg
Nia
NIb
CP
3' UTR

Figure 2

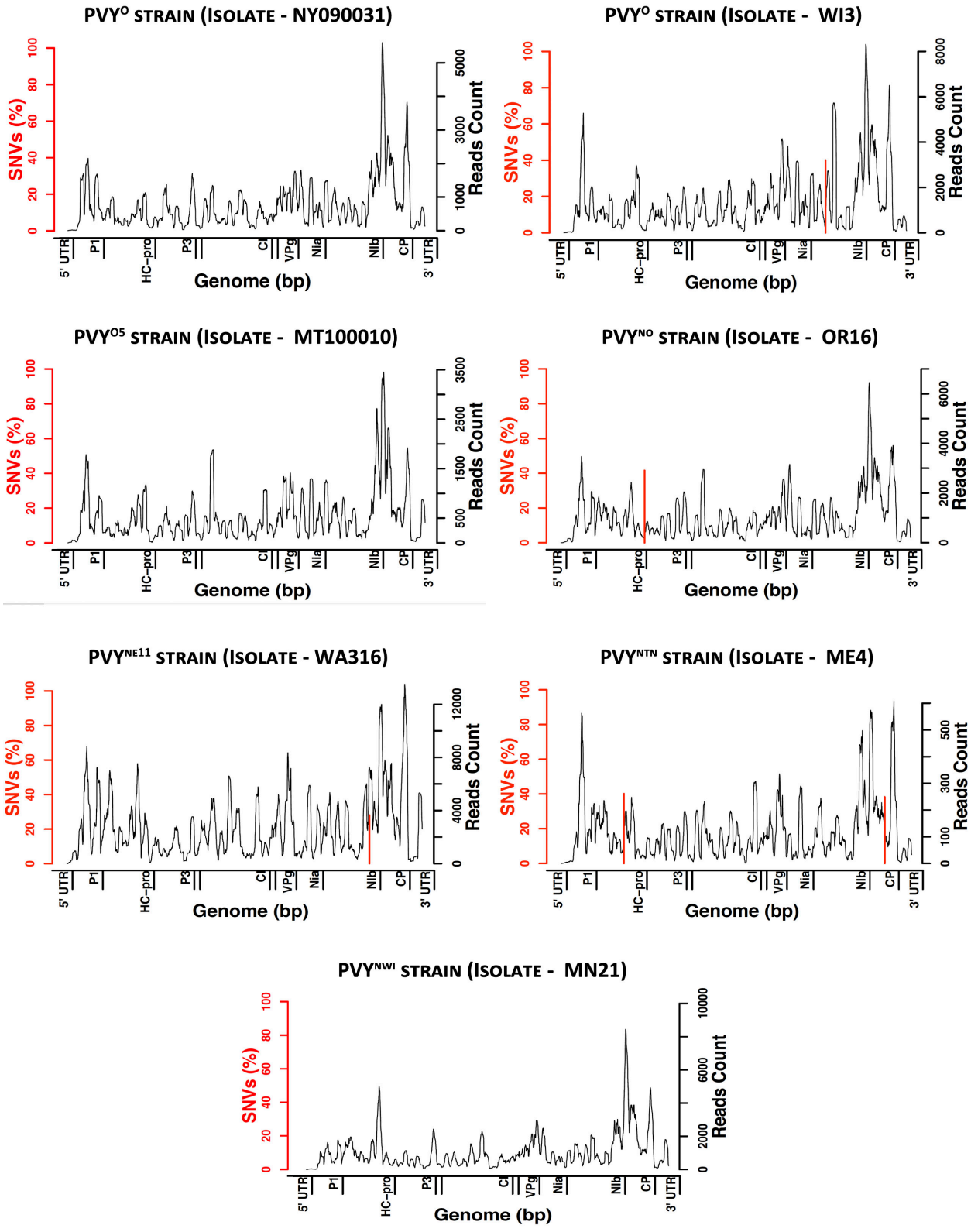

PVY^{O} STRAIN (ISOLATE - NY090031)
PVY^{O} STRAIN (ISOLATE - WI3)
PVY^{O5} STRAIN (ISOLATE - MT100010)
PVY^{NO} STRAIN (ISOLATE - OR16)
PVY^{NE11} STRAIN (ISOLATE - WA316)
PVY^{NTN} STRAIN (ISOLATE - ME4)
PVY^{NWI} STRAIN (ISOLATE - MN21)
SNVs (%)
Reads Count
Genome (bp)
5' UTR
P1
HC–pro
P3
CI
VPg
Nia
NIb
CP
3' UTR

Figure 3

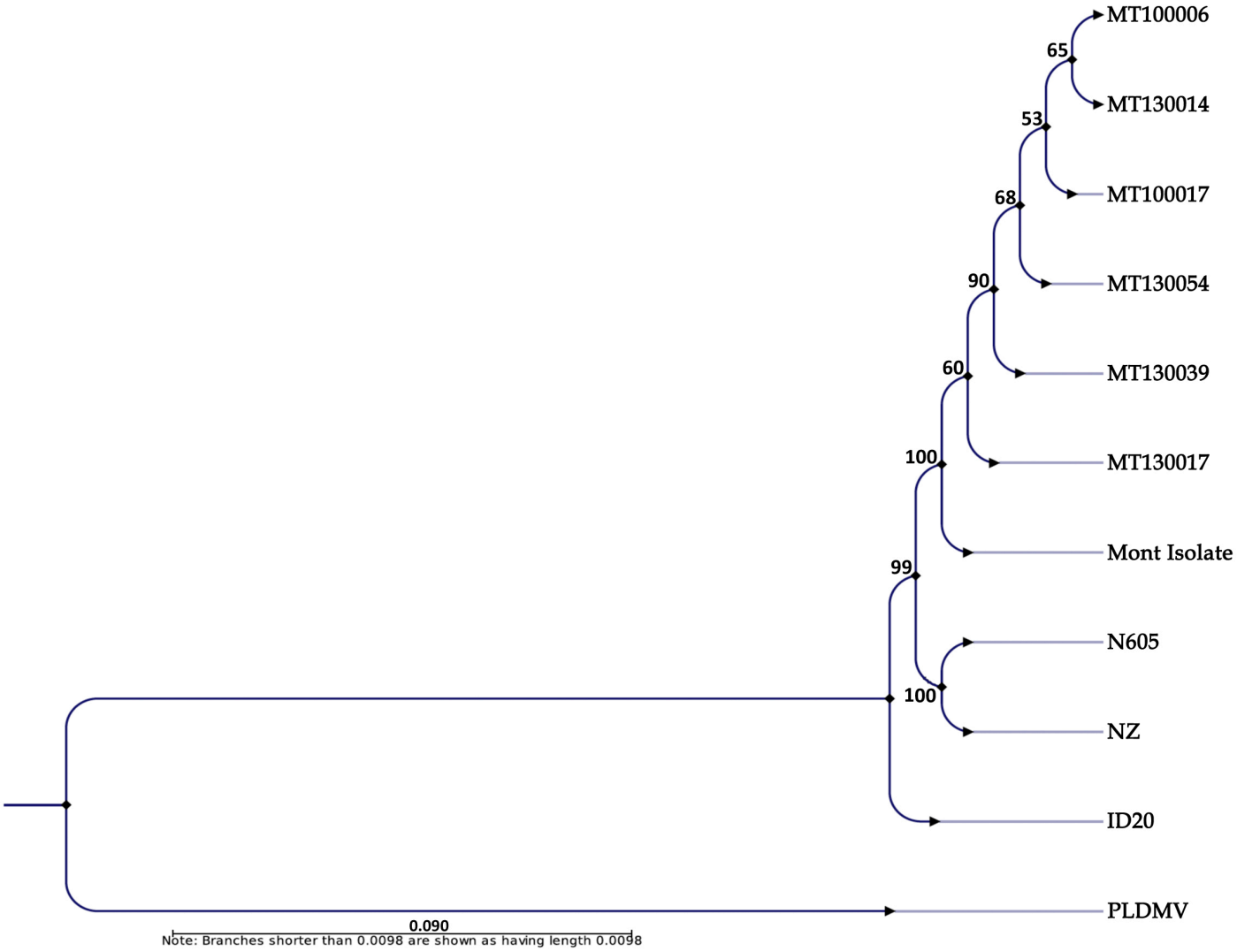


Figure 4

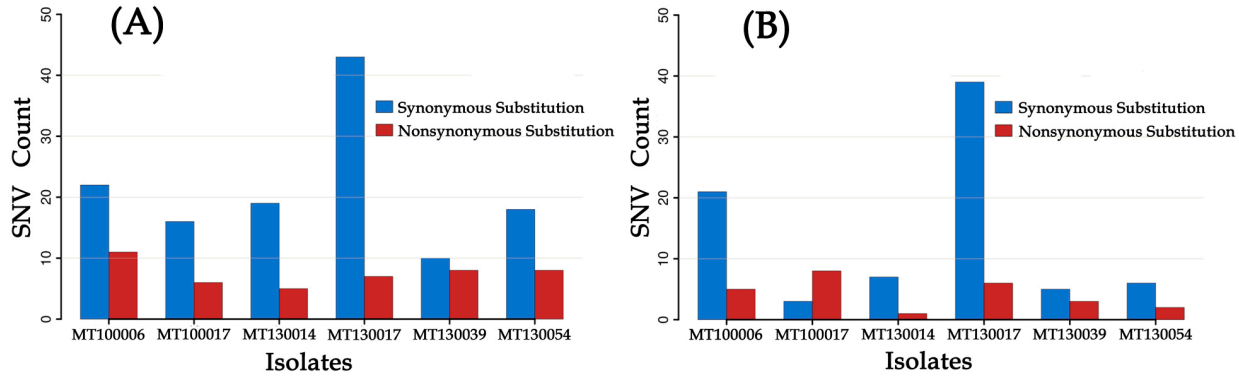


Figure 5

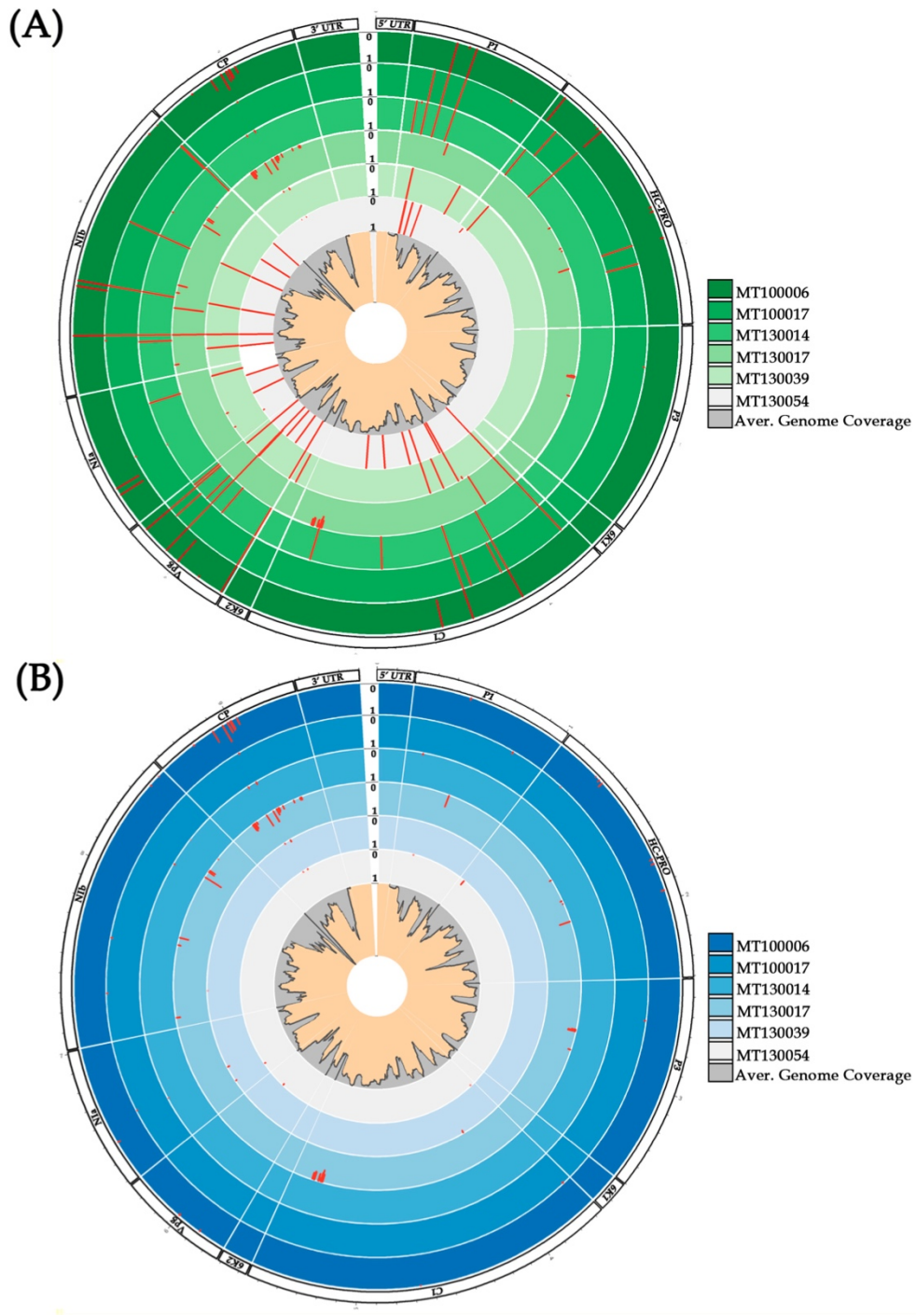

(A)
5' UTR
P1
HC-Pro
P3
6K1
CI
6K2
VPg
NIa
NIb
CP
3' UTR
MT100006
MT100017
MT130014
MT130017
MT130039
MT130054
Aver. Genome Coverage
(B)
5' UTR
P1
HC-Pro
P3
6K1
CI
6K2
VPg
NIa
NIb
CP
3' UTR
MT100006
MT100017
MT130014
MT130017
MT130039
MT130054
Aver. Genome Coverage

Figure 6

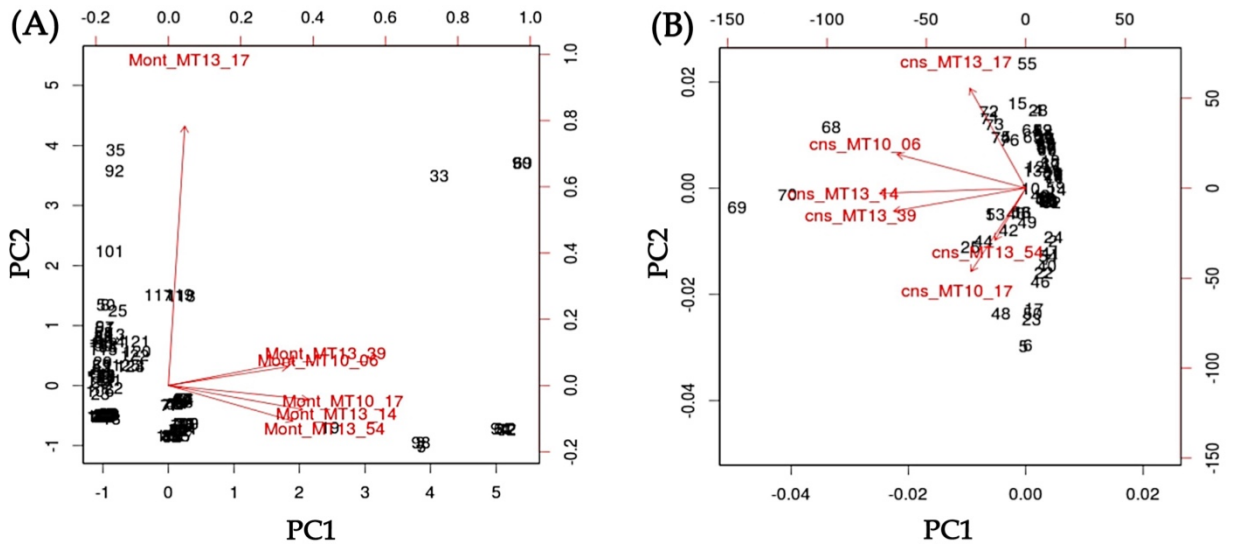


Figure 7

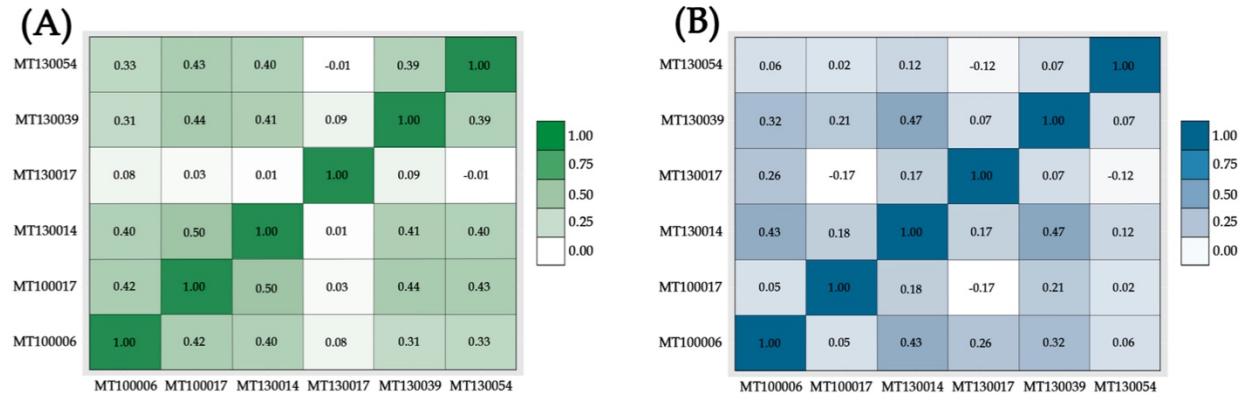
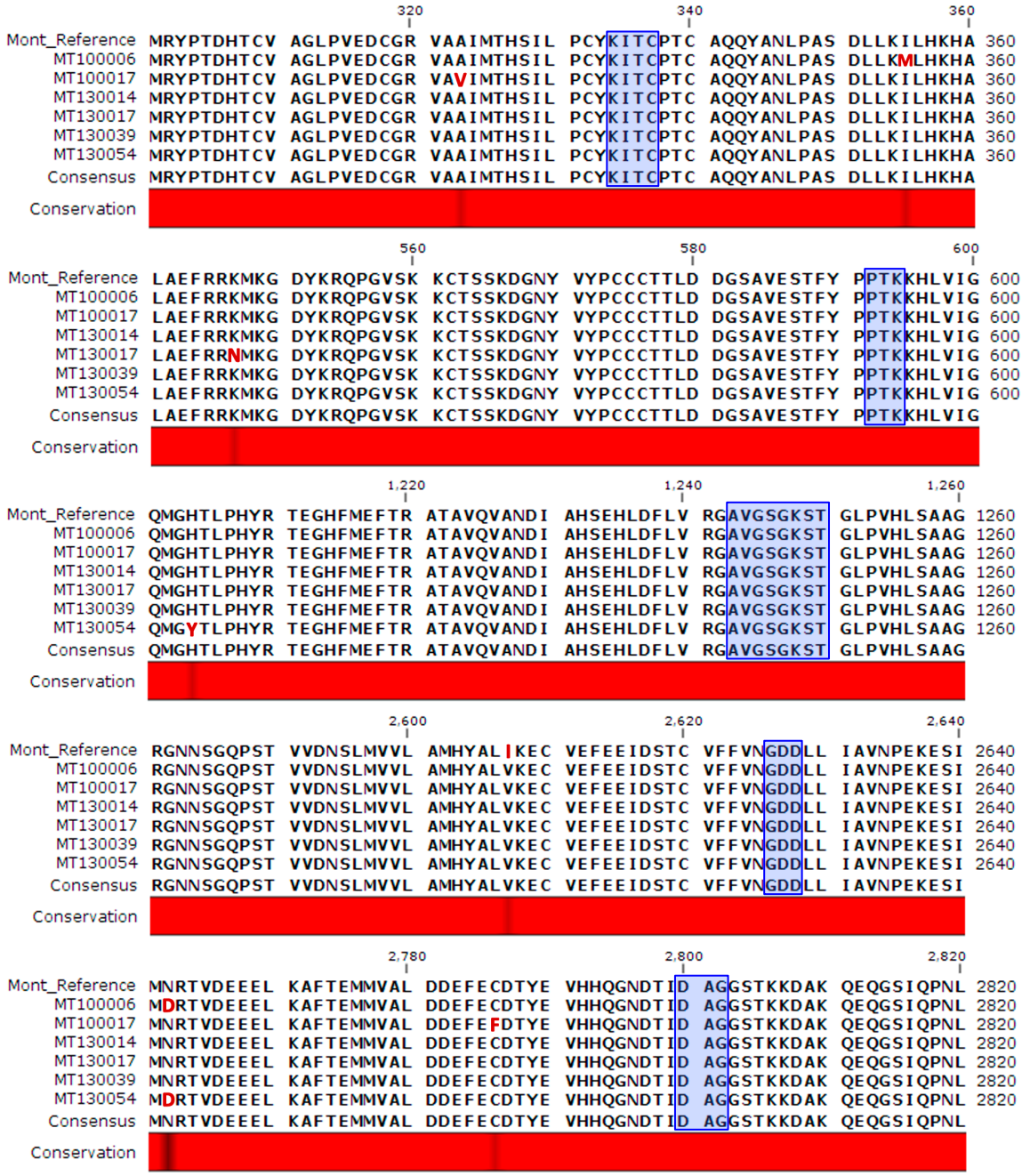


Figure 8



Supplemental figure 1

CHAPTER 5

Transmission mode as a driver of genetic diversity decline in potato virus Y quasispecies

Abstract

Surveys indicate that recombinant strains of *Potato virus Y* (PVY) have emerged in recent years to predominate in the U.S. potato crop, and that the genetic diversity abounds among and within PVY strains. To investigate how transmission modes (insect, vegetative propagation through tubers, and mechanical) are contributing to the evolution and diversification of PVY, we designed experiments to characterize the quasispecies of two PVY strains, PVY^{O} isolate WI3 and PVY^{NWi} isolate MN21, prior to and following passage of the virus by each transmission mode. Population diversity of PVY varied by the virus strain but did not differ among transmission modes. Most of the single the nucleotide variations (SNVs) detected were present at low frequencies (<10%) across all passages in both strains populations. Many SNVs detected in the source population carried through the passages in each transmission mode and virus strain. However, none of those SNVs reached 50% in frequency in WI3 populations. Four SNVs in MN21 populations reached over 50% in frequency, SNVs 386 and 5566 were uniquely found in aphid transmitted samples, 5187 was unique to mechanical inoculated samples, and SNV 5697 was unique to infected tuber samples. SNVs 5187 and 5697 were located on CI and VPg cistrons, respectively. CI is known to be involved in aphid transmission and cell-to-cell movement and VPg affects cell-to-cell and systemic movement of potyviruses. Overall, these data suggest that each transmission mode exerts unique selection pressures on the virus population and allows different mutations to accumulate and become fixed.

INTRODUCTION

The high replication rate of RNA viruses coupled with the poor proofreading ability of its RNA-dependent-RNA-polymerase often leads to a genetically diverse virus population in an infected

host (Holmes 2009). This expectation has spawned the "quasispecies" concept that the individuals in a virus population are not just a collection of random mutants, but an interactive group of variants (Eigen and Biebricher 1988). High heterogeneity within a quasispecies may allow the virus to quickly evolve and adapt to new and/or changing environments, i.e. achieving the so-called host adaptation (Vignuzzi, Stone et al. 2006). Two evolutionary processes, selection and genetic drift, are the major forces shaping virus quasispecies (García-Arenal, Fraile et al. 2001). Selection is a directional process that results in an increase in the frequency of the fittest variants frequency in the populations as the environment becomes more suitable. Genetic drift is driven by random changes in populations often the result of bottleneck events occurring during the virus life-cycle such as transmission between hosts and movement within the host (Zwart and Elena 2015). It has often been difficult to distinguish the effects of selection and genetic drift in plant virus evolution studies since the same forces (e.g. vector transmission and within-host movement) have been linked to both evolutionary processes (García-Arenal, Fraile et al. 2001). Furthermore, there have been technological challenges in measuring the complete genetic variation in a virus population (Zwart and Elena 2015). The advent of high-throughput sequencing (HTS) has allowed detection of viral mutations even in low frequencies (Beerenwinkel and Zagordi 2011). Coupled with novel and evolving bioinformatics algorithms, HTS has improved our ability to the study of the processes that shape virus quasispecies with an unprecedented level of genetic resolution (Acevedo and Andino 2014, Posada-Cespedes, Seifert et al. 2017).

The high population diversity found in a RNA virus quasispecies does not necessary reflect fitness gain. No mutations identified in a tobacco etch virus quasispecies infecting *Nicotiana tabacum* were found to be beneficial: cost of the mutations was lethal (41%) or deleterious (36%) and the remaining 23% were neutral, fitness was reduced an average of 41% (Carrasco, de la Iglesia et al. 2007). These results are consistent with studies on other RNA viruses (Elena, Fraile et al. 2014). Furthermore, synonymous mutations in RNA viruses are not always neutral. For instance, synonymous mutations in the PB2 polymerase gene of influenza A have strong impact on viral packing (Marsh, Rabadán et al. 2008), synonymous codons introduced in poliovirus caused attenuation (Coleman, Papamichail et al. 2008), and introducing synonymous changes

that yield differences in overall nucleotide composition or changes in GC content have been shown to play important roles in host - jump and host adaptation of different animal viruses (Holmes 2009). Additionally, the increasing discovery of functional RNA secondary structure within viral gene coding regions, strongly argue against synonymous mutation neutrality in virus populations (Simmonds and Smith 1999, Simmonds, Tuplin et al. 2004, McMullan, Grakoui et al. 2007). Hence, there is a need for a better understanding of viral population structure dynamics and of the forces that shape virus quasispecies to advance our knowledge on virus evolution.

Potato virus Y (PVY), the type species of the genus *Potyvirus* in the family *Potyviridae*, is a major virus disease of potatoes (*Solanum tuberosum* subsp. *tuberosum*) that exists as a complex of strains (Karasev, Hu et al. 2011). Genomic comparison studies have shown that the PVY^{O} and PVY^{N} strains have served as parents to numerous emerging recombinant strains, several of which have become prevalent in potato fields worldwide (Karasev and Gray 2013). PVY can be transmitted horizontally between plants mechanically and by aphids, and vertically through infected tubers (mother to daughter). The potyvirus genome contains one open reading frame that encodes a large polyprotein subsequently cleaved into ten putative proteins by three viral-encoded proteases (P1, HC-Pro and Nla) (Urcuqui-Inchima, Haenni et al. 2001), and an small out of frame ORF (PIPO) that is embedded in the P3-cistron and encodes the P3N-PIPO protein (Chung, Miller et al. 2008). Key motifs in the potyvirus polyprotein are linked to virus transmission by aphids and to virus movement within the host (Urcuqui-Inchima, Haenni et al. 2001). The DAG motif located near the surface-exposed N-terminus of the capsid protein interacts with the PTK motif on the C-terminus of the Helper component protease (HC-Pro). The KITC motif on the N-terminus of the HC-Pro interacts with binding domains on the aphid mouthparts allowing the HC-pro to act as a bridge connecting the virus particle to the aphid and facilitating transmission. The P3N-PIPO protein is associated with cell-to-cell viral movement by forming a protein complex with viral cytoplasmic inclusions (CI), capsid protein (CP), and host protein PA-CaP1, moving virions through plasmodesmata (Vijayapalani, Maeshima et al. 2012) and the HC-Pro facilitates the traffic of this virus-protein complex by altering the exclusion size limit of the plasmodesmata (Rojas, Zerbini et al. 1997). Furthermore, mutations introduced on VPg cistron of turnip mosaic virus significantly restricted cell-to-cell and systemic movement of the virus

(Dunoyer, Thomas et al. 2004), depicting an new function of VPg and the impact that mutations on those key cistrons may have on potyvirus movement in plants. All these features combined with the high diversity found in RNA viruses quasispecies, make PVY an ideal system to test important research questions e.g. “Does transmission mode shape virus quasispecies?”. Horizontal and vertical transmission may act on the PVY quasispecies overtime selecting mutations that give the virus a fitness advantage in each transmission mode.

To study how transmission modes (insect, vegetative propagation through tubers, and mechanical) may contribute to the evolution and diversification of PVY, experiments were designed to characterize the quasispecies (virus population diversity within a plant) of different PVY strains prior to and following passage of the virus by each transmission mode. We aimed to investigate in different plant tissues (leaf vs tuber) 1) if PVY quasispecies changed according to the transmission mode and 2) if specific mutations were selected in each transmission mode overtime. We used Illumina HTS technology and the latest bioinformatics algorithms to address those points. We initiated our experiment with a natural virus population that allowed us to study the dynamics of variations that were found in the source population and the new variations that rose in each passage overtime in each transmission mode.

MATERIAL AND METHODS

Experimental design. A single potato plant, the founding plant, was mechanically inoculated with each of the two virus isolates used in the study. These founding plants were then used as the source to mechanically inoculate three potato plants, each considered as a biological replication. These three plants were used as virus source for subsequent passages using each transmission mode (aphid transmission - AT, mechanical inoculation - MI, and infected tuber - IT). The virus quasispecies was characterized using HTS in each of the founding plants and a consensus sequence was determined and used to compare all of the quasispecies for that virus isolate from all subsequent passages. The founding virus quasispecies and consensus sequence was also compared with a full-length virus sequence determined previously by Sanger sequencing.

Each virus isolate was passed five times using mechanical and aphid transmission. A young, recently mature leaf was collected three weeks post inoculation (wpi) from each plant and an infected tuber was collected from each plant at harvest (four months post inoculation - 4 mpi). Five vertical passages of virus through tubers was planned, but a lack of viable tuber production after the second generation limited the number of passages to two (Figure 1).

Plant and PVY Strains Selection. Potato cv 'Goldrush', at five-to-six-leaf stage, were used in the experiments. All potato plants, 'Goldrush' used in this study originated from disease-free minitubers kindly provide by Chris Nobles, Cornell-Uihlein Potato Farm. Plants from mini tubers were cultivated in 10 gallons plastic pots, tested free from PVY using a PVY-specific triple-antibody sandwich ELISA (TAS-ELISA) (Agdia, Elkhart, IN), and used a source of cuttings to provide plants for the experiments. Cuttings taken from the mother plants were dipped in Hormex rooting hormone #1 (Brooker Chem. Corp., Chatsworth, CA), and planted into 96 well trays containing Cornell soil mix (Boodley and Sheldrake 1982). Three weeks later, well rooted plants were individually transplanted to four-liter plastic pots containing Cornell soil mix, grown for two weeks, then inoculated. All plants (healthy and infected plants) were maintained in an insect-free greenhouse under 16:8h light:dark conditions at 25 ±3^{o}C.

An ongoing national survey of PVY isolates affecting the U.S. seed potato crop has yielded a lab collection of thousands of isolates representing several PVY strains. A preliminary study was conducted to evaluate multiple PVY isolate quasispecies to aid the selection of strains and isolates for this study (Chapter 4). We were interested in strains and isolates that have a large and diverse quasispecies (high titer and many variations within the virus population). Based on the results from the preliminary study, we selected PVY^{O} isolate WI3 and PVY$^{N\text{-}Wi}$ isolate MN21. PVY$^{N\text{-}Wi}$ is a recombinant strain now prevalent in the US (Karasev and Gray 2013).

The isolates, originally collected from potato were maintained long term in lyophilized tobacco tissue stored at -40^{o}C. The strain identity of the isolates was confirmed using two strain-specific RT-PCR multiplex diagnostic assays (Lorenzen, Piche et al. 2006, Ali, Maoka et al. 2010). Lyophilized tissue was homogenized in 10 volumes of PBS (137 mM NaCl, 2.7 mM KCl, 10 mM

Na_2HPO_4, 2 mM KH_2PO_4, and pH adjusted to 7.4 with HCl) with a mortar and pestle. Then used to mechanically inoculate the founding potato plant by gently rubbing the viral inoculum solution with a cotton swab to the youngest fully expanded leaf using carborundum (600 mesh) as abrasive. Single plants were individually inoculated with both virus isolates described above. Three weeks post-inoculation, the youngest fully expanded leaf of each inoculated plant (TAS -ELISA tested to be infected) was collected and the terminal leaflet (~ 100 mg) used for total RNA extraction. The remainder of the leaf was used to mechanically inoculate three potato seedlings that served as the viral source plants for the subsequent passages. Three weeks after the inoculation, the youngest fully expanded leaf of each source plant was collected and the terminal leaflet (~ 100 mg) was sampled for total RNA extraction. Half of the remaining leaf was used to start the aphid transmission assay and the other half to start the mechanical inoculation assay. Plants were grown to maturity and the largest infected tuber harvested from each plant were collected and half of it used as the source of RNA to characterize the virus quasispecies and the other half was used to start the infected tuber transmission assay. A preliminary study found that large tubers from PVY infected plants had a higher PVY titer than small tubers and the virus was more concentrated in the distal end (the end far from the connection to the mother plant) of those tubers (data not shown). Individual leaf and tuber samples (~ 100 mg) collected for RNA extraction were ground in liquid nitrogen and stored in -80^{o}C for total RNA extraction.

Aphid transmission assay. Green peach aphids (*Myzus persicae*), reared on turnip (*Brassica rapa*), were used for all aphid transmission assays. Aphids were starved for two hours then allowed a 5-min acquisition access period on the PVY-infected source leaves (see description above). Ten aphids were moved to a clip cage enclosing a leaflet on each of the recipient plants and allowed a 24 hr inoculation access period. Aphids were killed by fogging plants with a mixture of pymetrozine, Endeavor 50WG, at 1.6 g $(93\ m^2)^{-1}$ and bifenthrin, Talstar P, at 15 ml $(93\ m^2)^{-1}$ and plants moved to a greenhouse. Three wpi, plants were tested using TAS -ELISA to determine infection status and if positive the youngest fully expanded leaf on three infected plants was collected. The terminal leaflet from the leaf was used for total RNA extraction and

the remainder of the leaf was used as virus source for aphid transmission for the next virus passage. For all passages, infected plants were allowed to grow to maturity (four months) and a sample from the distal end of the largest infected tuber was collected for total RNA extraction.

Mechanical inoculation assay. Half of the source leaf (see description above) was homogenized in 10 volumes of PBS and used to mechanically inoculate potato seedlings. Three wpi, plants were tested by TAS-ELISA to determine infection status and the youngest fully expanded leaf was collected from three plants. The terminal leaflet from the leaf was collected for total RNA extraction and the remainder of the leaf was used as the source for mechanical inoculation of the next virus passage. In all passages, infected plants were allowed to grow to maturity and the largest tuber was selected for RNA extraction as described above.

Infected tuber assay. The three original source plants inoculated from the founder plant were grown to maturity and the largest infected tuber from each plant was selected. The distal half of the tuber was cut longitudinally; one section was used for total RNA extraction and the other section was soaked for one hour in 2ppm gibberellic acid solution to help break dormancy (Alexopoulos, Aivalakis et al. 2008) and then used to produce the next plant generation. When the tuber began to sprout it was transplanted to a four-liter plastic pot containing Cornell soil mix. Emerging leaves were tested by TAS-ELISA to determine infection status and if positive all but the most vigorous sprout were removed and the plant left grown to maturity. Three weeks after the emergency of the sprouts, the terminal leaflet from the youngest fully expanded leaf was collected for total RNA extraction. At harvest, the largest infected tuber was collected and used for the next virus passage.

RNA extraction and viral nucleic acid enrichment. Total RNA was extracted from frozen grounded issue using the PureLink™ Plant RNA Reagent (Thermo Fisher Scientific, Waltham, MA-USA) following the manufacturer's directions. Total RNA concentration and quality was checked on a NanoDrop 200 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA-USA) and by agarose gel electrophoresis prior to PCR amplification. Several approaches were compared for library construction. Initially, we used a RNA-seq protocol (Zhong, Joung et al. 2011)

taking advantage of the poly A tail on the PVY genome. Over 36,000 (~ 0.8% of the total reads) PVY reads were obtained from the leaf samples and the average numbers of PVY reads from tuber tissue was less than 1,000. Subsequently, we attempted to enhance for PVY-specific reads in tuber tissue by using PVY-specific oligos attached to biotin molecules followed by capture with streptavidin-magnetic-beads to capture the biotin-oligo complexes from the total RNA samples. This enriched for PVY sequences from leaf samples (5% vs the 0.8%) but there was no significant enrichment from tuber samples. Next, we designed PVY oligos for each strain to cover 100% of the PVY ORF and allow overlapping regions of at least 76bp between amplicons (Supplemental Table 1). Briefly, first-strand cDNA was synthesized from the extracted RNA using a mix of random hexamers with anchored-dT primers following the protocol provided by the supplier using a ProtoScript® II First Strand cDNA Synthesis Kit (NEB, Ipswich, MA-USA). cDNA was then amplified (20 cycles), using the specifically designed PVY primer sets for each strain, via PCR using Phusion Green Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA-USA), following the manufacturer's protocols. Amplicons from each tissue sample were pooled, cleaned using ChargeSwitch PCR Clean-Up Kit (Thermo Fisher Scientific, Waltham, MA-USA), and quantified using a NanoDrop 200 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA-USA). Amplicons sizes were checked with gel electrophoresis. This approach yielded 95% enrichment of PVY reads per sample (~4,329,165) regardless the potato tissue type, leaf or tuber.

Library construction and sequencing. DNA samples were sheared by sonication using a Covaris S2 Focused-ultrasonicator (Covaris, Woburn, MA-USA), 130µL microTUBE AFA Fiber Pre-Slit Snap-Cap 6x16mm (25) (cat# 520045), and the microTUBE Holder (cat# 500114). Fifty µL of sample was sheared using the following parameters: Intensity of 5, duty cycle of 10%, 200 cycles per burst, temperature of 7^{o}C, and treatment time of 3 min. DNA fragment length distribution was assessed using a 2100 Bioanalyzer (Agilent, Santa Clara, CA-USA) with a target peak size ranging from 150 to 500 bp. Libraries were constructed from 1-2 µg of DNA following a published protocol (Zhong, Joung et al. 2011), with modifications. Fragmented DNA was purified and size selected using AMPure XP beads (NEB, Ipswich, MA-USA) before end-repair, dA-

tailing, and the universal TruSeq adapter ligation step. Size selection was done again using AM-Pure XP beads to filter out extra adapters and small size fragments. Finally, each library was PCR enriched 6-8 cycles with six-base single indexed primers. The barcoded libraries were pooled in equal molarity and quality checked using a 2100 Bioanalyzer (Agilent, Santa Clara, CA-USA), and sequenced on an Illumina HiSeq 2000/2500 system with the high-output mode; generating 100-bp single-end reads, at the Center for Advanced Technology (CAT) Institute of Biotechnology, Cornell University, Ithaca, NY-USA.

HTS data pre-processing and consensus genome reconstruction. Raw Illumina RNA-Seq reads were first demultiplexed using Illumina's CASAVA pipeline v1.8.2, and then checked for quality with FastQC (Andrews 2010). Filtering and trimming of adaptors and barcode sequences were done using fastq-mcf (Aronesty 2011). Trimmed reads containing any nucleotide with a PHRED quality score <30 and/or reads <70nt long were discarded.

Reference-based mapping for each of the strains was performed by aligning trimmed reads from the founding population (PVY reads from the founding plant) against the strain reference genome using BWA (Li and Durbin 2009) using the MEM alignment algorithm mode. Reference genome sequences used were: for PVY^{O}, isolate WI3 (GenBank: KY848031.1), and for $PVY^{N\text{-}Wi}$, isolate N1 (GenBank: HQ912863.1). A HTS consensus genome sequence for each individual strain was derived from the reference-based mapping of the read alignments by using Samtools/BCFtools pipeline (Li, Handsaker et al. 2009). Reads from the founding population were realigned to the HTS consensus sequence multiple times until all the high frequent changes (> 50%) in the population were incorporated to the HTS consensus sequences. Each HTS consensus sequence was then checked manually for possible indels and nucleotide ambiguity codes and used for downstream analysis.

Analysis of the PVY strain quasispecies. Reads from each sample were aligned against the corresponding strain HTS consensus genome sequence using BWA as described above. The coverage of each genome position for each sample was derived, based on the alignments of unique

RNA-seq reads to each strain specific HTS consensus sequence, using Bedtools genomecov (Quinlan and Hall 2010).

Single nucleotide variations (SNVs), variation in a single nucleotide without any limitations of frequency, in each sample population were identified using a stringent bioinformatics pipeline adapted from GATK Best Practices (Van der Auwera, Carneiro et al. 2013). PCR products duplicate were removed using Picard MarkDuplicates (Wysoker, Tibbetts et al. 2013) and base (Quality Score) recalibration (using BaseRecalibrator, Apply Recalibration, and AnalyzeCovariates tools) was conducted to detect and correct systematic errors in the base quality scores. Then, SNV calls were performed using LoFreq (Wilm, Aw et al. 2012) with the configuration set to eliminate SNVs with PHRED >30 and identified in <20 unique reads. Alignment maps were first visualized in IGV (Thorvaldsdóttir, Robinson et al. 2013) for overall alignment check. SNVs detected in each virus population were annotated for possible amino acid changes (synonymous/nonsynonymous, start codon gains or losses, and stop codon gains or losses) using SnpEff (Cingolani, Platts et al. 2012). All SNV maps and graphs from different analyses were created in R 3.4.3 (Team 2018) using in-house scripts.

Nucleotide diversity indices (π). For each PVY strain, the SNV dataset from each virus population was entered in SNPGenie software (Nelson, Moncla et al. 2015) to calculate nucleotide diversity indices (π) using the founding population consensus sequence as reference. We divided the dataset into SNVs present in the source population (standing SNVs) and those that appeared in subsequent populations (new SNVs) and calculated nucleotide diversities for all SNVs and for each divided dataset as performed in (Kutnjak, Elena et al. 2017). Results were plotted in R (Team 2018) using ggplot2 package 2.2.1 (Wickham 2009). Also in R, linear mixed-effect models analyses using the package lme4 1.1-15 (Bates, Mächler et al. 2015) and pairwise comparisons using emmeans package 1.1 (Lenth 2018) were conducted to test the significance at 95% of the difference between means. Normality was checked by plotting the histogram of the residues from each model. The π values were the response variables and passage and treatment were the explanatory variables.

Analyses of SNVs trajectories. To evaluate the fate of the SNVs overtime, a trajectory line connecting SNV frequencies in each passage was drawn for each SNV in R (Team 2018) using ggplot2 package 2.2.1 (Wickham 2009). Because most of the SNVs weren't present in our three bio-replications, we set up a criterion where for the SNV to be plotted it had to be present in at least two of the bio-replications, SNVs that didn't meet this criterion weren't plotted in the graphs. And we took an average of the frequencies of the SNVs from each bio-replication to plot the SNVs frequencies in the graphs. Also in R, we used the Kolmogorov-Smirnov D test (Lehmann 2006) to assess the similarity of virus populations based on the standing SNVs (maintained from the source population and carried through the passages). We were specifically interested in investigating if the selected SNVs and their frequencies were statistical different between transmission modes in each sample location (leaf and tuber) among passages.

RESULTS

Vertical and horizontal passaging of PVY strains in potato. To help establish our sampling and analysis strategy, a pilot experiment was conducted to evaluate the quasispecies populations of the WI3 isolate of PVY^{O} strain in a 'Goldrush' potato plant that developed over time and in different tissues. A mechanically infected potato seedling was grown to maturity. The terminal leaflet of the youngest fully expanded leaf was collected at 3 wpi weeks post inoculation. Four additional leaf samples were collected just prior to harvest (~4 months post inoculation); the terminal leaflet of the youngest fully expanded leaf and terminal leaflets from three evenly spaced leaf positions moving down the plant. Additionally, a tuber flesh sample was collected from the largest infected tuber harvested and from a sprout developing from that tuber after it had broken dormancy (Figure S1). RNA was extracted and the library was constructed using the RNA-seq protocol as described in the Materials and Methods. The number of reads per sample averaged 36,026 and 1,000 for leaf and tuber samples, respectively. Although the genome coverage was robust, many regions were described by less than 20 reads per site. This was especially true for the tuber sample and the low number of reads did not allow for the confident detection of SNVs in those regions as our stringent SNV calling pipeline requires at least 20 reads to be aligned per site (see material and methods). Nonetheless, the population structure was

not significantly different among the leaf and sprout samples, but the tuber sample did contain a very different population structure. A single nucleotide variation (SNV) was detected in all leaf samples and sprout, but not in the tuber sample. This single SNV was also detected in the WI3 isolate in separate experiment designed to compare the nucleotide diversity across a number of PVY isolates representing various strains (Chapter 3). A cluster of five SNVs were identified on the CP cistron and a single SNV on the NIb cistron in the tuber sample, none of which were detected in the leaf or sprout samples (Figure S1).

Based on the data from the pilot studies, two isolates of PVY representing the PVY^{O} and PVY^{NWi} strains (isolate WI3 and isolate MN21, respectively) were selected for use in the horizontal (mechanical and aphid transmission) and vertical (through tubers) passage experiments (Figure 1). Due to the differences observed in nucleotide diversity between the tuber sample and leaf/sprout samples in the pilot study samples were collected from tubers and leaves in each passage. Leaf samples were collected from each plant at 3 wpi and tuber samples were collected at harvest. Both leaf and tuber samples from each plant served as a source of RNA to characterize the virus quasispecies and leaf samples served as a virus source for the subsequent passage in mechanical inoculation (MI) and aphid transmission (AT) transmission modes. In infected tuber (IT) transmission, leaf and tuber samples were collected simultaneously at harvest and tuber samples served as a virus source for the subsequent passage. A total of 122 leaf and tuber samples were analyzed. Five passages were completed in the MI and AT experiments, although an outbreak of powdery mildew in the greenhouse prematurely killed the fifth passage plants prior to collecting tuber samples. Due to low tuber production and tuber vigor only two IT passages were completed. Cost constraints limited analysis of samples to passages one, two, three, and five.

PVY specific oligos enriched libraries for PVY sequences and increased the depth of HTS coverage. On average, 4,329,165 (~95% of the total reads per sample) 100bp long raw HTS reads representing 100% of the PVY genome were generated from each leaf and tuber sample using the PVY sequence enrichment protocol. This was a significant improvement over the other two methods, a RNA-seq protocol (Zhong, Joung et al. 2011), which yielded an average of 36,026

and ~ 1,000 raw PVY reads per leaf and tuber sample, respectively, and the biotin capture method which yielded an average of 178,329 and ~ 1,000 raw PVY reads per leaf and tuber sample, respectively. Using the PVY oligo enrichment method, we achieved a sequence depth ranging from 27 to 535,507 reads per genome site (Figure S2) after filtering and PCR duplication removal, which allowed the reliable detection of SNVs in each sample. Furthermore, it was possible to extract a consensus sequence for each virus strain from the founding population that had no gaps (Figure 1). It is noteworthy to mention that high quality total RNA extraction from potato tubers is extremely difficult. Several commercial RNA extraction kits (e.g Qiagen RNeasy Mini Kit and Invitrogen™ Total RNA Isolation Kit) yielded low concentration and degraded total RNA from tuber samples. The PureLink™ Plant RNA Reagent from Thermo Fisher Scientific was found to yield the highest concentrations and best quality of RNA and it also worked well for leaf samples.

Nucleotide diversity decreased after the first passage in each transmission mode. The SNP data set was used to measure the changes over time in the nucleotide diversity (π) (an estimate of population genetic diversity) of the PVY populations generated from each transmission mode passage. The nucleotide diversity averaged across all bioreps decreased significantly after passage one and stabilized in the subsequent passages ($p < 0.0001$ and $p < 0.02$ for leaf and tuber populations, respectively) for all transmission modes and for both PVY isolates (Fig. 2 and 3). The nucleotide diversity within the MN21 populations in leaf samples was significantly higher in passage zero for all transmission modes ($p = 0.0005$) and in passage one in AT ($p = 0.002$) than in tubers. The nucleotide diversity was higher in tuber samples than leaf samples in the IT in passage one ($p < 0.0001$). The WI3 nucleotide diversity was significantly higher in tuber samples than in leaf samples in passages one and two for IT and AT, respectively ($p = 0.0003$). There were no statistical differences in nucleotide diversity among leaf and tuber samples in any of the other passages for any treatment or PVY isolate. (Figure 2 A and B, and Figure 3 A and B). The MN21 nucleotide diversity was significantly higher than the WI3 nucleotide diversity in passage zero and one, but similar in the subsequent passages ($p < 0.0001$) in both leaf and tuber samples.

To investigate the effect of the fluctuation of emergence and disappearance of SNVs on PVY population diversity overtime, we calculated the nucleotide diversity from standing SNVs (SNVs present in the source inoculum – passage zero) and from new SNVs (SNVs appearing overtime in the passages). In both PVY isolates, the standing nucleotide diversity decreased after passage one in all transmission modes in leaf ($p < 0.01$) and tuber ($p < 0.0003$), except leaf in IT, which had a low nucleotide diversity in passage one and remained stable in passage two (Figure 2 C and D, and Figure 3 C and D). The nucleotide diversity calculated from the new SNVs dataset remained constant after passage one in all transmission modes for both PVY isolates in leaf and tuber samples. The new SNP nucleotide diversities were not significantly different among transmission modes in either PVY isolate strains at any sampling.

Maintenance and selection of SNVs through passages by each transmission mode. The frequency of SNVs identified in at least two bioreps was determined for each isolate and transmission mode to quantify their selection and loss during the passages (Figures 4 and 5). In the founding WI3 population, the number of SNVs identified in the leaf sample (n=548) was more than double that identified in the tuber sample (n=268) (Table S2). Subsequently, the average total number of SNVs identified after the transmission mode passages were higher in the tuber samples than in the leaf samples in both the AT and MI modes (315 vs 156 for AT and 277 vs 184 for MI). The averages were similar 232 vs 240 in tuber and leaf, respectively, for the IT mode. A majority of the increased SNV count for the AT and MI tuber samples was due to new SNVs, whereas the number of standing SNVs remained nearly constant (Table S2). Most of the SNVs were present at low frequencies (<10%) in the WI3 populations (Figure 4) with only two standing SNVs identified in the MI leaf samples increasing in frequency (near 25%) in later passages (Figure 4C). Three and one new SNVs increase in frequency from passage one to passage two in the IT mode for leaf and tuber samples (Fig 4 A and B).

In the founding MN21 population, the number of SNVs identified in the leaf sample (n=424) and tuber sample (n=395) were comparable (Table S2). Similar to what was observed for the WI3 populations following the transmission mode passages, the average number of SNVs was higher in tuber samples than leaf samples; 338 vs 212 for the AT mode, 260 vs 206 for the IT mode and

321 vs 238 for the MI mode. In contrast to what was observed for the WI3 populations, the both the standing SNVs and new SNVs contributed to the differences between the tuber and leaf samples (Table S2). While most of the SNVs in the MN21 populations were present at low frequencies (<10%) across all passages (Figure 5), there were considerably more SNVs, both standing and new, in the MN21 populations than in the WI3 populations that increased in frequency in subsequent passages, several attaining frequencies >80% that would suggest these changes were moving toward fixation in the population (Fig 5).

The abundance and distribution of maintained SNVs in the PVY genome differed between PVY^{O} and PVY$^{N\text{-}Wi}$ strains. To evaluate if standing SNVs (maintained from the source population and carried through the passages) and their frequencies were statistically different between transmission modes and each tissue type (leaf and tuber), a two-sample Kolmogorov-Smirnov D test (Lehmann 2006) was calculated in R (Team 2018). The SNVs and their frequencies from each passage and sample location in each transmission mode were input variables (data shown in Fig. 4 and 5). Only SNVs that were present in at least two of the bio-replications were considered and an average of the frequencies of the SNVs from each bio-replication were entered in the model (see D values and P-values in supplemental table 3). In the WI3 populations obtained from leaves the standing SNVs differ among transmission modes in passage one ($p < 0.0001$), but are not significantly different ($p > 0.54$) in subsequent passages. In contrast, the standing SNVs identified in WI3 populations from tuber samples, were significantly different in all passages for all transmission modes. For the MN21 populations, the same analysis revealed that, except for passage two in leaf samples, the standing SNVs and their frequencies different among each passage for each transmission mode. After a close inspection of the data, it was determined that more standing SNVs carried though all the passages of the MN21 populations than the WI3 populations.

Because the leaf and tuber samples were collected at different times after inoculation, 3 wpi and 4 mpi, respectively, the standing SNVs maintained through all passages were analyzed separately. Also, because we were not able to carry all transmission modes to five passages, we compared the standing SNVs in each transmission mode relative to the last passage (Figure 6, 7,

and Supplemental Table 4). For example, the standing SNVs identified throughout the five MI and AT passages in leaf samples were compared, whereas only those carried through the third passage were compared from tuber samples. SNVs carried through the first two passages were compared among all three transmission modes in both leaf and tuber samples. We considered only SNVs that were present in at least two bio-replications for this analysis and we took an average of the frequencies of the SNVs from each bio-replication to plot the SNVs frequencies in the graphs.

A total of 392 standing SNVs were carried through the last examined passage in both isolates combined (Table 1). Of those, 304 SNVs were synonymous and 88 were nonsynonymous. We detected more than four times the number of standing SNVs in MN21 populations (319 SNVs) than in WI3 populations (73 SNVs). The cistrons 6K1 (0), CP (1), and 6K2 (2) had the lowest number of standing SNVs and the Vpg (17), Nib, (17) and CI (13) cistrons had the highest number in the WI3 populations (Table 1). In the MN21 populations, the 6K1 (1), P3 (4), and Vpg (4) cistrons contained the lowest number of SNVs and the NIb (185), HC-Pro (42), and CP (33) cistrons had the highest number of SNVs.

We identified standing SNVs unique to each transmission mode and the distribution of those SNVs varied between the two PVY isolates (Figure 6, 7, and Supplemental Table 4). Most of those standing SNVs were kept in low frequency and none of them reached 50% in frequency in WI3 populations. However, in MN21 populations four SNVs (386, 5566, 5187, and 5697) reached over 50% in frequency and two of those (386 and 5566) reached over 98% in frequency. SNV 386 was unique to AT transmission mode in a leaf sample in the fifth passage and caused nonsynonymous mutation on cistron P1. SNV 5566 was also unique to AT transmission mode but found in a tuber samples in passage 3 and caused a synonymous mutation on cistron 6K2. SNV 5187 was unique to MI transmission mode and reached 65% in frequency in a leaf sample in passage 5, it caused a nonsynonymous mutation on CI cistron. SNV 5697 was unique to IT transmission mode and reached 60% in frequency in passage 2 on a leaf sample and caused a synonymous mutation on cistron Vpg (Supplemental Table 4).

DISCUSSION

During the course of RNA virus infections, a large number of viral genomes are replicated some of which are packaged into virions that can move within or between hosts. This coupled with high mutation rates, creates the raw materials that are responsible for the rapid adaptation often observed in RNA virus populations (Zwart and Elena 2015). In our experiments, we started with a high diversity inoculum source and evaluated how that diversity is shaped by each transmission mode over time: mechanical inoculation (MI), aphid transmission (AT), and vertical transmission through infected tubers (IT). In MI, even though many viral particles can be found in the inoculum (up to 10^9 tobacco mosaic virus (TMV) particles in a tobacco leaf (Sacristan, Diaz et al. 2011)), just a few units (2-20) of TMV were found to infect plant leaves through mechanical inoculation (Sacristán, Malpica et al. 2003). On average just 0.5-3.2 PVY particles were estimated to be transmitted to a new host by an aphid vector (Moury, Fabre et al. 2007), but in our experiments, that average can be increased 10 folds as we used 10 aphids per transmission event. Therefore, the number of founders between MI and AT are likely to be comparable in our experiments and may partially explain the similar nucleotide diversities found among our virus populations perpetuated by these two transmission modes. Also, the similarity in nucleotide diversity suggests that the bottlenecks induced by those transmission modes are perhaps comparable. Although other factors such as inoculum dose, founders genetic structure, timing of infection, and host-vector-virus interaction may contribute to the magnitude of the bottleneck(Holmes 2009, Gutiérrez, Michalakis et al. 2012, Elena, Fraile et al. 2014, McCrone and Lauring 2018). Also important was the observation that several SNVs were found in low frequency in the source population and became high in frequency (over 50% and reaching up to 99% in some cases) overtime in each transmission mode for each strain (Figures 4 and 5). This is a strong evidence that natural selection was occurring in our analyzed samples, consistent with what was reported for another potyvirus, zucchini yellow mosaic virus, when transmitted by aphids (Simmons, Dunham et al. 2012). That is even more evident when we consider the four SNVs only detected in one of the transmission modes (386 and 5566 in AT, 5187 in MI, and 5697 in IT) that were present in the MN21 source population and increased in frequency standing through the AT passages. Those SNVs need to be analyzed further to determine potential

biological function, especially SNV 5187 in CI, a cistron known to be involved in aphid transmission and cell-to-cell movement (Vijayapalani, Maeshima et al. 2012), and SNV 5697 in VPg, where mutations are known to have an effect on cell-to-cell and systemic movement (Dunoyer, Thomas et al. 2004).

The distribution of mutational fitness effects (DMFE) - the fraction of all possible mutations that are beneficial, neutral, and deleterious – dictate the evolutionary fate of a given viral population in a host (Dean and Thornton 2007). If a virus is well adapted to a given host, most of the mutations are expected to be either deleterious or lethal and are quickly purged from the population (Elena, Fraile et al. 2014), resulting in a decreased population diversity. We observed similar trends in that the nucleotide diversity (π) of all SNV datasets was high in the source population, significantly reduced after the first passage in each transmission mode, and stabilized overtime (see figures 2 A and B, 3 A and B). This may be an indication of host adaptation by the virus.

Our finding that the nucleotide diversity associated with new SNVs remained constant after passage one in each transmission mode contradicts Kutnjak *and et. al.* findings, they found an increase in the π of newly formed SNVs overtime. Nevertheless, the sampling time used by those authors were different than ours, they sampled between 6 to 21 days with a break of 3 days in between sampling times, whereas we sampled at 21 days after transmission for leaves and four months for tuber samples. It can be that our sampling times were just long enough for the virus population to stabilize while in the Kutnjak *and et. al.* sampling approach, the virus population was still recovering from the effects of transmission. It would be noteworthy to see what would happen to Kutnjak *and et. al.* virus populations if they continued with the experiments a little longer. Furthermore, a possible explanation for the high π detected in the source population of both strains may also be the fact that the virus inoculum had been stored in frozen tobacco tissue after collected originally from infected plants in potato fields. Viral fitness vary across all its hosts and host-switching may play an important role in changing virus populations (Elena, Fraile et al. 2014). However, we transferred the inoculum to a potato plant before the starting of the experiments and used systemically infected leaves from the infected potato

plant to inoculate the founding population of each strain to dilute possible effects of host-switching.

The higher nucleotide diversity in source population of PVY^{N-Wi} than PVY^{O}, may be an explanation to a current phenomenon, the change in prevalence of PVY strains in the potato fields worldwide (Karasev and Gray 2013, Quenouille, Vassilakos et al. 2013). From surveys conducted by our group since 2004 on over 8,000 samples, more than 90% of the PVY samples collected in recent years were infected with recombinant strains (including PVY^{N-Wi}); on the other hand, PVY^{O}, which was prevalent prior to 2005 is a minor component of the total PVY population affecting potato in the US. We hypothesize that the higher initial population diversity in PVY^{N-Wi} gives this strain a fitness advantage over PVY^{O}. From our results, a few pieces of evidence support this premise: (1) - PVY^{N-Wi} had a higher initial population diversity than PVY^{O} strain (Figure 2 and 3); (2) - more SNVs reached 50% frequency in PVY^{N-Wi} populations over time than in PVY^{O} (Figure 4 and 5); and (3) - more standing SNVs from source population carried through the last passage evaluated in PVY^{N-Wi} strain (319) than in PVY^{O} populations (73) (Table 2). Even though we did not attempt to estimate viral fitness in our experiments, it has been shown that that high virus population diversity increase virus fitness (Moya, Elena et al. 2000, Gutiérrez, Michalakis et al. 2012, Elena, Fraile et al. 2014, McCrone and Lauring 2018), which help to support out fitness advantage hypothesis. Furthermore, population diversity has been positively correlated with viral population size (Bailey, Lauck et al. 2014) and large population size was strongly correlated with the effects associated with beneficial mutations that become fixed in the population of vesicular stomatitis virus (VSV) (Miralles, Gerrish et al. 1999). Additionally, reduced poliovirus diversity leads to loss of neurotropism and an attenuated pathogenic phenotype (Vignuzzi, Stone et al. 2006). Our findings warrant the necessity of designing specific studies to evaluate fitness in both PVY isolate populations to better test our strain fitness advantage hypothesis.

Detecting unique standing SNVs in populations maintained by each transmission mode suggest that there is a specific selection pressure exerted from each transmission mode (Figure 6 Supplemental Table 4). This is supported by the fact that standing SNVs were detected in the PVY

genome cistrons known to be associated with aphid transmission (CP and HC-Pro) and cell to cell movement of the virus (CI and CP) (Andret-Link and Fuchs 2005). However, the signatures of selection - changes in the genome that have been increased in frequency in a population because of their functional importance in specific processes (Nielsen 2005) - seems to vary by PVY strain, as the standing SNVs in each transmission mode, passage, and sample location were different between the isolates analyzed in this study (Figure 6 and Supplemental Table 4). Furthermore, the fitness effects of these SNVs in the population of the two PVY isolates need to be examined in more detail. The fact that in both strains, the P3 and 6K1 cistrons had the lowest amount of SNVs detected, indicate that those genome regions are less prone for changes, an information that turned out to have vital practical uses in designing probes for PCR based tests. As an example, we found in our data analysis that a single point mutation (80% in frequency in the virus population) in the binding site of the forward primer, at position 5727 on the VPg cistron of the PVY^N genome, resulted in a faint band or no amplification, producing inconclusive results when an established multiplex PVY-PCR test (Lorenzen, Piche et al. 2006) was used, this was corrected by designing a degenerate primer for the site (data not show). Our results can be used from now on by scientists to design PVY primers targeting the regions on PVY genome that are less prone to the occurrence of mutations to avoid such problems in the future.

NIb had one of the highest amount of SNVs in both strains overall, implying that this cistron is flexible to changes (Figure 6). NIb codes for the potyviruses' RNA-dependent RNA polymerase (RdRp) (Hong and Hunt 1996) and it is regarded as one of the most constrained proteins expressed by the PVY genome and any variation on this region can cause significant changes in fitness (Moury, Desbiez et al. 2006). In fact, a single nonsynonymous point mutation on PVY NIb cistron was found to be enough to confer virulence on pepper plants bearing the *Pvr4* PVY-resistant gene (Janzac, Montarry et al. 2010). We detected fewer SNVs in the NIb cistron of the WI3 populations than in MN21 populations. However most of the SNVs in WI3 populations were over 10% in frequency while in the MN21 populations the SNVs were all below 10% in frequency. The high number of changes found in MN21 populations, even though in low frequency, may result in different versions of RdRp, which may be behind the higher diversity

found in the MN21 populations relative to the WI3 populations. One method to test this hypothesis would be to switch this cistron in both strains and see if $PVY^{N\text{-}Wi}$ expressing the PVY^{O} NIb would yield a less diverse population and vice and versa.

Our results indicate that the population diversity of PVY varies by the virus strain but doesn't differ among transmission modes. More importantly, we compiled evidence to suggest that the transmission mode plays an important role in shaping PVY populations and it appears that selection is one of the forces governing those differences in the analyzed samples. This was highlighted by the fact that we detected unique standing SNVs to each transmission mode. Taken all together, our findings add additional knowledge towards a deeper understanding of the complexity of RNA viruses' biology and the key role that virus population diversity plays in diagnostics and on the foundation of viral diseases management strategies.

Literature Cited

Acevedo, A. and R. Andino (2014). "Library preparation for highly accurate population sequencing of RNA viruses." Nat. Protocols **9**(7): 1760-1769.

Alexopoulos, A. A., G. Aivalakis, K. A. Akoumianakis and H. C. Passam (2008). "Effect of gibberellic acid on the duration of dormancy of potato tubers produced by plants derived from true potato seed." Postharvest Biology and Technology **49**(3): 424-430.

Ali, M. C., T. Maoka, K. T. Natsuaki and T. Natsuaki (2010). "The simultaneous differentiation of *Potato virus Y* strains including the newly described strain PVY^{NTN-NW} by multiplex PCR assay." Journal of Virological Methods **165**(1): 15-20.

Andret-Link, P. and M. Fuchs (2005). "Transmission specificity of plant viruses by vectors." Journal of Plant Pathology **87**(3): 153-165.

Andrews, S. (2010). "FastQC: A quality control tool for high throughput sequence data " Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Aronesty, E. (2011). "Command-line tools for processing biological sequencing data." ea-utils.

Bailey, A. L., M. Lauck, A. Weiler, S. D. Sibley, J. M. Dinis, Z. Bergman, C. W. Nelson, M. Correll, M. Gleicher, D. Hyeroba, A. Tumukunde, G. Weny, C. Chapman, J. H. Kuhn, A. L. Hughes, T. C. Friedrich, T. L. Goldberg and D. H. O'Connor (2014). "High Genetic Diversity and Adaptive Potential of Two Simian Hemorrhagic Fever Viruses in a Wild Primate Population." PLOS ONE **9**(3): e90714.

Bates, D., M. Mächler, B. Bolker and S. Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4." 2015 **67**(1): 48.

Beerenwinkel, N. and O. Zagordi (2011). "Ultra-deep sequencing for the analysis of viral populations." Current Opinion in Virology **1**(5): 413-418.

Boodley, J. W. and R. Sheldrake (1982). "Cornell peat-like mixes for commercial plant growing." Cornell Inf. Bull. 43.

Carrasco, P., F. de la Iglesia and S. F. Elena (2007). "Distribution of Fitness and Virulence Effects Caused by Single-Nucleotide Substitutions in *Tobacco Etch Virus*." Journal of Virology **81**(23): 12979-12984.

Chung, B. Y.-W., W. A. Miller, J. F. Atkins and A. E. Firth (2008). "An overlapping essential gene in the Potyviridae." Proceedings of the National Academy of Sciences **105**(15): 5897-5902.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu and D. M. Ruden (2012). "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w(1118); iso-2; iso-3." Fly **6**(2): 80-92.

Coleman, J. R., D. Papamichail, S. Skiena, B. Futcher, E. Wimmer and S. Mueller (2008). "Virus Attenuation by Genome-Scale Changes in Codon Pair Bias." Science **320**(5884): 1784-1787.

Dean, A. M. and J. W. Thornton (2007). "Mechanistic approaches to the study of evolution: the functional synthesis." Nature reviews. Genetics **8**(9): 675-688.

Dunoyer, P., C. Thomas, S. Harrison, F. Revers and A. Maule (2004). "A Cysteine-Rich Plant Protein Potentiates Potyvirus Movement through an Interaction with the Virus Genome-Linked Protein VPg." Journal of Virology **78**(5): 2301-2309.

Eigen, M. and C. Biebricher (1988). RNA Genetics, Vol. III, Variability of RNA Genomes. Eds E. Domingo, JJ Holland and P. Ahlquist. CRC Press Inc., Florida**:** 211-245.

Elena, S. F., A. Fraile and F. García-Arenal (2014). Chapter Three - Evolution and Emergence of Plant Viruses. Advances in Virus Research. K. Maramorosch and F. A. Murphy, Academic Press. **88:** 161-191.

García-Arenal, F., A. Fraile and J. M. Malpica (2001). "Variability and genetic structure of plant virus populations." Annual Review of Phytopathology **39**(1): 157-186.

Gutiérrez, S., Y. Michalakis and S. Blanc (2012). "Virus population bottlenecks during within-host progression and host-to-host transmission." Current Opinion in Virology **2**(5): 546-555.

Holmes, E. C. (2009). "The Evolutionary Genetics of Emerging Viruses." Annual Review of Ecology, Evolution, and Systematics **40**(1): 353-372.

Hong, Y. and A. G. Hunt (1996). "RNA Polymerase Activity Catalyzed by a Potyvirus-Encoded RNA-Dependent RNA Polymerase." Virology **226**(1): 146-151.

Janzac, B., J. Montarry, A. Palloix, O. Navaud and B. Moury (2010). "A Point Mutation in the Polymerase of Potato virus Y Confers Virulence Toward the Pvr4 Resistance of Pepper and a High Competitiveness Cost in Susceptible Cultivar." Molecular Plant-Microbe Interactions **23**(6): 823-830.

Karasev, A. and S. Gray (2013). "Continuous and Emerging Challenges of *Potato virus Y* in Potato." Annual Review of Phytopathology **51**: 571-586.

Karasev, A. V., X. Hu, C. J. Brown, C. Kerlan, O. V. Nikolaeva, J. M. Crosslin and S. M. Gray (2011). "Genetic Diversity of the Ordinary Strain of Potato virus Y (PVY) and Origin of Recombinant PVY Strains." Phytopathology **101**(7): 778-785.

Kutnjak, D., S. F. Elena and M. Ravnikar (2017). "Time-Sampled Population Sequencing Reveals the Interplay of Selection and Genetic Drift in Experimental Evolution of Potato Virus Y." Journal of Virology **91**(16).

Lehmann, E. L. (2006). Nonparametrics: Statistical Methods Based on Ranks, Springer-Verlag New York.

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows–Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and G. P. D. P. Subgroup (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Lorenzen, J., L. M. Piche, N. C. Gudmestad, T. Meacham and P. Shiel (2006). "A Multiplex PCR Assay to Characterize *Potato virus Y* Isolates and Identify Strain Mixtures." Plant Disease **90**(7): 935-940.

Marsh, G. A., R. Rabadán, A. J. Levine and P. Palese (2008). "Highly Conserved Regions of Influenza A Virus Polymerase Gene Segments Are Critical for Efficient Viral RNA Packaging." Journal of Virology **82**(5): 2295-2304.

McCrone, J. T. and A. S. Lauring (2018). "Genetic bottlenecks in intraspecies virus transmission." Current Opinion in Virology **28**: 20-25.

McMullan, L. K., A. Grakoui, M. J. Evans, K. Mihalik, M. Puig, A. D. Branch, S. M. Feinstone and C. M. Rice (2007). "Evidence for a functional RNA element in the hepatitis C virus core gene." Proceedings of the National Academy of Sciences **104**(8): 2879-2884.

Miralles, R., P. J. Gerrish, A. Moya and S. F. Elena (1999). "Clonal Interference and the Evolution of RNA Viruses." Science **285**(5434): 1745-1747.

Moury, B., C. Desbiez, M. Jacquemond and H. Lecoq (2006). Genetic Diversity of Plant Virus Populations: Towards Hypothesis Testing in Molecular Epidemiology. Advances in Virus Research, Academic Press. **67:** 49-87.

Moury, B., F. Fabre and R. Senoussi (2007). "Estimation of the number of virus particles transmitted by an insect vector." Proceedings of the National Academy of Sciences **104**(45): 17891-17896.

Moya, A., S. F. Elena, A. Bracho, R. Miralles and E. Barrio (2000). "The evolution of RNA viruses: A population genetics view." Proceedings of the National Academy of Sciences **97**(13): 6967-6973.

Nelson, C. W., L. H. Moncla and A. L. Hughes (2015). "SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data." Bioinformatics **31**(22): 3709-3711.

Nielsen, R. (2005). "Molecular Signatures of Natural Selection." Annual Review of Genetics **39**(1): 197-218.

Posada-Cespedes, S., D. Seifert and N. Beerenwinkel (2017). "Recent advances in inferring viral diversity from high-throughput sequencing data." Virus Research **239**: 17-32.

Quenouille, J., N. Vassilakos and B. Moury (2013). "*Potato virus Y*: a major crop pathogen that has provided major insights into the evolution of viral pathogenicity." Molecular Plant Pathology **14**(5): 439-452.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Rojas, M. R., F. M. Zerbini, R. F. Allison, R. L. Gilbertson and W. J. Lucas (1997). "Capsid protein and helper component-proteinase function as potyvirus cell-to-cell movement proteins." Virology **237**(2): 283-295.

Sacristan, S., M. Diaz, A. Fraile and F. Garcia-Arenal (2011). "Contact transmission of tobacco mosaic virus: A quantitative analysis of parameters relevant for virus evolution." J Virol **85**(10): 4974-4981.

Sacristán, S., J. M. Malpica, A. Fraile and F. García-Arenal (2003). "Estimation of population bottlenecks during systemic movement of tobacco mosaic virus in tobacco plants." Journal of Virology **77**(18): 9906-9911.

Simmonds, P. and D. B. Smith (1999). "Structural constraints on RNA virus evolution." J Virol **73**(7): 5787-5794.

Simmonds, P., A. Tuplin and D. J. Evans (2004). "Detection of genome-scale ordered RNA structure (GORS) in genomes of positive-stranded RNA viruses: Implications for virus evolution and host persistence." RNA **10**(9): 1337-1351.

Simmons, H. E., J. P. Dunham, J. C. Stack, B. J. A. Dickins, I. Pagán, E. C. Holmes and A. G. Stephenson (2012). "Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus." Journal of General Virology **93**(8): 1831-1840.

Team, R. D. C. (2018). A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing.

Thorvaldsdóttir, H., J. T. Robinson and J. P. Mesirov (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." Briefings in Bioinformatics **14**(2): 178-192.

Urcuqui-Inchima, S., A.-L. Haenni and F. Bernardi (2001). "Potyvirus proteins: a wealth of functions." Virus Research **74**(1–2): 157-175.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel and M. A. DePristo (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics, John Wiley & Sons, Inc.

Vignuzzi, M., J. K. Stone, J. J. Arnold, C. E. Cameron and R. Andino (2006). "Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population." Nature **439**(7074): 344-348.

Vijayapalani, P., M. Maeshima, N. Nagasaki-Takekuchi and W. A. Miller (2012). "Interaction of the Trans-Frame Potyvirus Protein P3N-PIPO with Host Protein PCaP1 Facilitates Potyvirus Movement." PLoS Pathog **8**(4): e1002639.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York.

Wilm, A., P. P. K. Aw, D. Bertrand, G. H. T. Yeo, S. H. Ong, C. H. Wong, C. C. Khor, R. Petric, M. L. Hibberd and N. Nagarajan (2012). "LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets." Nucleic Acids Research **40**(22): 11189-11201.

Wysoker, A., K. Tibbetts and T. Fennell (2013). "Picard tools version 1.90." http://picard.sourceforge.net.

Zhong, S., J. G. Joung, Y. Zheng, Y. R. Chen, B. Liu, Y. Shao, J. Z. Xiang, Z. Fei and J. J. Giovannoni (2011). "High-throughput illumina strand-specific RNA sequencing library preparation." Cold Spring Harb Protoc(8): 940-949.

Zwart, M. P. and S. F. Elena (2015). "Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution." Annual Review of Virology **2**(1): 161-179.

Figure Legends

Figure 1: Experimental Design. PVY infected lyophilized tobacco tissue was use for mechanically inoculating a single potato plant, the founding population plant. The virus from the founding plant was then mechanically inoculated into three potato plants (source plants), each plant was considered a replication and those plants were used as virus source for each transmission mode: AT - aphid transmission (represented by the aphid cartoon), MI - mechanical inoculation (represented by mortar and pestle cartoon), and IT - infected tuber (represented by the potato tuber cartoon). Each time the virus was transmitted to a new plant (in all three transmission modes) we counted it as a passage. A leaf and a tuber sample (represented by the red circle) were collected from each plant to compare the virus quasispecies in those locations and among each transmission mode over time. We didn't deep sequences samples from passage 4 and in IT experiments, plants didn't survive the third passage onward so we analyzed only samples from passage 1 and 2. In AT and MI, no tuber was produced in passage 5, then we didn't deep sequenced those tuber samples either. The procedure was carried on for two different PVY strains, PVY^{O} (isolate WI3) and $PVY^{N\text{-}Wi}$ (isolate MN21). Potato seedlings, cv 'Goldrush', at five-to-six-leaf stage, that was tested free from PVY infection by monoclonal double-antibody sandwich ELISA (Agdia, Elkhart, IN-USA), was used in the experiments.

Figure 2: PVY^{O} population diversity dynamics. Average population nucleotide diversity (π) was calculated for all SNVs in leaves (A) and tubers (B) samples, for SNVs detected in the source population (standing SNVs) in leaves (C) and tubers (D) samples, and for new arising SNVs (new SNVs) in leaves (E) and tubers (F) samples. Drawn lines connect π values calculated in each passage. Purple lines are the nucleotide diversities of infected tuber transmission mode populations - replications 1 (IT-1), 2 (IT-2), and 3 (IT-3). Blue lines are the nucleotide diversities of mechanical inoculation populations - replications 1 (MI-1), 2 (Mi-2), and 3 (MI-3). Green lines are the nucleotide diversities of aphid transmission populations - replications 1 (AT-1), 2 (AT-2), and 3 (AT-3). In X axis, 0 = Source population, 1 = Passage 1, 2 = Passage 2, 3 = Passage 3, and 5 = passage 5.

Figure 3: PVY^{N-Wi} population diversity dynamics. Average population nucleotide diversity (π) was calculated for all SNVs in leaves (A) and tubers (B) samples, for SNVs detected in the source population (standing SNVs) in leaves (C) and tubers (D) samples, and for new arising SNVs (new SNVs) in leaves (E) and tubers (F) samples. Drawn lines connect π values calculated in each passage. Purple lines are the nucleotide diversities of infected tuber transmission mode populations - replications 1 (IT-1), 2 (IT-2), and 3 (IT-3). Blue lines are the nucleotide diversities of mechanical inoculation populations - replications 1 (MI-1), 2 (Mi-2), and 3 (MI-3). Green lines are the nucleotide diversities of aphid transmission populations - replications 1 (AT-1), 2 (AT-2), and 3 (AT-3). In X axis, 0 = Source population, 1 = Passage 1, 2 = Passage 2, 3 = Passage 3, and 5 = passage 5.

Figure 4: Fates of arising mutations in PVY^{O} populations. Frequency of SNVs identified through all passages of PVY^{O} populations in leaves and tubers samples in the three transmission modes **(A-B)** Infected Tuber, **(C-D)** Mechanical Inoculation, and **(E-F)** Aphid Transmission. Blue trajectories represent SNVs that were present in the source population and in each subsequent passage. Red trajectories represent SNVs that were not present in the source population but were detect in each subsequent passage. Black lines represent SNVs that present in some passages but not in others. Solid lines are nonsynonymous changes and dashed lines are synonymous changes.

Figure 5: Fates of arising mutations in PVY^{N-Wi} populations. Frequency of SNVs identified through all passages of PVY^{O} populations in leaves and tubers samples in the three transmission modes **(A-B)** Infected Tuber, **(C-D)** Mechanical Inoculation, and **(E-F)** Aphid Transmission. Blue trajectories represent SNVs that were present in the source population and in each subsequent passage. Red trajectories represent SNVs that were not present in the source population but were detect in each subsequent passage. Black lines represent SNVs that present in some passages but not in others. Solid lines are nonsynonymous changes and dashed lines are synonymous changes.

Figure 6: Genome location of the standing SNVs (SNVs detected in the source population and carried through the last passage) in each transmission mode in leaf samples. SNVs unique to MI and AT detected on passage 5 in PVY^{O} (A) and in PVY^{N-Wi} (B). Unique SNVs to MI, AT, and IT detected on passage 2 in PVY^{O} (C) and in PVY^{N-Wi} (D). MI = Mechanical Inoculation, AT = Aphid Transmission, and IT = Infected Tuber.

Figure 7: Genome location of the standing SNVs (SNVs detected in the source population and carried through the last passage) in each transmission mode in tuber samples. SNVs unique to MI and AT detected on passage 3 in PVY^{O} (A) and in PVY^{N-Wi} (B). Unique SNVs to MI, AT, and IT detected on passage 2 in PVY^{O} (C) and in PVY^{N-Wi} (D). MI = Mechanical Inoculation, AT = Aphid Transmission, and IT = Infected Tuber.

Supplemental Figure 1: Distribution and frequency of reads and SNVs throughout the genome of PVY^{O} strain. Reads were aligned against PVY^{O} strain NGS consensus sequence extracted from sample (A) by aligning reads against the WI3 isolate reference genome (GenBank: KY848031.1) of PVY^{O} strain. SNVs are represented by the red histogram bars with frequency scale 0 (none of the reads aligned to that region had the SNV) to 100 (100% of reads aligned to that region had the SNV). The grey curve represents the genome coverage. A single potato seedling cv 'Goldrush' was mechanically inoculate with PVY^{O} strain WI3 isolate. A = Sample collected from the top leaf at three weeks after inoculation. B = Sample collected from the top leaf at four months after inoculation (harvest). C = Sample collected from the middle-top leaf at four months after inoculation (harvest). D = Sample collected from the middle-bottom leaf at four months after inoculation (harvest). E = Sample collected from the bottom leaf at four months after inoculation (harvest). F = Sample collected from the tuber at four months after inoculation (harvest). G = Sample collected from the sprout at four months after inoculation (harvest).

Supplemental Figure 2: Sequence reads coverage and depth for samples from all transmission modes and locations. AT = Aphid Transmission, MI = Mechanical Inoculation, and IT = Infected Tuber.

Tables

Table 1: Distribution throughout of the PVY genome of the SNVs that were maintained from source populations and carried through the last examined passage in each sample location (leaf and tuber) in each PVY strain. O = PVY^{O} strain and N-Wi = $PVY^{N\text{-}Wi}$ strain. MI = Mechanical Inoculation, AT = Aphid Transmission, and IT = Infected Tuber.

			Cistrons																				
			P1		HC-Pro		P3		6K1		CI		6K2		Vpg		Nia		Nib		CP		
Strain	Treatment	Location	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Syn	Nonsyn	Total
O	AT	Leaf	-	-	-	-	1	-	-	-	-	-	-	-	1	1	-	-	-	1	-	-	**4**
O	AT	Tuber	-	-	-	-	1	-	-	-	3	2	1	-	3	-	3	-	-	-	-	-	**13**
O	IT	Leaf	2	1	1	-	2	2	-	-	1	2	-	-	6	1	-	-	2	5	-	-	**25**
O	IT	Tuber	1	-	-	-	1	-	-	-	1	1	-	-	1	-	1	-	4	-	-	-	**10**
O	MI	Leaf	-	-	2	-	1	-	-	-	-	-	-	-	1	-	-	-	1	1	-	-	**6**
O	MI	Tuber	1	-	1	-	1	-	-	-	3	-	1	-	3	-	2	-	2	-	-	1	**15**
	Sub-Total		**4**	**1**	**4**	**0**	**7**	**2**	**0**	**0**	**8**	**5**	**2**	**0**	**15**	**2**	**6**	**0**	**9**	**7**	**0**	**1**	**73**
N-Wi	AT	Leaf	-	1	2	1	-	-	-	-	-	1	-	-		-	-	1	5	1	2	-	**14**
N-Wi	AT	Tuber	4	1	10	3	3	-	-	-	7	1	2	-	1	-	2	2	60	13	11	6	**126**
N-Wi	IT	Leaf	-	-	4	1	1	-	-	-	-	2	3	-	1	-	1	-	17	4	2	1	**37**
N-Wi	IT	Tuber	3	-	4	2	-	-	1	-	1	2	2	-	1	-	-	1	43	7	3	1	**71**
N-Wi	MI	Leaf	-	-	1	1	-	-	-	-	-	2	1	1	-	-	-	1	22	4	3	-	**36**
N-Wi	MI	Tuber	1	3	11	2	-	-	-	-	1	2	1	-	1	-	-	-	6	3	3	1	**35**
	Sub-Total		**8**	**5**	**32**	**10**	**4**	**0**	**1**	**0**	**9**	**10**	**9**	**1**	**4**	**0**	**3**	**5**	**153**	**32**	**24**	**9**	**319**
	Total		**12**	**6**	**36**	**10**	**11**	**2**	**1**	**0**	**17**	**15**	**11**	**1**	**19**	**2**	**9**	**5**	**162**	**39**	**24**	**10**	**392**

Supplemental Table 1 – SNVs count summary. SNVs detected in each strain, transmission mode, sample location, and passage. For each time point, the total number of detected SNVs were divided in standing SNVs (SNVs present in the source population) and new SNVs (SNVs newly appeared in each time point).

Supplemental Table 2 – Two-sample Kolmogorov-Smirnov D test Results. SNVs selected from the source population (and carried through the passages) and their frequencies were compared between transmission modes and sample locations (leaf and tuber) overtime to test if the same SNVs were, statistically, selected in each time point.

Supplemental Table 3 – List of selected SNVs (SNVs detected in the source population and carried through the passages) in each transmission mode. Comparisons were made according to the last possible passage in each transmission mode and sample location (leaf and tuber). We compared in mechanical inoculation (MI) and aphid transmission (AT) the SNVs that were present in the source population and carried through the fifth passage in leaf samples. In tuber

samples, we compared MI and AT the SNVs carried thought the third passage. Then, we compared between all three transmission modes MI, AT, and infected tuber (IT) the SNVs the carried through up to second passages in both leaf and tuber samples.

Supplemental Table 4 – Primer sets used in the PVY sequences enrichment step.

Figures

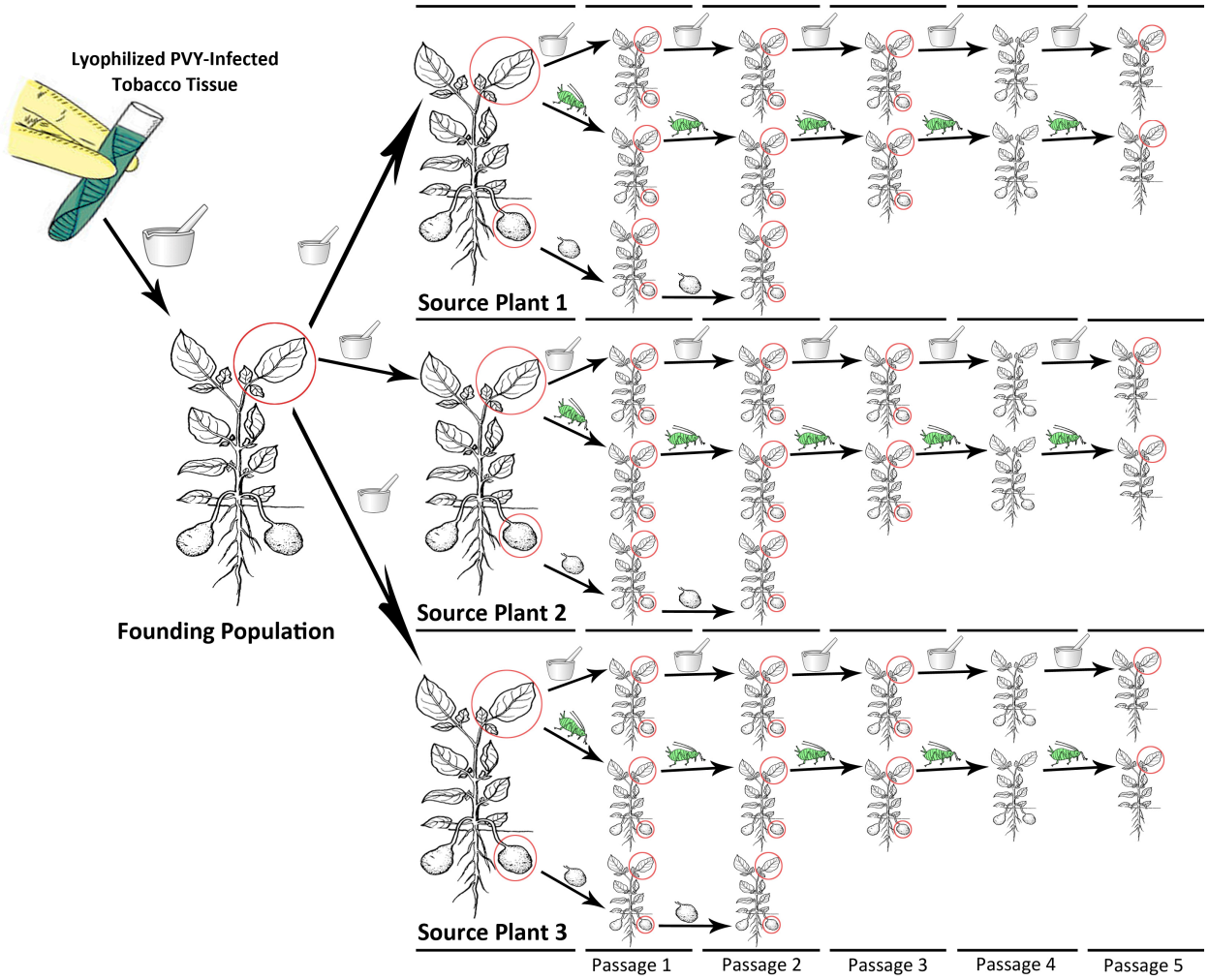


Figure 1

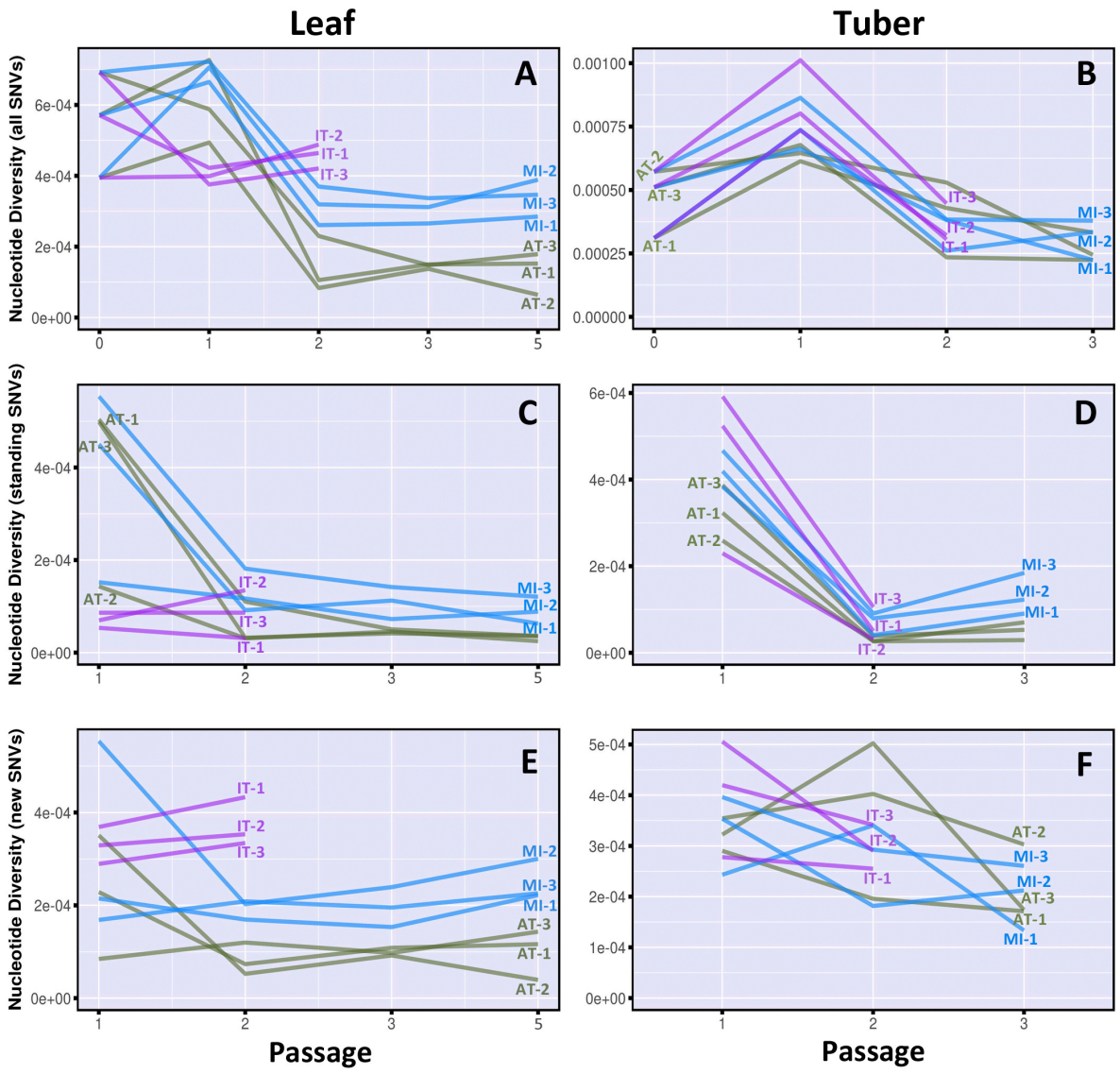

Leaf
Tuber
A
B
C
D
E
F
Nucleotide Diversity (all SNVs)
Nucleotide Diversity (standing SNVs)
Nucleotide Diversity (new SNVs)
Passage
Passage
IT-1
IT-2
IT-3
MI-1
MI-2
MI-3
AT-1
AT-2
AT-3

Figure 2

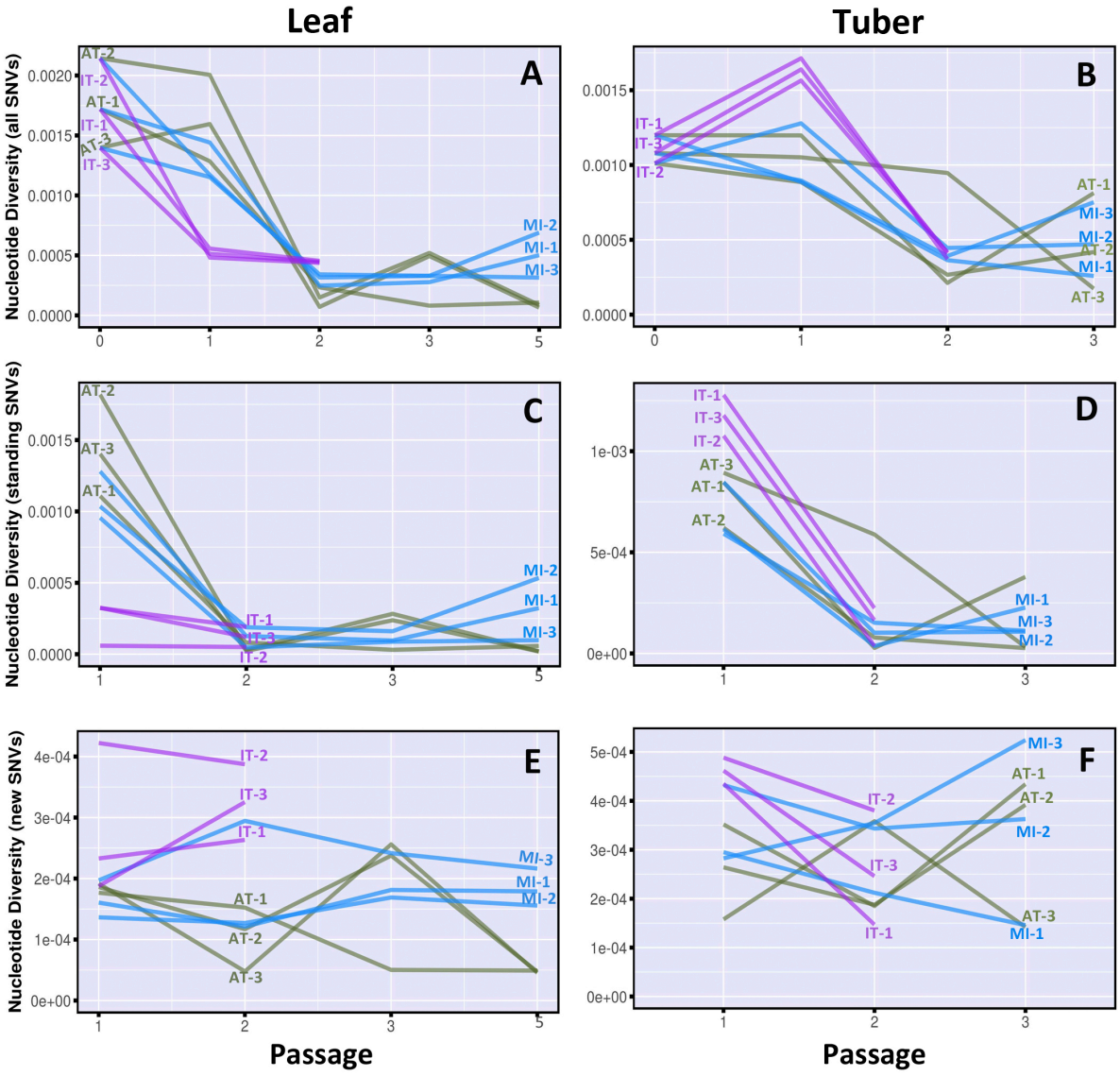

Leaf
Tuber
A
B
C
D
E
F
Nucleotide Diversity (all SNVs)
Nucleotide Diversity (standing SNVs)
Nucleotide Diversity (new SNVs)
Passage
Passage
AT-1
AT-2
AT-3
IT-1
IT-2
IT-3
MI-1
MI-2
MI-3

Figure 3

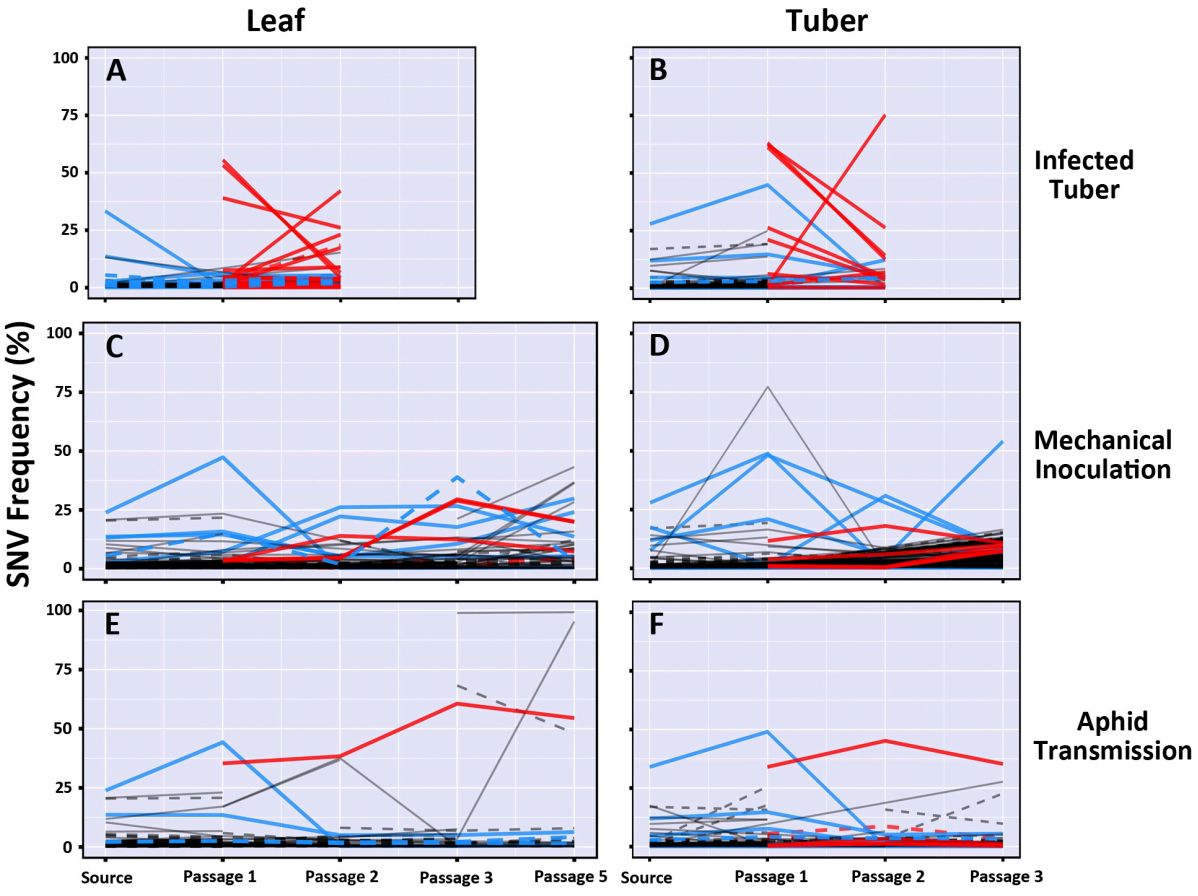

Leaf
Tuber
A
B
C
D
E
F
100
75
50
25
0
SNV Frequency (%)
Infected Tuber
Mechanical Inoculation
Aphid Transmission
Source
Passage 1
Passage 2
Passage 3
Passage 5

Figure 4

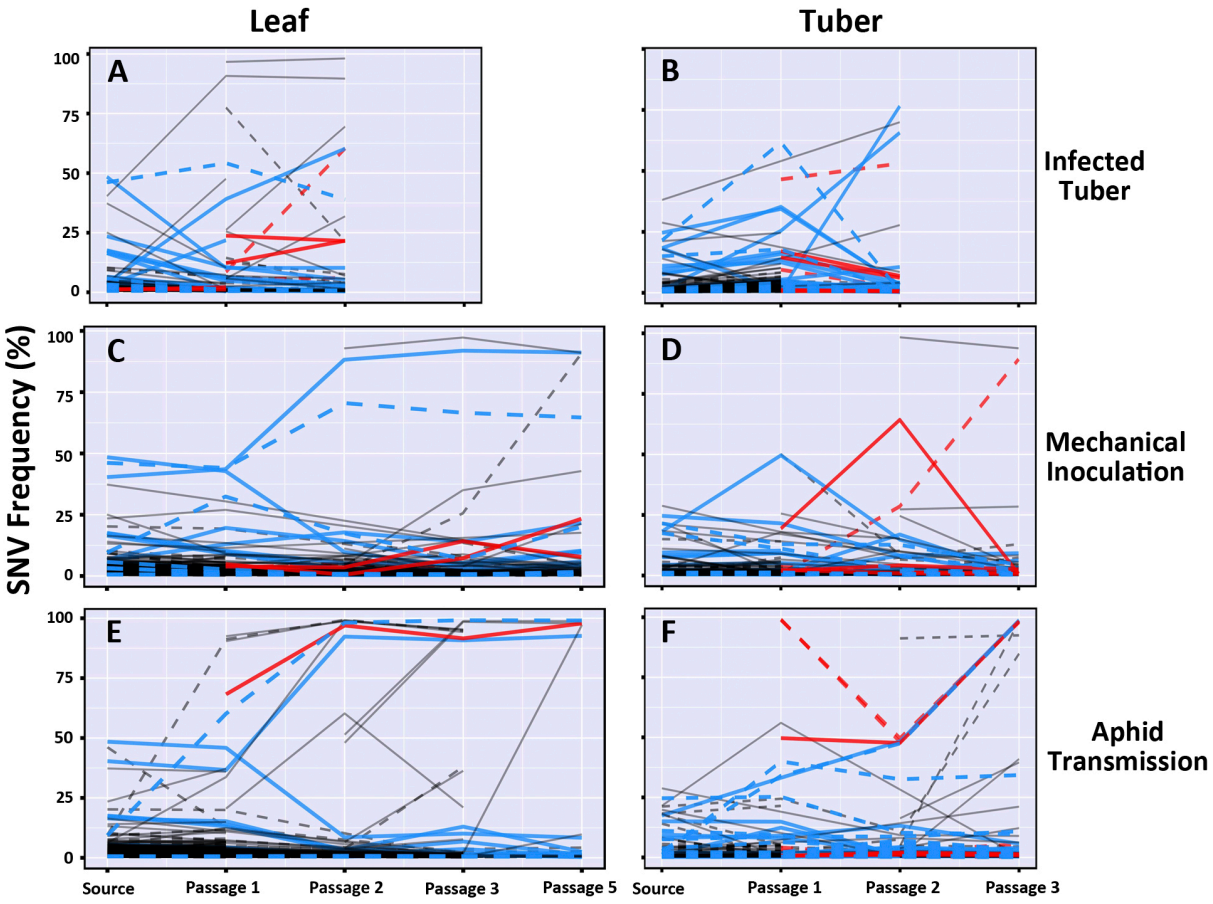

Leaf
Tuber
A
B
C
D
E
F
Infected Tuber
Mechanical Inoculation
Aphid Transmission
SNV Frequency (%)
100
75
50
25
0
Source
Passage 1
Passage 2
Passage 3
Passage 5

Figure 5

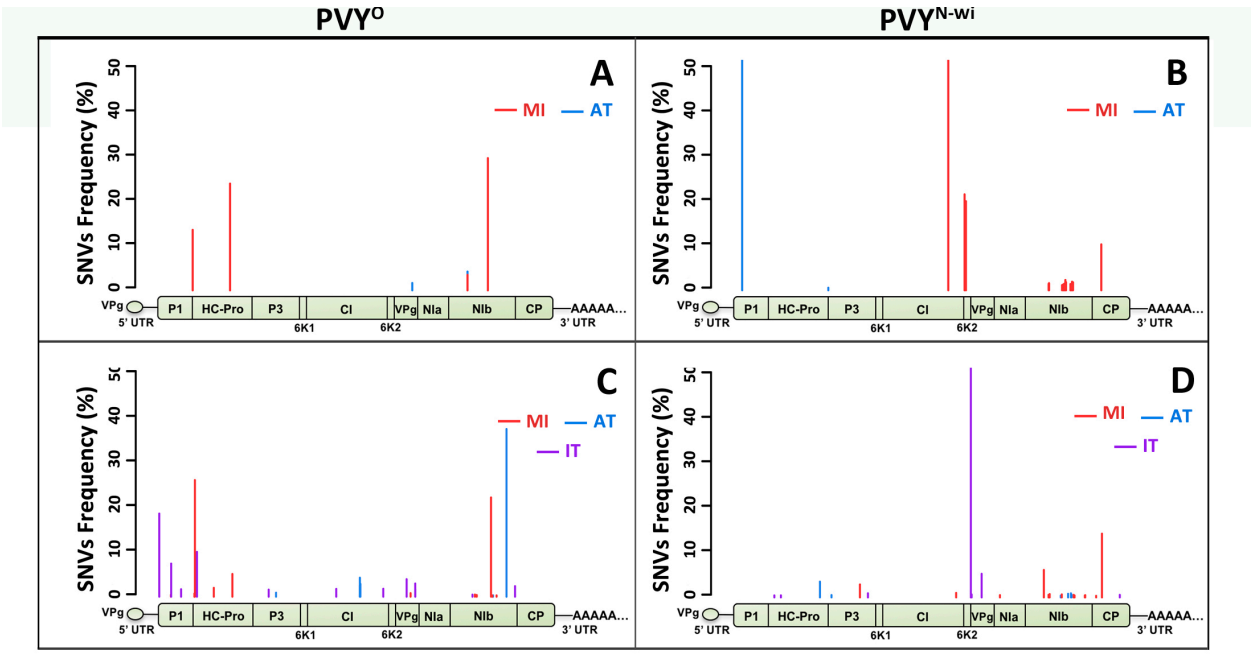


Figure 6

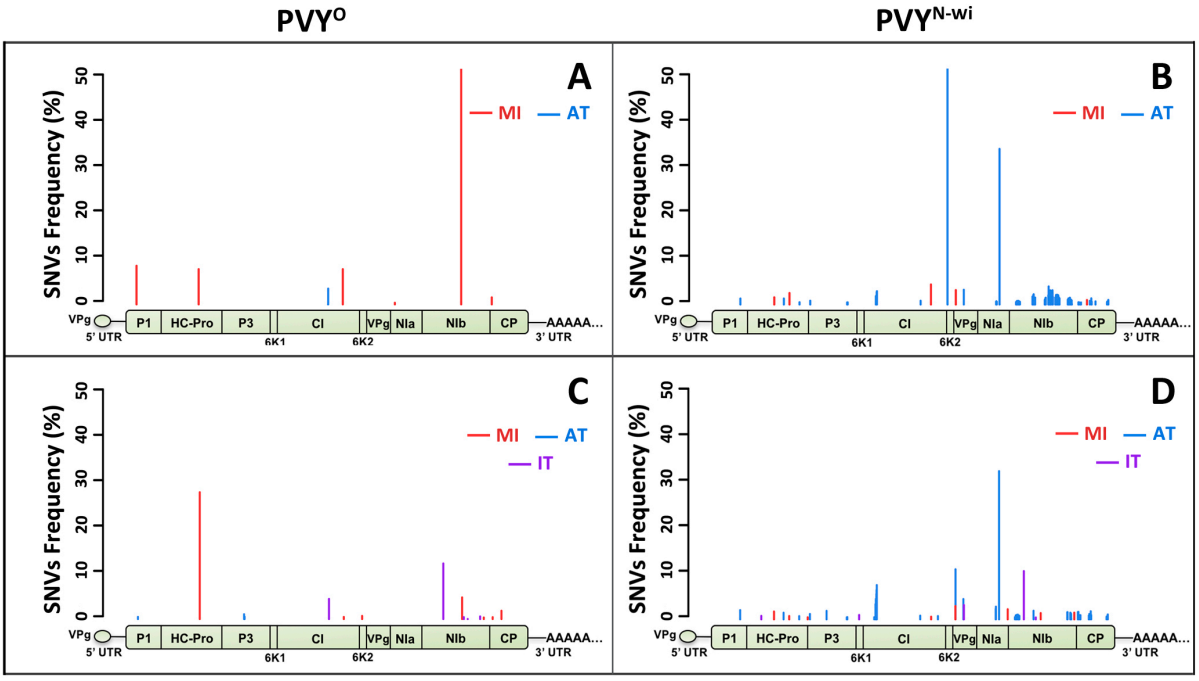
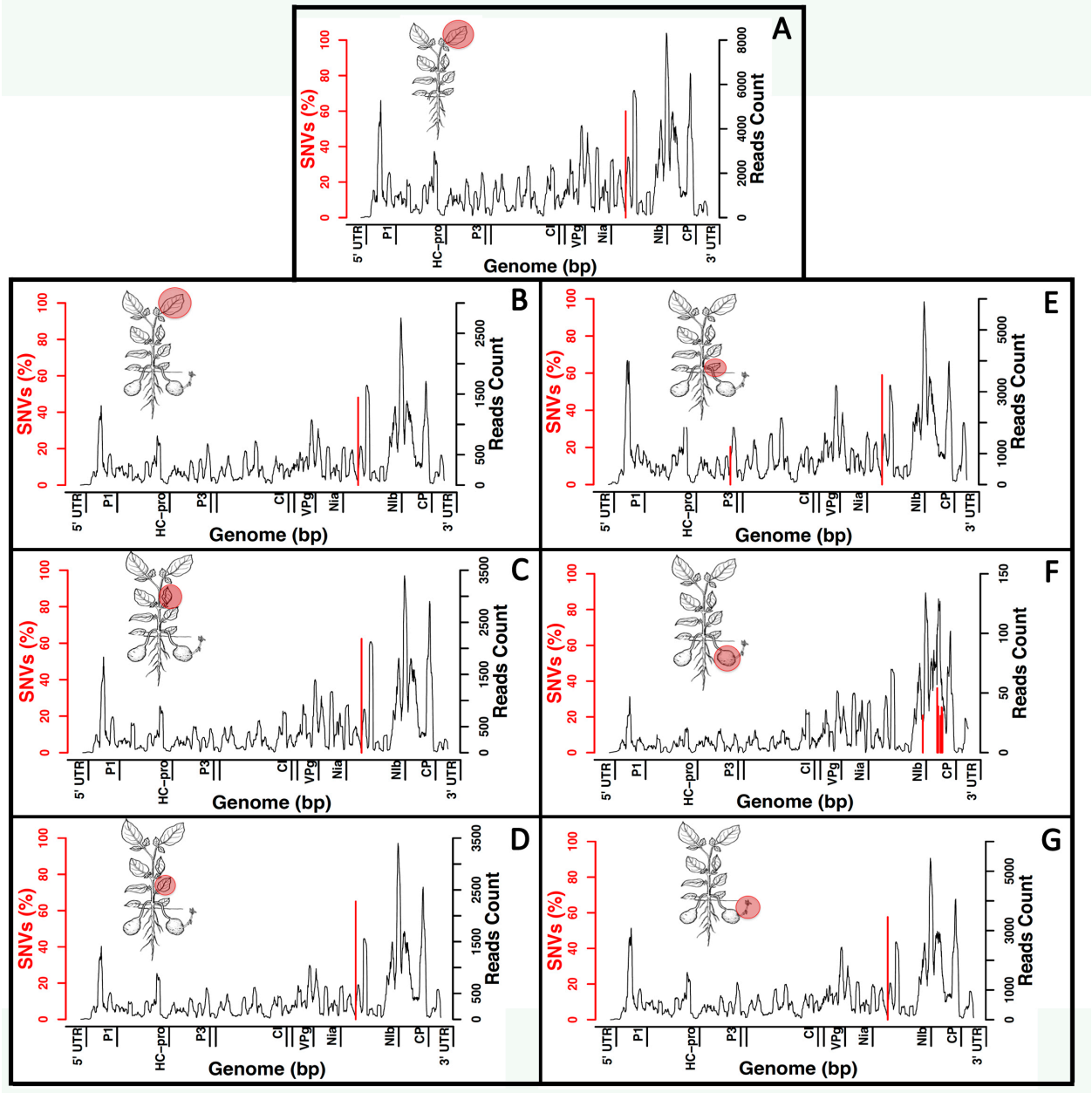
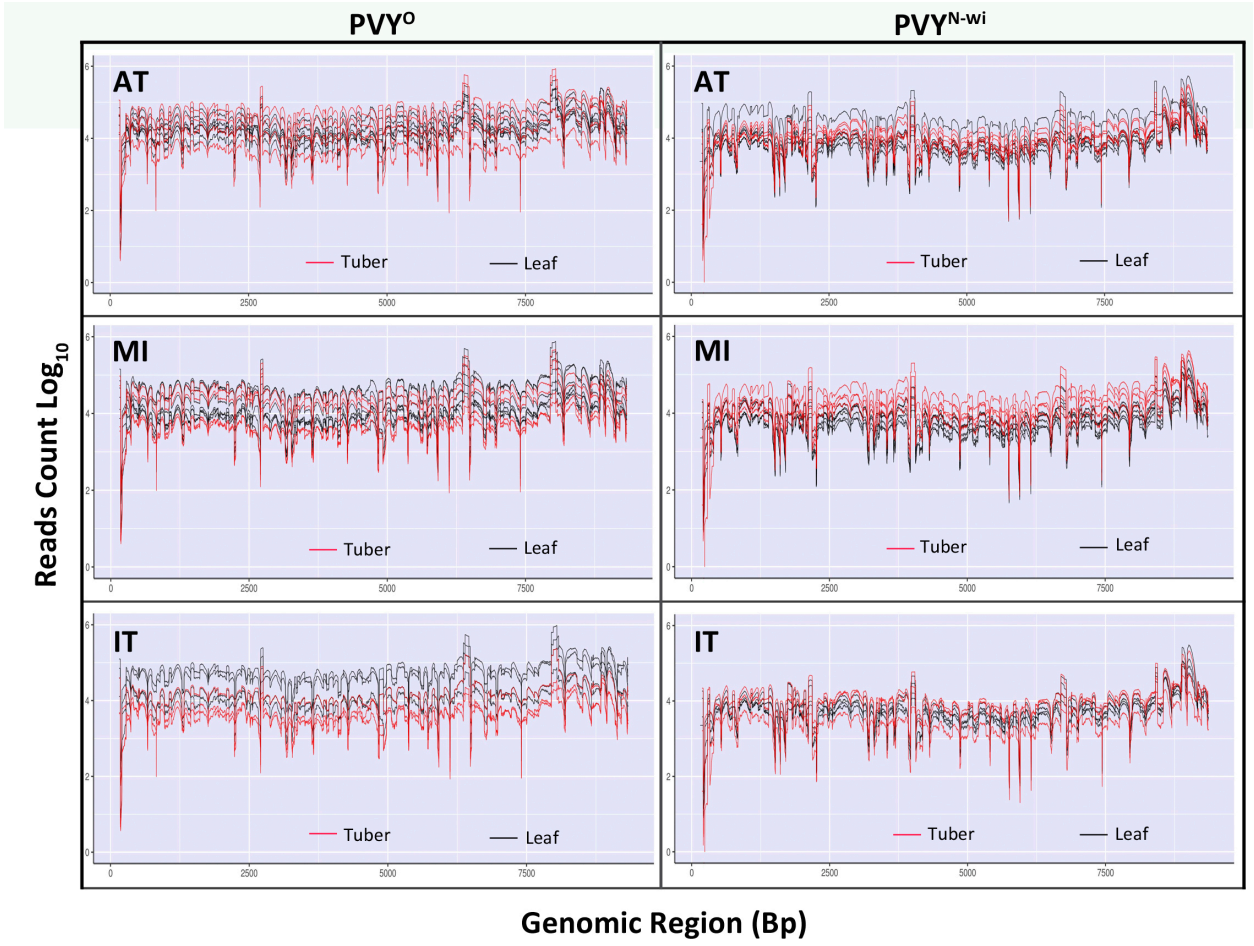


Figure 7



Supplemental Figure S1



Supplemental Figure S2

CONCLUDING REMARKS

PVY has emerged as a major problem to the US potato seed industry and the rise of the recombinant strains is making traditional PVY control methods obsolete. Rogueing based solely on symptom expression is no longer efficient and is contributing to the selection of strains that are asymptomatic in many potato cultivars. A direct solution is to develop detection tools that are more efficient in detecting PVY strains on the farm to facilitate effective removal of infected plants in the fields. However, the PVY epidemic is evolving constantly due to changes in the virus populations and fitness. Fundamental studies to investigate the evolution and fitness of PVY and how these changes contribute to PVY survival will facilitate continued improvement of detection tools and management strategies. We need to constantly improve our technologies to catch up with this fast-evolving virus, if we want to manage it in potatoes

A long term PVY management strategy is to breed for resistance, which hasn't been a priority for potato breeders, even though several molecular markers for PVY resistance genes have successfully been developed in potatoes. The development of molecular markers for PVY symptom expression - e.g. Mosaic, PTNRD, and tuber quality problems – that are reliable and easy to use would be an excellent tool to add to marker-assisted selection programs. Phenotyping potato crosses for reaction to PVY is a specialized and difficult task; therefore, a simple-to-use PVY molecular marker would make it easier for technicians and breeders alike to predict PVY symptom expression in their potato clones.