UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE FITOPATOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM FITOPATOLOGIA

IVAIR JOSÉ DE MORAIS JÚNIOR

Exploring the genomic diversity of potato virus Y and its interaction with hosts

BRASÍLIA – DF
2024

IVAIR JOSÉ DE MORAIS JÚNIOR

Exploring the genomic diversity of potato virus Y and its interaction with hosts

Tese de doutorado apresentada ao Programa de Pós-Graduação em Fitopatologia da Universidade de Brasília, para obtenção do título de Doutor em Fitopatologia.
Orientador: Profa. Dra. Alice Kazuko Inoue Nagata
Coorientador: Prof. Dr. Santiago F. Elena Fito

BRASÍLIA – DF
2024

AGRADECIMENTOS

Nem sempre foi fácil, mas também não foi sempre difícil. Ficaria triste se essa tese fosse medida apenas pelos meus esforços ou pelos resultados obtidos, pois muito do que consegui não teria sido possível sozinho. Muitas pessoas cruzaram meu caminho ao longo desses anos; algumas chegaram e ficaram, trazendo risos e uma companhia que muito apreciei. Outras, por sua vez, tornaram a jornada um pouco mais desafiadora. A todas elas, meu sincero obrigado. Por bem ou por mal, cada uma contribuiu para que eu chegasse até aqui.

Esta tese não foi escrita por uma única mão (ou cérebro), mas pelas mãos de muitas pessoas, fruto dos esforços coletivos. Em especial, agradeço aos meus orientadores. Nunca tive a infelicidade de encontrar um mau orientador; pelo contrário, todos foram excelentes e me ensinaram a enxergar a ciência com entusiasmo.

À Alice Nagata, que não só me orientou, mas também me ajudou a crescer, estando ao meu lado em todos os momentos. Ela sempre sorriu com minhas pequenas conquistas e me impulsionou rumo ao desconhecido e promissor futuro que vivenciei no doutorado.

A Santiago Elena, que em va acollir, acompanyar, orientar i em va posar a la meua disposició tot el que estava al seu abast perquè poguera realitzar un treball excel·lent.

A Paqui de la Iglesia, sense la qual encara estaria fent les meues primeres passes; aquest treball mai hauria estat possible sense la seua ajuda.

Agradeço também, de forma especial, aos meus amigos do Laboratório de Virologia da Embrapa Hortaliças: Dorian, Jonas, Yanca, Bárbara, Gabriela, Valdemir, Malu, AnaLis, Erich e Wandressa. I als companys del Laboratório de Virología Evolutiva y de Sistemas, que em van ajudar en una adaptació que no va ser fàcil.

Aos meus pais, que jamais imaginaram ter um filho Doutor, mas que me apoiaram em todos os meus caminhos, por mais incertos que fossem.

Aos meus amigos Lais e Maike, que me ofereceram suporte emocional nos momentos mais difíceis.

Aos que permaneceram ao meu lado e que amo mais que tudo, meus queridos Cacau e Neve, e aos que se foram, especialmente Ada e Darwin.

Muitos nomes, todos lembrados com gratidão e felicidade.

A todos que me viram chorar e sorrir, enlouquecer e progredir, meu mais profundo agradecimento.

Resumo

O potyvirus potato virus Y (PVY) já foi considerado um dos cinco vírus mais importantes entre os vírus de plantas, devido à sua capacidade de infectar uma ampla gama de hospedeiros, ser transmitido por várias espécies de afídeos e causar prejuízos significativos em culturas de importância agronômica. Sua rápida adaptação a novos ambientes, altas taxas de mutação e recombinação resultam em uma nuvem de mutantes conhecida como vírus *quasispecies*, capazes de se adaptar a condições diversas sobrevivendo no ambiente e se dispersando a novas regiões. Embora no passado alguns programas de melhoramento tenham focado no desenvolvimento de materiais com resistência à infecção por PVY, ainda existe uma grande lacuna de conhecimento sobre a interação do vírus com diferentes hospedeiros e as alterações genômicas resultantes dessa interação.

Para abordar essas questões, iniciamos a análise de isolados de PVY coletados em campos de produção de tomate (PVYSl), batata (PVYSt) e pimentão (PVYCa). Desenvolvemos um protocolo de sequenciamento genômico utilizando a tecnologia Nanopore de modo a sequenciar simultaneamente os genomas de PVYCa, PVYSt e PVYSl, reduzindo significativamente os custos operacionais. Foi também realizada uma avaliação de resistência a infecção por PVY de cultivares de tomate, pimentão e linhagens de *Solanum* spp. do Banco de Germoplasma do Instituto Agronômico de Campinas. Observamos que nenhuma das cultivares de tomate, pimentão e acessos de banco de germoplasma avaliados apresentou resistência à infecção por PVY, indicando a necessidade urgente de busca por materiais com algum nível de resistência para o mercado e para programas de melhoramento.

Para a análise da influência do hospedeiro nas modificações genômicas, um experimento foi realizado com dois isolados virais (PVYNb, coletado em *Nicotiana benthamiana*, e PVYSt) com 10 passagens virais sucessivas por inoculação mecânica em plantas de *N. benthamiana*, tomateiro e batateira. PVYNb e PVYSt mostraram comportamentos distintos: diminuição e extinção da infecção viral em tomateiro, aumento expressivo em *N. benthamiana* e manutenção moderada em batateira. PVYNb apresentou maior especialização com mais SNPs fixados, indicando maior capacidade adaptativa a novos ambientes e hospedeiros, enquanto PVYSt se mostrou mais generalista com menos SNPs fixados. Além disso, investigamos a geração e

manutenção de genomas defectivos virais (GDVs) em diferentes populações de PVY. Foram identificados GDVs nas populações de PVY, cuja produção foi dependente do isolado viral, modo de transmissão, órgão da planta, passagem realizada e hospedeiro. Nossos achados fornecem informações para a elaboração de novas abordagens de recomendações de manejo e controle do PVY, promovendo avanços na sustentabilidade da produção agrícola.

Palavras-chaves: Alteração genômica, Ecologia de vírus, Evolução de vírus, Genomas defectivos virais (GDVs), Interação vírus-hospedeiro, Sequenciamento Nanopore, Suplantação de resistência

Exploring the genomic complexity of potato virus Y and its interaction with hosts

Abstract

The potyvirus potato virus Y (PVY) has been considered one of the five most important plant viruses due to its ability to infect a wide range of hosts, be transmitted by various aphid species, and cause significant damage to crops. Its rapid adaptation to new environments, high mutation and recombination rates result in a cloud of mutants known as *viral quasispecies*, capable of adapting to diverse conditions, surviving in the environment, and spreading to new regions. Although some breeding programs in the past focused on developing materials resistant to PVY infection, there remains a substantial gap in knowledge about the virus' interaction with different hosts and the resulting genomic alterations from this interaction.

To address these issues, we initiated the analysis of PVY isolates collected from tomato (PVYSl), potato (PVYSt), and pepper (PVYCa) production fields. A genomic sequencing protocol was developed using Nanopore technology so we could simultaneously sequence the genomes of PVYCa, PVYSt, and PVYSl, significantly reducing operational costs. We also evaluated the resistance to PVY infection of tomato and pepper cultivars, and *Solanum* spp. lines from the Germplasm Bank of the Instituto Agronômico de Campinas. None of the evaluated tomato, pepper cultivars, and germplasm bank accessions showed resistance to PVY infection, indicating an urgent need to find materials with some level of resistance for the market and breeding programs.

To analyze the host influence on genomic modifications, an experiment was conducted with two viral isolates (PVYNb, collected from *Nicotiana benthamiana*, and PVYSt) with 10 successive viral passages through mechanical inoculation in *N. benthamiana*, tomato, and potato plants. PVYNb and PVYSt exhibited different behaviors, displaying a decrease and extinction of viral infection in tomato plants, a significant increase in *N. benthamiana*, and moderate maintenance in potato plants. PVYNb showed greater specialization with more fixed SNPs, indicating a higher adaptive capacity to new environments and hosts, while PVYSt was more generalist with fewer fixed SNPs. Additionally, we investigated the generation and maintenance of defective viral genomes (DVGs) in different PVY populations. DVGs were identified in PVY

populations, whose production depended on the viral isolate, transmission mode, plant organ, passage performed, and host.
Our findings provide valuable information for developing new management and control recommendations for PVY, promoting advances in the sustainability of agricultural production.

Keywords: Genomic variation, Defective viral genomes (DVGs), Virus ecology, Virus evolution, Virus-host interaction, Nanopore sequencing, Overcoming resistance

Summary

Introduction **1**

References **8**

General objective **12**

Specific objectives: **12**

UNRAVELING THE DYNAMICS OF HOST SPECIFICITY AND RESISTANCE RESPONSES TO POTATO VIRUS Y, AND IMPLICATIONS FOR CROP MANAGEMENT **13**

Abstract **14**

Main text **15**

References **27**

DETECTING AND SEQUENCING THE WHOLE-GENOME OF DISTINCT POTATO VIRUS Y ISOLATES USING A PCR-NANOPORE APPROACH **38**

Abstract **38**

Introduction **40**

Materials and methods **42**

Virus isolates collection 42

Nanopore primer design 43

Amplification of target region with PCR 43

ONT sequencing strategy 44

Sequencing analysis and genome assembly 44

Total RNA purification and Illumina sequencing 45

Genome analysis 45

Design and evaluation of PVY-specific primers 46

Results **46**

Field collection and identification of PVY isolates 46

Nanopore primer design and evaluation 47

27 Nanopore sequencing results 49
28 Illumina sequencing results 54
29 Genome comparison and phylogenetics 54
30 PVY-specificity primers 56

31 **Discussion** **57**

32 **References** **62**

33 **EXPERIMENTAL EVOLUTION OF HOST RANGE FOR TWO ISOLATES OF *POTYVIRUS***
34 ***YITUBEROSI*** **80**

35 **Abstract** **80**

36 **Introduction** **81**

37 **Materials and methods** **83**
38 Plants and growth environment 83
39 Isolates, inoculation and collection 84
40 Primer design and RNA amplification 84
41 Evolution experiment 86
42 Transmission rate experiment 87
43 HTS and sequence analyses 88

44 **Results** **89**
45 Test of transmission efficiency across host species 89
46 Variation of viral loads along the passage through the plants 92
47 Infectivity of evolved lineages depends on both the evolved environment and the test host 95
48 Changes in virulence and symptomatology 96
49 Genome alterations 99

50 **Discussion** **105**

51 **References** **110**

52 ***IN SILICO* EVIDENCES FOR THE PRESENCE OF DEFECTIVE VIRAL GENOMES (DVGS) IN**
53 **POTATO VIRUS Y-INFECTED PLANTS** **123**

54 **Abstract** **123**

55 **Introduction** **125**

56 **Materials and methods** **127**
57 Datasets 127
58 DVGs identification 128
59 DVG analysis 129

60 **Results** **129**
61 DVGs profile 129
62 Clustering DVGs 134
63 DVGs formation and distribution 136
64 DVGs diversity 139

65 **Discussion** **141**

66 **References** **145**

67 **Concluding remarks** **162**
68
69

Introduction

Since Virology emerged as a scientific field in the late 19th and early 20th centuries (Burrell et al. 2017), our understanding of viruses as pathogens affecting all kinds of organisms has evolved dramatically. Early studies relied on physical, biological, and electron microscopy techniques (Zuo et al. 2024). Today, methods such as genetic engineering and deep sequencing have accelerated the advances in the field. For example, we can now edit genomes, gain insights into viral infection dynamics and host responses through single-cell sequencing, and visualize high-resolution 3D structures of viral particles using Cryo-EM. Technology facilitated the development of antiviral drugs targeting various stages of the viral infection cycle. Recombinant DNA technology and mRNA vaccine platforms have further helped vaccine development, as evidenced by the rapid creation of COVID-19 vaccines (Karikó et al. 2005; Baden et al. 2021; Polack et al. 2021).

While initial virus research primarily focused on plants, the significant impact of viruses on human health often overshadows plant virology. However, plant viruses may also cause huge damages, exemplified by important diseases such as tobacco mosaic virus (TMV) in tobacco (Chen et al. 2014), potato virus Y (PVY) in potatoes (Nolte et al. 2004), African cassava mosaic virus (ACMV) in cassava (Legg et al. 2011), barley yellow dwarf virus (BYDV) in cereals (Choudhury et al. 2018), rice tungro virus (RTV) in rice (Hibino et al. 1991), tomato yellow leaf curl virus (TYLCV) in tomatoes (Papayiannis et al. 2011), banana bunchy top virus (BBTV) in bananas (Dale 1987), plum pox virus (PPV) in stone fruits (Németh 1994), and papaya ringspot virus (PRSV) in papaya (Jain et al. 2004). These viruses not only reduce crop quality and yield but also impact food security and livelihoods in affected regions.

To mitigate these impacts, we must implement strategies such as enhanced surveillance and implement diagnostics, quarantine and sanitation measures, breeding for resistance, and integrated pest management (Strange and Scott 2005). Yet, there are numerous questions missing of appropriate answers. We are within a small bubble of knowledge that, despite recent advances, still holds many mysteries. Our innate curiosity drives us to explore and seek for answers to many questions. Numerous questions remain unanswered in Virology: What is the exact origin of viruses? How do

they evolve so rapidly? Why do some viruses cause severe diseases while others do not? What determines their host range? How do viruses manipulate host cellular machinery so effectively? What drives the emergence and re-emergence of viral diseases? How do viruses cross species barriers enabling them to infect new hosts? And what unknown viruses live in unexplored habitats?

Driven by these questions, our research aims to fill gaps in the overall comprehension of the virus genome and host range, starting from collection of virus isolates in commercial fields for detailed analysis of genome alterations during host-virus interactions. Our work, conducted over four years, focused on potato virus Y (PVY; species *Potyvirus yituberosi*, genus *Potyvirus*, family *Potyviridae*), a pathogen often associated with substantial crop losses (Kerlan et al. 2008) having caused significant impacts in the past on tomato and pepper crops in Brazil. But since the 1960s, concerns about PVY in these crops have diminished due to the development of resistant cultivars. As a result, a few studies have been conducted on PVY in crops other than its primary host, the potato.

Looking the database available in GenBank, out of 585 PVY genomes available, only 18 viruses were isolated from tomatoes (*Solanum lycopersicum*), six from peppers (*Capsicum annuum*), while the isolates from potatoes (*Solanum tuberosum*) add up to 466 items of this list. Thus, it is evident that studies on potatoes have been prioritized. The other isolates on the list include those collected from *Capsicum baccatum* (n = 1), *Datura metel* (n = 1), *Nicotiana tabacum* (n = 67), *Physalis peruviana* (n = 7), *S. americanum* (n = 1), *S. bataceum* (n = 7), *S. nigrum* (n = 7), *S. phureja* (n = 1), *S. quitoenses* (n = 1), *S. sisymbriifolium* (n = 1), and *Curcubita pepo* (n = 1).

In a way, it is understandable that numerous studies have been conducted on potatoes, since currently there are no cultivars with strong resistance to PVY, and it can significantly reduce production both qualitatively (Nolte et al. 2004) and quantitatively (Beczner et al. 1984). However, despite the predominance of large-scale agricultural systems for potato production, other vegetable crops cultivated nearby may act as PVY reservoirs for potato plants.

In recent years, during field trips of our group, necrotic symptoms were observed in tomato plants. The disease is known as "Mexican fire" disease. This symptom was then demonstrated to be associated to PVY infection (Lucena et al. 2024). The most

concerning aspect is that this symptom is increasingly being found in tomato plants, which could represent a threat to tomato cultivation. This has raised an alert about the potential risks that PVY could pose to these crops which are often grown by small-scale and low-income producers.

Due to this, we sought to fill the gap left by breeding programs by testing widely used tomato and pepper cultivars in production fields. In a preliminary discriminant analysis of principal components (DAPC) using all available genomes in GenBank that have the annotation of the host from which the isolate was collected (n = 492), we observed that the host in which the PVY was collected have influenced the clustering (Fig 1.). Isolates from peppers and tomatoes tend to be in the same group, separated from the potato isolates. This genetic differentiation highlighted by DAPC prompted us to question the role and forces that the host might exert on distinct PVY populations.

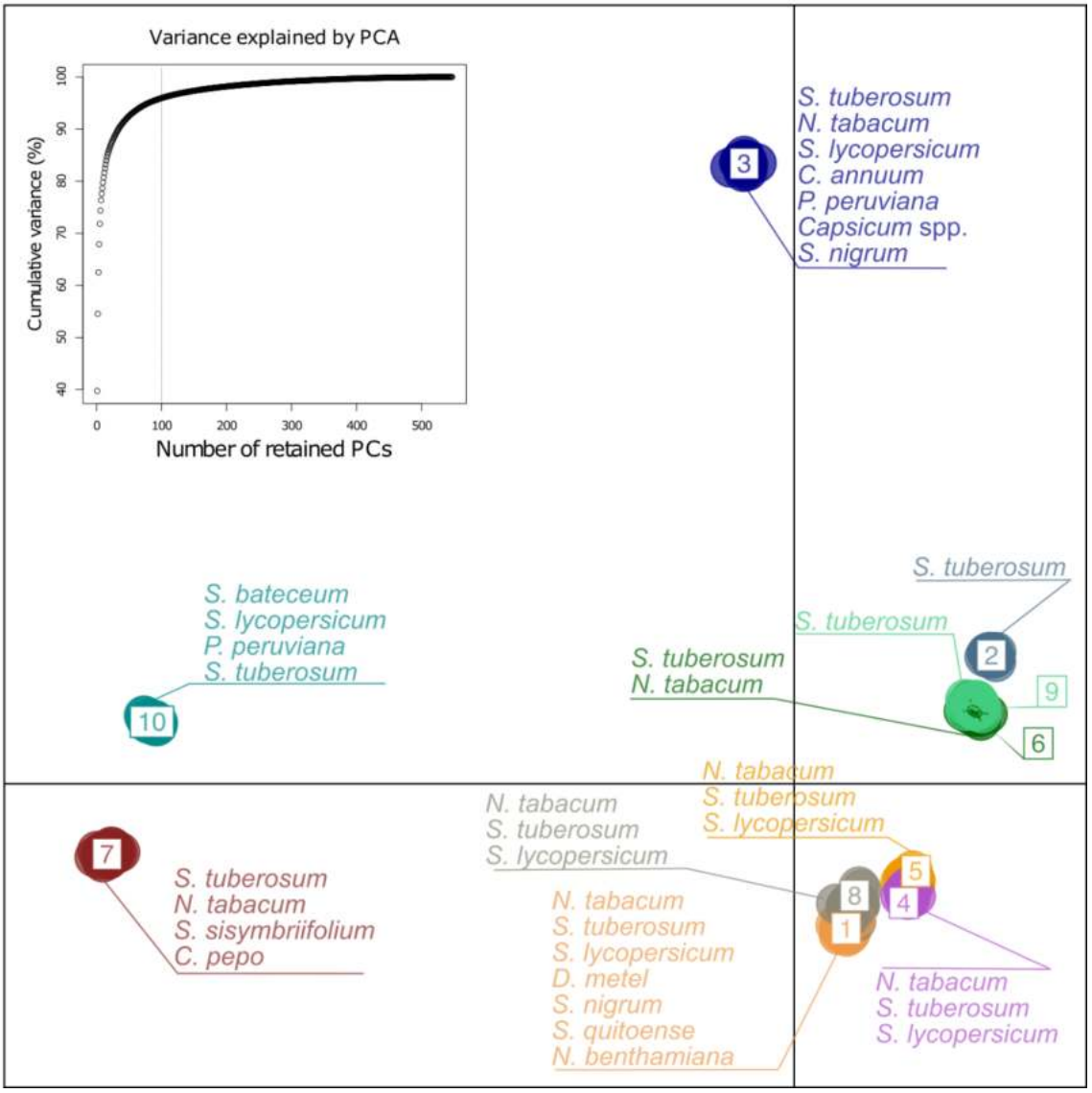


Fig 1. Scatter plot of Discriminant Analysis of Principal Components (DAPC) based on 492 potato virus Y (PVY) genomes obtained from GenBank. Only isolates with known host information were included in the analysis. The dataset was divided into 10 distinct subpopulations, each represented by a different color. Arrows within each cluster point to the specific host species from which the isolates were collected.

From this, we hypothesized that the host could alter the evolutionary course of PVY populations. Understanding that isolates collected from different crops may be distinct and that these differences cannot be detected by tools such as ELISA or RT-PCR, one of our objectives was to develop a quick, easy, and affordable way to sequence the entire genome of isolates using Nanopore sequencing.

Still trying to understand how the virus-host interaction works, we hypothesized that during the passage of the virus through different hosts, these hosts would modulate the composition and fitness of the viral population, leading to viral specialization and fixation of few haplotypes, or the contrary, an expansion of the genetic variability associated to generalism. In pilot trials, we observed that some isolates collected from different crops were not able to infect certain plant species, which could be related to genomic changes or selection mechanisms. From this, we decided to study how the adaptation of two PVY isolates collected from two different plant species evolve in distinct hosts.

Indeed, due to the wide host range of PVY, the virus faces significant tradeoffs in fitness when infecting different hosts (Elena et al. 2014). For instance, adaptations that enhance PVY fitness in one host, such as the evolution of resistance-breaking strains in potato (*Solanum tuberosum*), may come at the expense of reduced fitness when the virus infects other solanaceous hosts like tomato (*Solanum lycopersicum*).This phenomenon underscores the importance of host-specific adaptations in PVY evolution. PVY classification has often been linked to symptomatology and infectivity in different hosts, suggesting a correlation between viral isolates and host range properties (Quenouille et al. 2013). For example, some PVY isolates that are infectious to potato tend to be poorly infectious to pepper (Romero et al. 2001), while isolates from Chile show limited infectivity to potatoes and are predominantly restricted to that region (Moury 2010).

The adaptability of PVY is further enhanced by its RNA-dependent RNA polymerase (RdRp), which is particularly error-prone, leading to high mutation rates (Drake 1993; Sanjuán 2012). These high mutation rates are among the highest observed in nature and contribute significantly to the rapid evolution and adaptability of RNA viruses (Sanjuán and Domingo-Calap 2016), enabling them to swiftly evade host immune responses and develop resistance to antiviral treatments. Moreover, RNA viruses, including PVY, mutate more rapidly than DNA viruses (Drake et al. 1998). Single-stranded RNA viruses exhibit higher mutation rates than their double-stranded counterparts, and there is a negative correlation between genome size and mutation rate. This suggests that viral genetic diversity is influenced by both virus- and host-dependent factors and evolves in response to selective pressures (Sanjuán and Domingo-Calap 2016).

Recombination also plays a critical role in the evolutionary process of PVY. Recombination is responsible of generating variability that can drive adaptation or even the emergence of new species (Padidam et al. 1999; Inoue-Nagata et al. 2006; Fiallo-Olivé et al. 2019; Lal et al. 2022). In RNA viruses, the recombination occurs when the RdRp associated with a nascent transcript dissociates from one template and associates with another (Kirkegaard and Baltimore 1986), Together with mutation, these evolutionary parameters, influenced by past selection, work to maintain a mutation-selection balance, an equilibrium where the population remains resilient against deleterious mutations. This balance, shaped by the interplay of limited genetic capacity, high mutation rates, and population size dynamics, forges a close relationship between the biology of RNA viruses like PVY and their evolutionary dynamics (Dolan et al. 2018).

In large viral populations like those of PVY, the diversity generated by these high mutation rates results in a network of mutant genotypes surrounding a dominant sequence. This network enables various interactions among the viral variants, such as antagonism, cooperation, and recombination. During transmission bottlenecks, the phenomenon of en bloc transmission helps preserve population size and diversity, facilitating coinfection and mitigating the effects of genetic drift (Dolan et al. 2018).

As a result, PVY populations are constantly generating mutants with varying levels of infectivity. The high mutation rates in these populations regularly give rise to such variants, each with its own potential impact on the virus's overall fitness (Holland et al. 1982). The concept of quasispecies, first proposed in the 1970s for bacteriophage Qβ replicating in *Escherichia coli*, is crucial for understanding this diversity. It refers to a model where a viral population exists not as a single, homogeneous entity but as a cloud of genetically diverse variants centered around a consensus or "master" sequence (Eigen 1971; Eigen and Schuster 1977). Within this quasispecies cloud, individual variants can interact through mechanisms like recombination, complementation, and selection, enabling the viral population to maintain its fitness despite the accumulation of deleterious mutations. Although initially described for bacteriophages, the quasispecies concept has since become fundamental to understanding RNA virus evolution and adaptability.

Indeed, the selection pressure have been shaping the PVY population from its emergence in the Andes of South America to its introduction in Europe in the 16th century (Torrance and Talianksy 2020), making PVY an important viruses in agricultural systems. The fitness landscape for PVY is thus intricately shaped by its interactions with different plant hosts, mutation rates, and the genetic diversity of the viral population. As a quasispecies, PVY comprises numerous coexisting variants within a population. These variants explore the fitness landscape, with some mutations leading to increased fitness. The rugged nature of this fitness landscape allows PVY to rapidly adapt to new environmental conditions, such as shifts in host species or changes in agricultural practices.

Based on all that we know about viral variability, we became curious to understand other factors that might be involved in the success or failure of the virus-host interaction, such as the effect of defective genomes (DVGs) in the wild-type virus replication process. For potato cultivation, tubers are usually used, which are often propagated over several generations. Our hypothesis is that DVGs are pervasively generated in PVY populations, particularly in potatoes that are vegetatively propagated, considering that they have fewer selective bottlenecks compared to those transmitted by aphid vectors. Therefore, we sought to identify the DVGs component of PVY populations maintained under different transmission protocols. Specifically, we examined populations isolated from potato tubers or leaves, inoculated via aphid vector or mechanically, maintained in different hosts, and after consecutive viral passages.

This thesis has been divided into four chapters, all centered on PVY and its interaction with different hosts. In the first chapter, we performed a screening for PVY resistance in commercial tomato and pepper cultivars, and also in tomato accessions used in breeding programs at the Instituto Agronômico de Campinas (IAC). In the second chapter, we developed a methodology for PVY genome sequencing using Nanopore MinION. In the third chapter, we performed a viral passage evolution experiment with different hosts and two different viral isolates. Finally, in the fourth chapter, we evaluated the emergence and maintenance of defective viral genomes (DVGs) in different viral populations and their potential impact on PVY population modulation.

Together, our experiments aimed to deepen our understanding of PVY genetic diversity and adaptability in different host environments. By leveraging advanced sequencing technologies and rigorous experimental approaches, we seek to unravel the complexities of virus-host interactions, uncovering the mechanisms that drive viral evolution and specialization. In summary, in this thesis we managed to: (i) identify host preferences and genetic variability by screening a range of tomato and pepper cultivars, we determined the host preferences of different PVY isolates and explored the genetic variability among these isolates; (ii) develop rapid genome sequencing methods establishing a streamlined, cost-effective methodology for sequencing PVY genomes using Nanopore MinION technology, facilitating rapid and accurate genomic analysis; (iii) investigate viral replication dynamics through viral passage experiments, we evaluated how different hosts influence PVY replication rates and the genetic bottlenecks that shape viral populations; and (iv) examine the role of defective viral genomes (DVGs) studying the emergence and impact of DVGs on PVY populations, particularly in relation to their role in modulating viral replication and host adaptation.

References

Baden LR, Sahly HME, Essink B, et al (2021) Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. New England Journal of Medicine, 384, 5, 403-416. http://dx.doi.org/10.1056/nejmoa2035389.

Beczner L, Horváth J, Romhányi I, et al (1984) Studies on the etiology of tuber necrotic ringspot disease in potato. Potato Research, 27, 339–352. https://doi.org/10.1007/BF02357646

Burrell CJ, Howard CR, Murphy FA (2017) History and impact of virology. In: Fenner and White's Medical Virology. Elsevier, pp 3–14. https://doi.org/10.1016/B978-0-12-375156-0.00001-1

Chen W, Liu W, Jiao H, et al (2014) Development of a concentration method for detection of tobacco mosaic virus in irrigation water. Virologica Sinica 29:155-161. https://doi.org/ 10.1007/s12250-014-3461-7

Choudhury S, Larkin P, Meinke H et al. (2018) *Barley yellow dwarf virus* infection affects physiology, morphology, grain yield and flour pasting properties of wheat. Crop & Pasture Science 70:16-25. https://doi.org/10.1071/CP18364

Dale JL (1987) Banana bunchy top: an economically important tropical plant virus disease. Advances in virus research 33:301-325. https://doi.org/10.1016/s0065-3527(08)60321-8

Dolan PT, Whitfield ZJ, Andino R (2018) Mechanisms and concepts in RNA virus population dynamics and evolution. Annual Review of Virology 5:69–92. https://doi.org/10.1146/annurev-virology-101416-041718

Drake JW (1993) Rates of spontaneous mutation among RNA viruses. Proceedings of the National Academy of Sciences 90:4171–4175. https://doi.org/10.1073/pnas.90.9.4171

Drake JW, Charlesworth B, Charlesworth D, Crow JF (1998) Rates of spontaneous mutation. Genetics, 148:1667-86. https:// 10.1093/genetics/148.4.1667

Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58:465–523. https://doi.org/10.1007/BF00623322

Eigen M, Schuster P (1977) A principle of natural self-organization. Naturwissenschaften 64:541–565. https://doi.org/10.1007/BF00450633

Elena SF, Fraile A, García-Arenal F (2014) Evolution and emergence of plant viruses. Advances in Virus Research 88:161-191. https://doi.org/ 10.1016/B978-0-12-800098-4.00003-9

Fiallo-Olivé E, Trenado HP, Louro D, Navas-Castillo J (2019) Recurrent speciation of a tomato yellow leaf curl geminivirus in Portugal by recombination. Scientific Reports 9:1332. https://doi.org/10.1038/s41598-018-37971-z

Hibino H, Ishikawa K, Omura T (1991) Characterization of rice tungro bacilliform and rice tungro spherical viruses. Phytopathology 81:1130-1132.

Holland J, Spindler K, Horodyski F, et al (1982) Rapid evolution of RNA genomes. Science (1979) 215:1577–1585. https://doi.org/10.1126/science.7041255

Inoue-Nagata AK, Martin DP, Boiteux LS, et al (2006) New species emergence via recombination among isolates of the Brazilian tomato infecting Begomovirus complex. Pesquisa Agropecuária Brasileira 41:1329–1332. https://doi.org/10.1590/S0100-204X2006000800018

Jain RK, Nasiruddin KM, Sharma KM, et al (2004) First report of occurrence of Papaya ring spot virus infecting Papaya in Bangladesh. Plant disease 88:221. https://doi.org/10.1094/PDIS.2004.88.2.221C

Karikó K, Buckstein M, Ni H, Weissman D (2005) Suppression of RNA recognition by toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity 23:165–175. https://doi.org/10.1016/j.immuni.2005.06.008

Kerlan C, Moury B (2008) Potato virus Y. Encyclopedia of Virology, 4, 287-296.

Kirkegaard K, Baltimore D (1986) The mechanism of RNA recombination in poliovirus. Cell 47:433–443. https://doi.org/10.1016/0092-8674(86)90600-8

Lal A, Kil E-J, Vo TTB, et al (2022) Interspecies recombination-led speciation of a novel geminivirus in Pakistan. Viruses 14:2166. https://doi.org/10.3390/v14102166

Legg JP, Jerimiah SC, Obiero HM, et al (2011) Comparing the regional epidemiology of the cassava mosaic and cassava brown streak virus pandemics in Africa. Virus Research 159:161-170. https://doi.org/10.1016/j.virusres.2011.04.018

Lucena VS, Nakasu EYT, Pereira JL, et al (2024) Emergence of potato virus Y outbreaks in tomatoes in Brazil, the disease and spread. bioRxiv. https://doi.org/10.1101/2024.05.17.594728

Moury B (2010) A new lineage sheds light on the evolutionary history of Potato virus Y. Molecular Plant Pathology 11:161–168. https://doi.org/10.1111/j.1364-3703.2009.00573.x

Neméth M (1994) History and importance of plum pox in stone-fruit production. Bulletin OEPP EPPO Bulletin 24:525-536. https://doi.org/10.1111/j.1365-2338.1994.tb01065.x

Nolte P, Whitworth JL, Thornton MK, McIntosh CS (2004). Effect of seedborne potato virus y on performance of russet burbank, russet norkotah, and shepody potato. Plant Disease, 88 (3), 248-252. https://doi.org/10.1094/PDIS.2004.88.3.248

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265:218–225. https://doi.org/10.1006/viro.1999.0056

Quenouille J, Vassilakos N, 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. https://doi.org/10.1111/mpp.12024

Romero A, Blanco-Urgoiti B, Soto MJ, et al (2001) Characterization of typical pepper-isolates of PVY reveals multiple pathotypes within a single genetic strain. Virus Research 79:71–80. https://doi.org/10.1016/S0168-1702(01)00300-8

Sanjuán R (2012) From molecular genetics to phylodynamics: evolutionary relevance of mutation rates across viruses. PLoS Pathogens 8:e1002685. https://doi.org/10.1371/journal.ppat.1002685

Sanjuán R, Domingo-Calap P (2016) Mechanisms of viral mutation. Cellular and Molecular Life Sciences 73:4433–4448. https://doi.org/10.1007/s00018-016-2299-6

Strange RN, Scott PR (2005) Plant Disease: a threat to global food security. Annual Review of Phytopathology 43:83–116. https://doi.org/10.1146/annurev.phyto.43.113004.133839

Papayiannis LC, Katis NI, Idris AM, Brown JK (2011) Identification of weed hosts of tomato yellow leaf curl virus in Cyprus. Plant Disease 95:120-125. https://doi.org/10.1094/PDSI-05-10-0346

Polack FP, Thomas SJ, Kitchin N, et al (2020) Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. New England Journal of Medicine, 383, 27, 2603-2615. http://doi.org/10.1056/nejmoa2034577

Torrance L, Talianksy ME (2020) Potato Virus Y Emergence and evolution from the Andes of South America to become a major destructive pathogen of potato and other solanaceous crops worldwide. Viruses 12. https://doi.org/10.3390/V12121430

Zuo K, Gao W, Wu Z, et al (2024) Evolution of virology: science history through milestones and technological advancements. Viruses 16:374. https://doi.org/10.3390/v16030374

General objective

Study the effects of the host on the evolution of PVY by analyzing aspects of infectivity, host specificity, transmissibility, and genome variation.

Specific objectives:

Chapter I

1. Assess the infectivity of PVY isolates across distinct host species.
2. Evaluate the resistance of commercial tomato and pepper cultivars, along with germplasm bank accessions, against PVY infection.

Chapter II

3. Design and validate specific primers for the detection of PVY using RT-PCR.
4. Develop a cost-effective sequencing protocol utilizing the Nanopore technology.

Chapter III

5. Monitor and quantify the infection ability of two PVY isolates across multiple host combinations in a viral passage experiment.
6. Investigate the genomic alterations on PVY isolates as they infect different hosts at selected passage points.
7. Evaluate the impact of distinct PVY isolates on the phenotype of infected plants during successive viral passages through the hosts.
8. Determine the role of host species in driving the generation of viral genomic variability.

Chapter IV

9. Identify and characterize the formation of defective viral genomes (DVGs) using data from Chapter III and those available in databases.
10. Detect the generation of DVGs according to the interaction of the PVY strains/isolates with diverse hosts, with different transmission modes and plant organs, and after mechanically passaging under distinct experimental conditions.

Chapter I

Unraveling the dynamics of host specificity and resistance responses to potato virus Y, and implications for crop management

Ivair José de Morais[1,2], Dorian Yest Melo Silva[1,2], Barbara Mavie Camargo[1], André Luiz Lourenção[3], Alice Kazuko Inoue-Nagata[2]

[1]Departamento de Fitopatologia, Instituto de Ciências Biológicas, Universidade de Brasília (UnB), 70910-900, Brasília, Distrito Federal, Brazil

[2]Embrapa Hortaliças, 70351-970, Brasília, Distrito Federal, Brazil

[3]Departamento de Entomologia e Acarologia, Escola Superior de Agricultura "Luiz de Queiroz" (ESALQ), Universidade de São Paulo, 13418-900, Piracicaba, São Paulo, Brazil

Corresponding author: Alice Kazuko Inoue-Nagata

e-mail: alice.nagata@embrapa.br

Submitted to Tropical Plant Pathology

Abstract

Potato virus Y (PVY), a virus member of the family *Potyviridae*, poses a significant threat to global agriculture, affecting crops such as potato, tomato, pepper, and tobacco. Despite its economic importance, there remains a critical gap in understanding the dynamics of PVY-host interactions and the development of effective management strategies. This study aimed to comprehensively characterize PVY isolates from sweet pepper, potato, and tomato plants, elucidating their infectivity and adaptation across diverse host species and cultivars. Initially, using antigen-trapped ELISA, we determined the optimal detection timeframe and leaf sampling strategy for detection of PVY by serological assays, showing that some hosts require a minimum incubation period and leaf selection for a reliable virus detection. By comparing PVY isolates from distinct hosts, we demonstrated that the choice of the isolate is crucial for resistance evaluations. Additionally, inoculation trials across various plant species elucidated differences in infectivity and adaptation among PVY isolates. Resistance trials in commercial cultivars of tomato and pepper plants and wild *Solanum* spp. accessions revealed susceptibility across all tested materials, challenging previous assumptions of resistant cultivars and accessions. These findings underscore the urgency of addressing PVY spread and understanding host-virus interactions to identify resistant genotypes for commercial use and for developing breeding programs directed to PVY isolates present in Brazil.

Keywords: host range, plant breeding, *Potyviridae*, *Potyvirus*, resistance screening, viral adaptation

Main text

According to the latest update from the International Committee on Taxonomy of Viruses, potato Y virus (PVY) is classified as species *Potyvirus yituberosi* (genus *Potyvirus*, family *Potyviridae*). It possesses a positive single-stranded RNA genome of approximately 9.7 kb in length, encoding 11 mature proteins (Inoue-Nagata et al. 2022). Ten proteins, P1, HC-Pro, P3, 6K1, CI (cylindrical inclusion), 6K2, NIa, VPg, NIb (viral polymerase), and CP (capsid protein), are derived from the cleavage of a larger polyprotein by viral proteases. One protein, PIPO, is generated by a polymerase slippage mechanism and is expressed as the trans-frame protein P3N-PIPO. PVY stands as a serious viral threat in global agriculture, affecting crops such as potato, tomato, pepper, and tobacco (Quenouille et al. 2013). In fact, PVY has been considered a major threat to global potato production due to its high prevalence and ability to rapidly spread through fields (Karasev and Gray 2013). Its detrimental impact on crop yield is also relevant in tomato and pepper crops, underscoring the necessity for comprehensive research to identify resistant cultivars amidst its high prevalence and rapid spread in fields (Karasev and Gray 2013). Despite its importance, the current tomato portfolio of cultivars lacks a comprehensive description of resistance against PVY, thereby requiring further investigation.

Studies have revealed the substantial economic losses PVY can induce, with sweet pepper crops experiencing yield reductions ranging from 20 to 70% upon infection, particularly severe during early stages (Avilla et al. 1997). While the exact economic impact on tomato crops remains unquantified, its significant effects are well-documented (Quenouille et al. 2013). Thus, PVY remains a relevant concern to agriculture, threatening both yield and economic stability.

Historically, PVY posed a significant threat to Brazilian agriculture during the 1960s and 1970s. However, the development of resistant tomato cultivars, such as those in the Ângela group, and hybrid peppers has substantially mitigated its impact (Nagai and Costa 1969; Nagai 1971). The rare reports of PVY occurrence in Brazilian tomato and pepper fields further diminished its economic significance in these crops (Meissner et al., 1990). Yet, recent observations suggesting a new disease named “Mexican Fire” in plants infected with PVY, highlight the resurgence of PVY in tomato fields,

underscoring the potential re-emergence of this virus as a serious threat in Brazil (Lucena et al. 2024).

PVY is a generalist virus and exhibits a broad host range, experimentally infecting over 400 species across 30 families (Edwardson and Christie 1997; Jeffries 1998) and understanding the host range of viruses is crucial for virus diagnosis (Dijkstra 1992; McLeish et al. 2019). However, the determinants of host range in plant virus genomes and their implications for virus fitness and pathogenicity remain largely unknown. Despite this, it is known that the inability of a virus to infect a particular plant host may arise from various factors, including the failure to complete essential steps of the infection cycle, such as replication or systemic movement, or the presence of active and specific resistance mechanisms within the plant (Kang et al. 2005). Additionally, host range expansion is a common phenomenon among plant viruses, often at the cost of reduced fitness in the original host (Agudelo-Romero and Elena 2008; Bedhomme et al. 2012; García-Arenal and Fraile 2013). Furthermore, after serial passages in a specific host, the infectivity in the original host can diminish, suggesting potential constraints on a virus adapted to one host's ability to infect another one within its host range (Yarwood 1979). This implies that a virus adapted to one host may not necessarily be able to infect another host within its host range.

Nevertheless, even among generalist viruses, significant host-virus associations exist, with host specialization emerging as a successful strategy for increased prevalence (Malpica et al. 2006). Such specialization often involves genetic changes within the virus genome, potentially leading to alterations in host range. Additionally, host jumping and adaptation within plant species are not sporadic events in plant virus evolution but rather significant drivers of viral emergence (Vassilakos et al. 2016). These events carry epidemiological consequences, impacting viral survival and spread. Therefore, elucidating virus-host interactions holds immediate implications for control measures.

PVY exists as a complex of strains, delineated based on host range, serological properties and molecular characteristics (Singh et al. 2008). These strains are generally classified as PVY^{C}, PVY^{O}, and PVY^{N}. Studies investigating different PVY strains have revealed exceptional nucleotide diversification through mutation and/or recombination, enabling adaptation to new cultivars or diverse environments and resulting in varying

degrees of infectivity (Karasev and Gray 2013; Nigam et al. 2019). PVY^O and PVY^N predominantly comprise potato isolates, which are less adept at infecting peppers, while PVY^C primarily consists of pepper isolates with limited adaptation to potato (Moury 2010). However, it is noteworthy that the PVY^C clade also includes potato-infecting isolates (Dullemans et al. 2011). Interestingly, in tomato fields, a PVY^C isolate from commercial tomato production was grouped within the same clade as potato-infecting isolates but exhibited an inability to infect potatoes (Chikh-Ali et al. 2016). In addition to the C, O, and N strains, a large number of recombinants can be easily found, particularly in potato production fields, where they are more prevalent than non-recombinant strains (Galvino-Costa et al. 2012; Karasev et al. 2011). This prevalence poses a challenge for developing PVY-resistant potatoes, as there are currently no resistant cultivars available.

Phylogenetically, the host species appears to significantly influence the distribution of PVY, as evidenced by studies demonstrating differential infectivity among isolates across hosts (Cuevas et al. 2012). This effect becomes apparent when certain isolates successfully infect one host while failing to do so in others (Green et al. 2017).

Therefore, our study aims to address basic concepts of virus-hosts interaction at a mechanically inoculation and detection level, filling this gap in knowledge by understanding (1) the dynamics between three PVY isolates, identified in three distinct host species, and (2) the capacity to infect its original host and other hosts. We also consider the recent increase in incidence of PVY in tomato crops (Lucena et al. 2024) and search for resistant commercial sweet pepper and tomato cultivars, alongside wild solanum lines utilized in breeding programs. Our findings yield valuable insights that can contribute to breeding programs and help understanding the intricate dynamics of PVY-host interactions.

First, we used three PVY isolates collected from different hosts: PVYCa collected from a sweet pepper (*Capsicum annuum*) plant, PVYSt from a potato (*Solanum tuberosum*) plant and PVYSl from tomato (*Solanum lycopersicum*), all of them from the district of PAD-DF, close to Brasília, the Federal District in Brazil. Seeds were sown in polystyrene trays containing 128 cells and subsequently transplanted to 500 mL pots, containing organic potting mix and substrate (1:1 ratio), and kept in a greenhouse.

For all trials, the detection was done by antigen-trapped ELISA in nitrocellulose membranes (dot-ELISA) using a house made polyclonal antibody at a concentration of 1 µg/mL, as described in Nagata et al. (1995). This antibody detects both PVY^{O} and PVY^{N} strains (Inoue-Nagata et al., 2001; Fonseca et al., 2005). The crude sap diluted in 0.5x PBS of each sample was applied on a nitrocellulose filter and treated with 1 µg/mL anti-PVY after blocking with skimmed milk, and later with anti-rabbit IgG alkaline phosphatase-conjugated antibody produced in goat (Sigma-Aldrich), diluted 1:30,000. Samples were considered positive if a purple color developed after incubation with a solution with nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP) by visual inspection.

In a pilot test, the detection of PVY in sweet pepper cv. Ikeda, our model cultivar, by serology proved to be challenging due low level of detection in early post-inoculation stages (*data not shown*). Due to this, the optimal time for inoculum collection was determined by testing the second and third leaves of plants 3, 5, 7, 9, 11, and 13 days post-inoculation (d.p.i.) of cv. Ikeda, using 10 plants each. Our aim was to determine which leaves, and the minimal time to collect samples to avoid false negative results. The inoculation was done using leaves of infected plants ground (~1:10) in 0.05 M phosphate buffer, pH 7.0, in plants with 2-4 true leaves. At this stage, we used the PVYCa and PVYSt isolates due to their ability to infect pepper plants (*data not shown*).

The serological test demonstrated that PVY remained undetectable until 13 d.p.i. under the tested conditions, regardless of the PVY isolate. This implies that the virus remains below detection levels in the plant until at least 11 d.p.i. Notably, no infections were observed until 11 d.p.i., with positive detections emerging only two days later (Sup. Fig. 1). While the dot-ELISA method is commonly employed due to its cost-effectiveness and suitability as an initial screener for a large number of plants, our results suggest that PVY detection is only reliably possible after at least 13 d.p.i., indicating a narrow window for serological detection within this timeframe considering the sweet pepper cultivar Ikeda. Consequently, screening plants for PVY during the early stages of infection may yield false negative results, as the virus may be present in the field but remain undetectable at these early stages.

In serological tests, a single leaf, preferably the youngest, is typically collected for detection. We conducted experiments to determine which of the younger leaves is most

suitable for the detection test. For PVYCa, the virus was detected in the second youngest leaf in 4 out of 5 inoculated plants, and in the third leaf in 2 out of the same 5 plants. For PVYSl, 4 positives out of 5 were detected in the second leaf, while 1 out of 5 were detected using the third leaf (Sup. Fig. 1). The detection test was performed at 13 d.p.i. In conclusion, our findings suggest that for the detection of PVY in sweet pepper plants using dot-ELISA, testing should be conducted at least 13 d.p.i., preferably using the second youngest leaf. Note that our experiments were exclusively conducted with Ikeda peppers, as detection in tomato and potato cultivars posed no challenges during previous laboratory tests (*data not shown*). Therefore, all PVY detections in our experiments were performed with at least 13 d.p.i. and using the second youngest leaf.

To investigate whether the host from which the PVY isolate originated influences resistance responses, PVYCa, PVYSt and PVYSl were used for inoculation of 27-30 plants of sweet pepper cv. Ikeda, potato cv. Atlantic and tomato cv. Santa Clara.

Sweet pepper plants were infected with PVYCa (8 positives out of 27, Infection Rate (IR) of 30%) and PVYSl (6/29, IR=21%), but not with PVYSt. Tomato plants were infected by all isolates: PVYCa (11/30, IR=37%), PVYSt (18/30, IR=60%) and PVYSl (20/30, IR=67%). Potato plants were infected by PVYSt (17/28, IR=61%), but neither PVYCa nor PVYSl infected them. This suggested a strong specificity of the isolates to the hosts (Gebre Selassie et al. 1985; Fereres et al. 1993; Romero et al. 2001; Moury 2010). None of the combinations yielded a 100% IR. Interestingly, PVYCa was unable to establish infection in potato plants, while PVYSt failed to infect sweet pepper plants, indicating a clear distinct interaction between these two viruses and hosts. Actually, the responses of pepper and tomato plants against the inoculation of PVYCa and PVYSl were similar, and clearly differed from the ones of PVYSt.

A Generalized Linear Model (GLM) with a binomial distribution was fitted to assess the interaction effects of species, virus isolate, and host on the infection proportion. The significance of the model coefficients was evaluated to determine the effect of each factor and their interactions on the infectivity. The model showed that the original hosts generally had higher infection proportions compared to non-original hosts, with some exceptions, such as tomato infected with PVYSt (Fig. 1). It was also possible to detect three different patterns in non-original hosts, in which pepper plants could be infected by PVYCa and PVYSl, tomatoes by all isolates and potatoes only by

PVYSt. This is consistent with the expectation than viruses are better adapted to their original hosts.

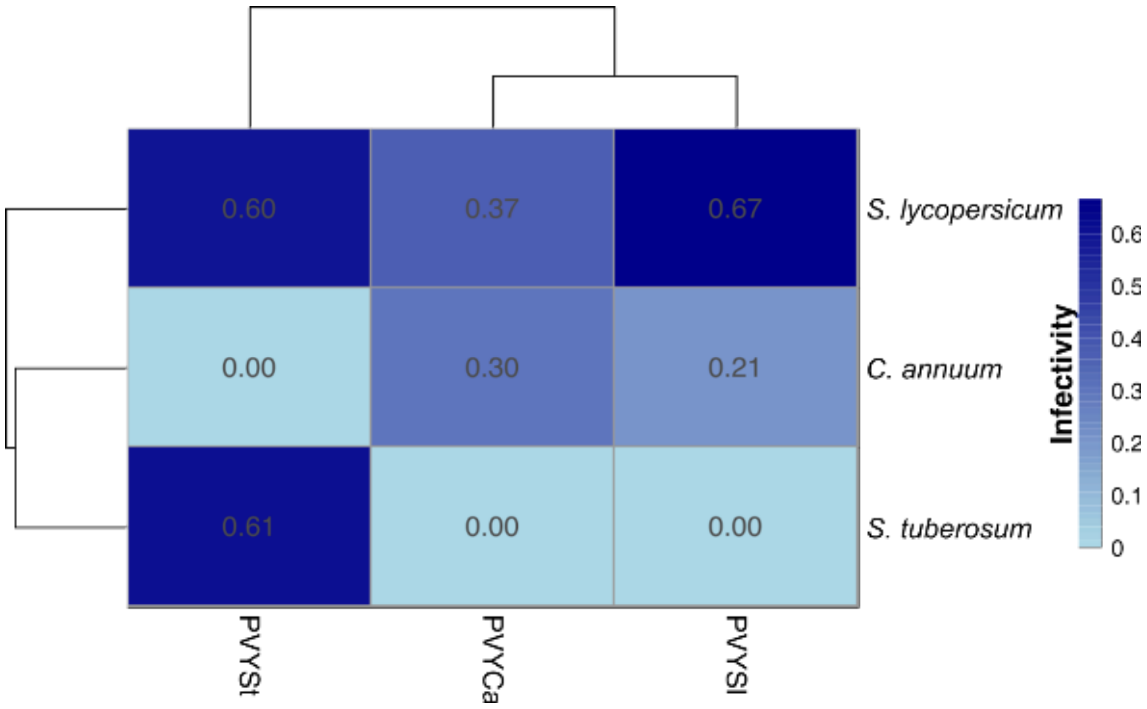


Fig 1. Infectivity proportions of the three plant species for each PVY isolate. Darker blue shades represent a higher number of infected plants, while lighter shades represent fewer infected plants.

Further studies could explore the mechanisms behind the observed infection patterns, such as differences in plant immune responses or viral replication efficiency in original versus non-original hosts. Understanding these underlying factors could improve the prediction of viral spread and the development of resistant plant cultivars. In conclusion, our findings confirm the importance of considering isolate specificity in screening and management strategies for disease control (reviewed in Karasev and Gray 2013).

Our systematic evaluation of diverse host-pathogen interactions aimed to uncover potential cross-species transmission patterns of PVY and their implications for disease spread and management strategies. We observed that hosts (genotype, physiology and phenology) may influence and shape the PVY population, with certain isolates showing limited impact on specific hosts upon initial infection. This phenomenon suggests the

presence of antagonistic pleiotropy, wherein mutations beneficial in one host may be detrimental in another (Whitlock 1996). Furthermore, phylogenetic analysis revealed a strong correlation between PVY phylogeny and host species origin, with pepper isolates clustering together and no specificity observed for PVY isolates in tomatoes (Quenouille et al. 2013). Based on the evidence that the choice of the isolate is crucial for screening purposes, we selected PVYCa to test sweet pepper cultivars and PVYSl to test tomato cultivars. We did not screen potato cultivars for resistance to PVY as all commercial cultivars are known to be susceptible (Karasev and Gray 2013).

Seeds of commercial cultivars of sweet pepper (n = 5) and tomato (n = 18) were searched in the market and subjected to inoculation trials, conducted twice, in Autumn and Summer, to ensure consistent results. Inoculations were performed and symptoms recorded, both in a greenhouse environment. Based on paired t-test (-1.2371, p-value 0.2297), Wilcoxon signed-rank test (29.5, p-value 0.2781), and Cohen's d (-0.264), there was no statistically significant difference rates between 1st and 2nd repetitions. The results from both trials were similar, prompting the calculation of the IR based on combined data.

Sweet pepper cultivars were inoculated with PVYCa, resulting in infection across all five cultivars. The IR ranged from 45% to 82%, averaging 74% (Fig. 2, green bars). Notably, severe symptoms such as blistering and interveinal chlorosis, along with leaf abscission and severe damage, were observed, particularly in cv. Ikeda (Sup. Fig. 2). Despite displaying strong symptoms, cv. Ikeda exhibited the lowest infection rate among all cultivars (45% IR).

The absence of resistant sweet pepper cultivars contradicts the description of these cultivars as resistant to PVY infection, according to the seed company. This discrepancy underscores the importance of using multiple isolates during cultivar screening, considering potential infection barriers. Indeed, previous studies have demonstrated such barriers, such as the findings that isolates from potatoes poorly infect pepper plants, consistent with our results (Blanco-Urgoiti et al. 1998; Romero et al. 2001; Moury 2010).

Tomato cultivars (18 in total) were mechanically inoculated with PVYSl in the greenhouse, with all cultivars displaying susceptibility to the virus. The infection rates

were even higher compared to sweet peppers, with ten cultivars exhibiting 100% IR, and the lowest rate at 88%, averaging 96% for all cultivars (Fig. 2, blue bars).

Despite the high infection rates, tomato cultivars exhibited mild symptoms (Sup. Fig. 3). This raised concerns about the detection of PVY in tomato fields, as visual inspections may miss strains inducing mild or no symptoms, potentially serving as undetected inoculum sources.

There are no studies that elucidate these questions in commercial cultivars, primarily because PVY is well studied in potatoes but not in other crops. In these cases, ELISA detection methodology can be used, ruling out false negatives based on symptomatology. Although the observation of mild symptoms in tomato plants has already been reported (Costa et al. 1960) and is in agreement with the results found here, the appearance of strong symptoms of necrosis caused by PVY, present in the middle third of the plant in tomato production fields, cannot be ruled out (Lucena et al. 2024). This means that the symptoms development may be related to the viral isolate, the cultivar, environment aspects, simultaneous mixed infections (for example the combination of PVY and potato virus X (Vance 1991) or PVY and potato spindle tuber viroid (PSTVd) (Qiu et al. 2014) or a combination of them or unknown factors.

Although there is no information regarding the resistance to PVY infection in any of these 18 tomato cultivars, they were chosen due to the agronomic characteristics they possess, but more importantly to the resistance to other pathogens. Altogether, they are resistant to bacteria, fungi, nematode or even virus infection. This includes the resistance of BRS Sena to begomoviruses, Itaipava and Viradoro to tospoviruses, Grazianni and Vento to tobamoviruses, Serato to tospoviruses and tobamoviruses and Cariri, Candieiro, Durino, Milão, Monza, Parma, Protheus, Santyno, and Tyson to begomoviruses, tobamoviruses and tospoviruses.

However, our tests revealed that none of the commercial sweet pepper or tomato cultivars exhibited resistance to the tested PVY isolates, highlighting the necessity of seeking new materials through breeding programs. This emphasizes the urgency of addressing PVY susceptibility in commercial cultivars to mitigate potential production losses and ensure crop health. Note that these cultivars, when infected, may serve as a reservoir of the PVY isolates.

As no commercial cultivar was resistant to PVY infection, wild lines of *Solanum* spp. accessions were screened in an attempt to obtain potential resistance sources. Fourteen wild tomato materials from the Instituto Agronômico de Campinas Germplasm Collection of *Solanum* species were screened: *Solanum pimpinellifolium* (PI 126 931, LA 722, LA 1584 and PI 126 925), *S. habrochaites* (PI 134 418 and PI 127 826), *S. lycoperiscum* (Ângela Hiper), *S. pennellii* (LA 716) and *S. peruvianum* (LA 462-2, PI 127 830, PI 270 435, IAC 237, LA 444-1 and PI 128 659). The wild tomato species were tested once, due to limited seed availability, with Ângela Hiper being the exception and tested three times.

All accessions were susceptible to PVY infection with IR between 22% to 100%: *S. pimpinellifolium* (*n* = 4) presented 97% of IR, *S. lycopersicum* (*n* = 1) presented 97%, *S. habrochaites* (*n* = 2) presented 77%, *S. pennellii* (*n* = 1) presented 100% and *S. peruvianum* (*n* = 6) presented 69% of IR (Fig. 3, golden bars). The only material that showed low IR was LA444-1 (*S. peruvianum*) with 22%. Although some accessions exhibited chlorosis, veinal chlorosis and leafroll, most of them exhibited no symptoms at all, suggesting tolerance of these accessions (Sup. Fig. 4). Despite their susceptibility, these accessions may still be important in the search for resistance against PVY due to the lower IR compared to other tomato cultivars (Tukey`s HSD = 0.1, p-value=0.041).

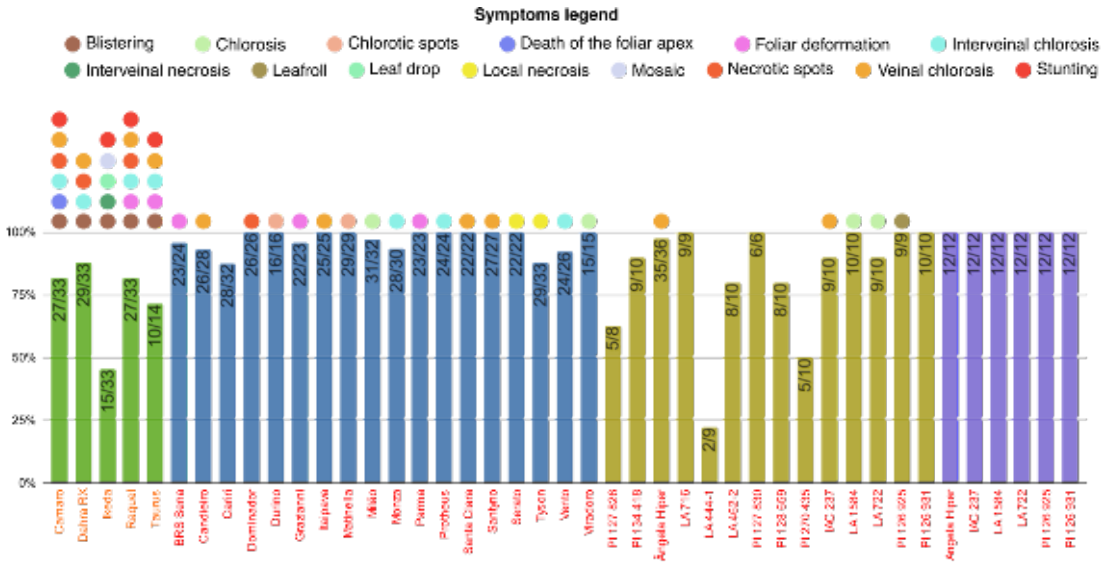


Fig 3. Comparative infection rates of tomato and pepper cultivars, and wild *Solanum* spp. accessions. Sweet pepper cultivars are represented by green bars, tomato cultivars

by blue bars, wild *Solanum* spp. accessions by golden bars, and their second-generation plants by purple bars. The number inside each bar indicates the number of plants positive for PVY infection followed by a slash and the total number of tested plants. Sweet pepper and tomato cultivars were evaluated in two different seasons in the greenhouse, with the number of plants representing the sum of positive and tested plants. The isolate used for inoculation is indicated by distinct colors on the *x*-axis: PVYCa in orange and PVYSl in red. Colored circles above the graph denote the presence of symptoms, with an absence of a circle indicating no symptoms.

Previous studies had identified the wild tomato LA444-1 as resistant against PVY based on the absence of symptoms using visual evaluation (Lourenção et al. 2005). Our findings demonstrate that while LA444-1 may serve as a potential source of resistance to PVY due to its lower IR, though it remains susceptible to PVY. This underscores the challenge of selecting the isolates for resistance tests, and also of relying solely on visual cues to determine resistance, especially when infected plants exhibit only mild or no symptoms, as described previously in others wild tomato accessions (Palazzo et al. 2008).

Interestingly, the cultivar Ângela Hiper, historically valued for its resistance to PVY, displayed unexpectedly high IR of 98%. Since the 1960s, significant efforts had been made to introgress PVY resistance into the tomato cultivar Santa Cruz, which was highly susceptible to this important disease. In the 1970s, through backcrossing between Santa Cruz and PI 126410 (*S. peruvianum*), a new cultivar called Ângela (Nagai and Costa 1969) was released. It was quickly adopted by tomato growers due to its resistance to PVY, *Fusarium oxysporum* f. sp. lycopersici race 1, and *Stemphylium solani*, as well as its high yield. Between 1975 and 1988, it was used on 75-80% of the total stalked (fresh market) tomato acreage. This initial success spurred the development of new cultivars, such as Ângela Hiper (Nagai et al. 1992), derived from the original. However, despite its past success, our extensive testing consistently revealed high levels of susceptibility (averaging 98%).

This result aligns with previous studies on screening wild tomato species for resistance, in which 19 *Solanum* spp. accessions were found to be susceptible to PVY, sometimes showing symptoms and other times remaining asymptomatic (Palazzo et al.

2008). However, the results obtained here indicate a higher level of susceptibility among the accessions, with a greater number of positive plants, suggesting that this virus isolate PVYSl has a potential to infect other tomato cultivars believed to be resistant to PVY infection.

To validate our findings and rule out the possibility of genetic segregation, we conducted an additional experiment with wild tomato accessions. We generated seeds from six autopollinated non-infected wild tomato plants, including one *S. lycopersicum*, one *S. peruvianum*, and four *S. pimpinellifolium* accessions. These seeds were then sown and subjected to PVY inoculation. All six cultivars exhibited a minimum IR of 90%, mirroring the parental generation's susceptibility. After mechanical inoculation with PVYSl, all first-generation plants displayed 100% IR, consistent with the parental generation, indicating no genetic segregation (Fig. 2, purple bars). As observed in the previous trial, no symptoms were observed in any of these plants. Thus, the susceptibility was confirmed for all commercial and wild tomato accessions to PVY infection. These findings collectively suggest that, although some may present escapes of infection, there are currently no known sources of resistance to the isolate PVYSl in tomatoes.

Based on the previously inoculation trials, PVYCa and PVYSl have similar infectivity properties, so PVYSt and PVYSl were tested in inoculation trials of various plant species. Our tests encompassed plant species from the Solanaceae family (*Datura metel*, *D. stramonium*, *Nicotiana benthamiana*, *N. glutinosa*, *N. rustica*, *N. sylvestris*, *Nicandra physalodes*, *Physalis pubescens* and *Solanum americanum*), Amaranthaceae (*Chenopodium amaranticolor* and *C. quinoa*) and Malvaceae (*Sida rhombifolia*). Due to low seed availability, not all hosts were tested with both isolates.

Our data show evidence that PVY infected hosts within the Solanaceae and Amaranthaceae families (Fig. 3), consistent with previous reports cataloging these plants as hosts of PVY (Edwardson and Christie 1997). However, *S. rhombifolia* (Malvaceae family) plants were not infected with PVY, corroborating existing reports that malvaceous plants are not hosts of PVY (Coutts and Jones 2014). PVYSl and PVYSt differed in the rate of IR in the tested hosts, in which PVYSl demonstrated to be more adapted to different hosts, compared to PVYSt (Fig. 3).

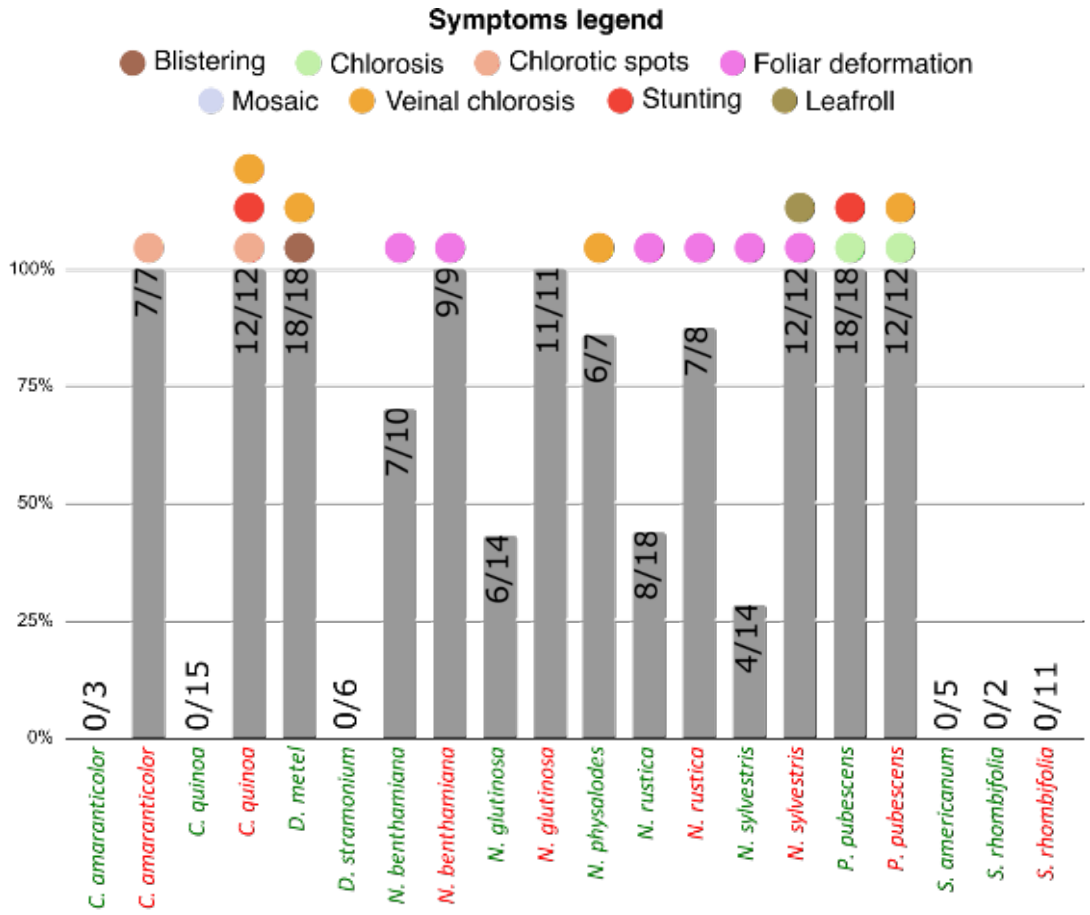


Fig 3. Experiments for determination of infection rates in indicator plants using PVYSt and PVYSl isolates from potato and tomato, respectively. Each PVY isolate is depicted by a distinct color on the *x*-axis: PVYSt in green and PVYSl in red. The number inside each bar indicates the number of plants positive for PVY infection followed by a slash and the total number of tested plants. Colored circles positioned above the graph denote the presence of symptoms, with an absence of a circle indicating no symptoms.

While both PVY isolates successfully infected most tested plants, the two exceptions were *C. amaranticolor* and *C. quinoa* plants. These two indicator plants are commonly used as test plants due to the production of easily countable local lesions after mechanical inoculation (Hollings 1956). They displayed unique symptoms upon inoculation with PVYSl. Initially, chlorotic spots with a red halo appeared on older leaves, which gradually evolved into systemic symptoms spreading throughout the plant (Sup. Fig. 6). This result contradicts previous knowledge of the local infection caused by PVY (Palazzo et al. 2008), demonstrating a concern with the use of model plants and

their applications. Importantly, this result was only observed when using PVYSl, while PVYSt was not able to infect this host, once again proving the importance of isolate choice. According to our results, it is crucial to exercise caution when performing detection tests, preferably conducting pilot tests to minimize the risk of false negative results and ensure accuracy.

While extensive research was conducted to elucidate the interactions between potato and PVY, such as transgenic approaches overexpressing PVY-derived coat protein, PVY-specific dsRNA (for RNAi), modified plant eIF4E, clustered regularly interspaced short palindromic repeats (CRISPR/Cas) and spray-induced gene silencing (SIGS) (Romano et al. 2001; Zimnoch-Guzowska et al. 2013; Valkonen et al. 2017; Torrance and Talianksy 2020), other crops such as tomatoes and peppers have received comparatively less attention. This highlights the need for increased research focus on tomato and pepper to develop effective PVY management strategies.

The absence of resistant materials from commercial or breeding programs underscores the urgency of addressing the spread of PVY in tomato and pepper production fields, as it allows the virus to persist. Furthermore, our findings highlight the variability in host range adaptation among different isolates of the same species, emphasizing the need for thorough testing using diverse isolates.

Organisms continually evolve and adapt to new environments, resulting in the emergence of new characteristics, including changes in their ability to infect hosts. Therefore, a more dynamic approach to understanding the interaction between the virus and its host is essential.

A comprehensive understanding of PVY and its adaptation across various host systems is vital for developing effective control strategies against this pathogen. Integration of advanced molecular techniques with a deep understanding of viral dynamics across diverse hosts is key to mitigating the impact of PVY and safeguarding global agricultural systems from its detrimental effects.

References

Agudelo-Romero P, Elena SF (2008) The degree of plant resilience to infection correlates with virus virulence and host-range. Spanish Journal of Agricultural Research 6:160. https://doi.org/10.5424/sjar/200806S1-384

Avilla C, Collar JL, Duque M, Fereres A (1997) Yield of bell pepper (*Capsicum annuum*) inoculated with CMV and/or PVY at different time intervals. Journal of Plant Diseases and Protection 104:1–8

Bedhomme S, Lafforgue G, Elena SF (2012) Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. Molecular Biology Evolution 29:1481–1492. https://doi.org/10.1093/molbev/msr314

Blanco-Urgoiti B, Sánchez F, Román CPS, et al (1998) Potato virus Y group C isolates are a homogeneous pathotype but two different genetic strains. Journal of General Virology 79:2037–2042. https://doi.org/10.1099/0022-1317-79-8-2037

Chikh-Ali M, Vander Pol D, Nikolaeva OV, et al (2016) Biological and molecular characterization of a tomato isolate of potato virus Y (PVY) of the PVYCA lineage. Archives of Virology 161:3561–3566. https://doi.org/10.1007/s00705-016-3071-9

Costa AS, Carvalho AMB, Kitajima EW (1960) Risca do tomateiro em São Paulo, causada por estirpe do vírus Y. Bragantia 19:1111–1128. https://doi.org/10.1590/S0006-87051960000100067

Coutts BA, Jones RAC (2014) Potato virus Y: contact transmission, stability, inactivation, and infection sources. Plant Disease 99:387–394. https://doi.org/10.1094/PDIS-07-14-0674-RE

Cuevas JM, Delaunay A, Visser JC, et al (2012) Phylogeography and molecular evolution of potato virus Y. PLoS One 7:e37853. https://doi.org/10.1371/journal.pone.0037853

Dijkstra J (1992) Importance of host ranges and other biological properties for the taxonomy of plant viruses. Archives of Virology pp 173–176. https://doi.org/10.1007/978-3-7091-6920-9_14

Dullemans AM, Cuperus C, Verbeek M, van der Vlugt RAA (2011) Complete nucleotide sequence of a potato isolate of strain group C of Potato virus Y from 1938. Archives of Virology 156:473–477. https://doi.org/10.1007/s00705-010-0853-3

Edwardson JR, Christie RG (1997) Viruses infecting peppers and other solanaceous crops. In: Florida Agricultural Experiment Station Monograph Series 18-II. University of Florida, Gainesville. Pp 424-524

Fereres A, Perez P, Gemeno C, et al (1993) Transmission of a Spanish pepper- and potato-PVY isolates by aphid (Homoptera: Aphididae) vectors: epidemiological implications. Environmental Entomology 22:1260-1265. https://doi.org/10.1093/ee/22.6.1260

Fonseca LN, Inoue-Nagata AK, Nagata T, et al (2005) Diferenciação de estirpes de Potato virus Y (PVY) por RT-PCR. Horticultura Brasileira, Brasília 4:904-910. https://doi.org/10.1590/S0102-05362005000400008

Galvino-Costa S, dos Reis Figueira A, de Assis Câmara Rabelo-Filho F, et al (2012) Molecular and serological typing of potato virus Y isolates from Brazil reveals a diverse set of recombinant strains. Plant disease 10:1451-1458. https://doi.org/10.1094/PDIS-02-12-0163-RE

García-Arenal F, Fraile A (2013) Trade-offs in host range evolution of plant viruses. Plant Pathology 62:2–9. https://doi.org/10.1111/ppa.12104

Gebre Selassie K, Marchoux G, Delecolle B, et al (1985) Variabilité naturelle des souches du virus Y de la pomme de terre dans le cultures de piment du sud-est de la France. Caractérisation et classification en pathotypes. Agronomie 5:621-630. https://doi.org/ 10.1051/agro:19850708

Green KJ, Chikh-Ali M, Hamasaki RT, et al (2017) Potato virus Y (PVY) isolates from *Physalis peruviana* are unable to systemically infect potato or pepper and form a distinct new lineage within the PVY C strain group. Phytopathology 107:1433–1439. https://doi.org/10.1094/PHYTO-04-17-0147-R

Hollings M (1956) *Chenopodium amaranticolor* as a test plant for plant viruses. Plant Pathology 5:57–60. https://doi.org/10.1111/j.1365-3059.1956.tb00085.x

Inoue-Nagata AK, Jordan R, Kreuze J, et al (2022) ICTV Virus Taxonomy Profile: *Potyviridae* 2022. Journal of General Virology 103:001738. https://doi.org/10.1099/jgv.0.001738

Inoue-Nagata AK, Fonseca MEN, Lobo TOTA, et al (2001) Analysis of the nucleotide sequence of the coat protein and 3'-untranslated region of two Brazilian Potato virus Y isolates. Fitopatologia Brasileira 26:45-52. https://doi.org/10.1590/S0100-41582001000100008

Jeffries CJ (1998) FAO/IPGRI Technical guidelines for the safe movement of potato germplasm. Food and Agriculture Organization of the United Nations, Rome/International Plant Genetic Resources Institute, Rome, Italy 19

Kang BC, Yeam I, Jahn MM (2005) Genetics of plant virus resistance. Annual Review of Phytopathology 43:581–621. https://doi.org/10.1146/annurev.phyto.43.011205.141140

Karasev AV, Xiaojun H, Brown CJ, et al (2011) Genetic diversity of the ordinary strain of potato virus Y (PVY) and origin of recombinant PVY strains. Phytopathology 7:778-785. https://doi.org/10.1094/PHYTO-10-10-0284

Karasev AV, Gray SM (2013) continuous and emerging challenges of potato virus Y in potato. Annual Review of Phytopathology 51:571–586. https://doi.org/10.1146/annurev-phyto-082712-102332

Lourenção AL, Siqueira WJ, Melo AMT, et al (2005) Resistência de cultivares e linhagens de tomateiro a tomato chlorotic spot virus e a potato virus Y. Fitopatol Brasileira 30:609–614. https://doi.org/10.1590/S0100-41582005000600007

Lucena VS, Nakasu EYT, Pereira JL, et al (2024) Emergence of potato virus Y outbreaks in tomatoes in Brazil, the disease and spread. bioRxiv. https://doi.org/10.1101/2024.05.17.594728

Malpica JM, Sacristán S, Fraile A, García-Arenal F (2006) Association and host selectivity in multi-host pathogens. PLoS One 1:e41. https://doi.org/10.1371/journal.pone.0000041

McLeish MJ, Fraile A, García-Arenal F (2019) Evolution of plant-virus interactions: host range and virus emergence. Current Opinion in Virology 34:50-55. https://doi.org/10.1016/j.coviro.2018.12.003

Meissner PEF, Cupertino FP, Avila AC (1990) Purificacao e serologia de dois isolados do virus Y da batata obtidos de pimentão e de tomate no Distrito Federal. Fitopatologia Brasileira 15:235-237.

Moury B (2010) A new lineage sheds light on the evolutionary history of potato virus Y. Molecular Plant Pathology 11:161–168. https://doi.org/10.1111/j.1364-3703.2009.00573.x

Nagai H (1971) Novas variedades de pimentão resistentes ao mosaico causado por vírus Y. Bragantia 30:91–100. https://doi.org/10.1590/S0006-87051971000200001

Nagai H, Costa AS (1969) Incorporação de resistência ao mosaico Y em tomateiro. Bragantia 28:219–226. https://doi.org/10.1590/S0006-87051969000100017

Nagai H, Lourenção AL, Siqueira WJ (1992) Tomato breeding for resistance to diseases and pests in Brazil. Acta Horticultae 301:91–97. https://doi.org/10.17660/ActaHortic.1992.301.10

Nagata T, Inoue AK, Dusi A, Kitajima EW (1995) Bidens mosaic potyvirus newly isolated from pea, its characteristics and serological relationship with other potyviruses. Fitopatologia Brasileira 20:473-478

Nigam D, LaTourrette K, Souza PFN, Garcia-Ruiz H (2019) Genome-wide variation in potyviruses. Frontiers in Plant Science 10. https://doi.org/10.3389/fpls.2019.01439

Palazzo SRL, Colariccio A, Marchi ATM (2008) Reação de acessos de *Lycopersicon* spp. a um isolado de potato virus y (PVY^O) de tomateiro. Bragantia 67:391–399. https://doi.org/10.1590/S0006-87052008000200015

Qiu CL, LV WH, LV DQ et al (2014) Symptoms of four potato varieties infected with Potato spindle tuber viroid (PSTVd). Journal of Plant Protection Research 6:159-163

Quenouille J, Vassilakos N, 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. https://doi.org/10.1111/mpp.12024

Romano E, Ferreira AT, Dusi AN, et al (2001) Extreme resistance to two Brazilians strains of potato virus Y (PVY) in transgenic potato cv Achat expressing the PVY^O coat protein gene from argentine PVY^O strain. Horticultura Brasileira 2:118-122

Romero A, Blanco-Urgoiti B, Soto MJ, et al (2001) Characterization of typical pepper-isolates of PVY reveals multiple pathotypes within a single genetic strain. Virus Research 79:71–80. https://doi.org/10.1016/S0168-1702(01)00300-8

Singh RP, Valkonen JPT, Gray SM, et al (2008) Discussion paper: The naming of potato virus Y strains infecting potato. Archives of Virology 153:1–13. https://doi.org/10.1007/s00705-007-1059-1

Torrance L, Talianksy ME (2020) Potato Virus y emergence and evolution from the Andes of South America to become a major destructive pathogen of potato and other solanaceous crops worldwide. Viruses 12. https://doi.org/10.3390/V12121430

Valkonen JPT, Gebhardt C, Zimnoch-Guzowska E, Watanabe KN (2017) Resistance to Potato virus Y in potato. In: Potato virus Y: biodiversity, pathogenicity, epidemiology and management. Springer International Publishing, Cham, pp 207–241

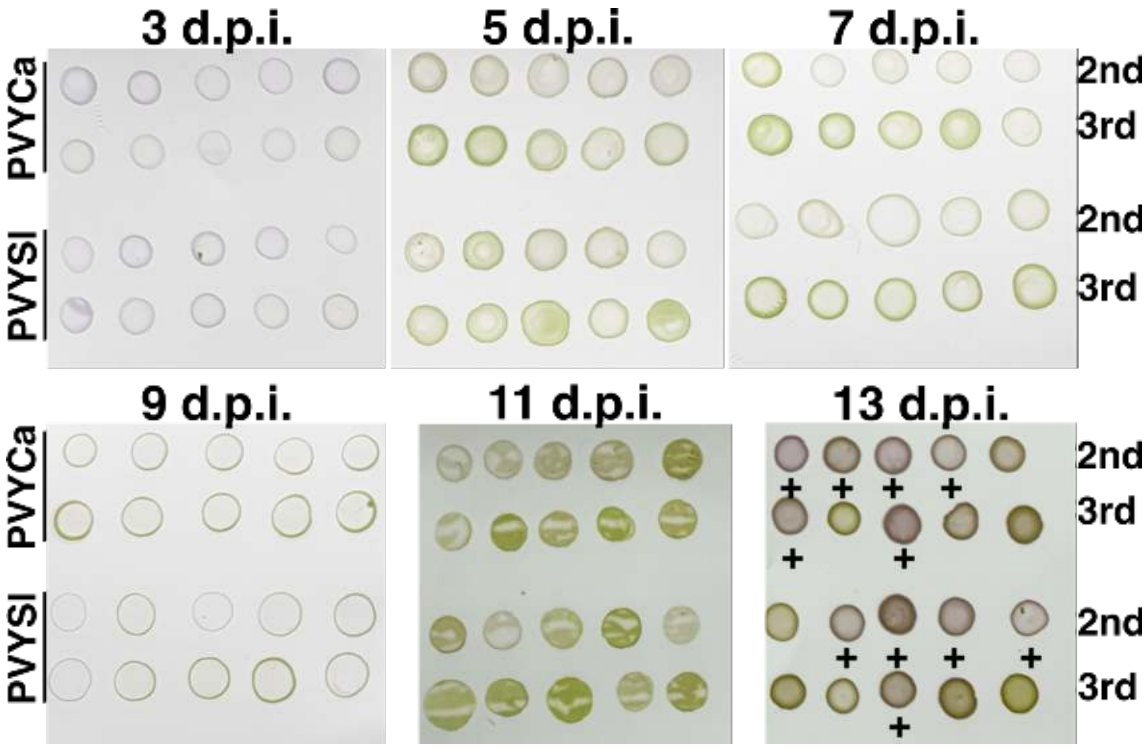
Vance BV (1991) Replication of potato virus X RNA is altered in coinfections with potato virus Y. Virology 181:486-494. https://doi.org/10.1016/0042-6822(91)90589-4

Vassilakos N, Simon V, Tzima A, et al (2016) Genetic determinism and evolutionary reconstruction of a host jump in a plant virus. Molecular Biology and Evolution 33:541–553. https://doi.org/10.1093/molbev/msv222

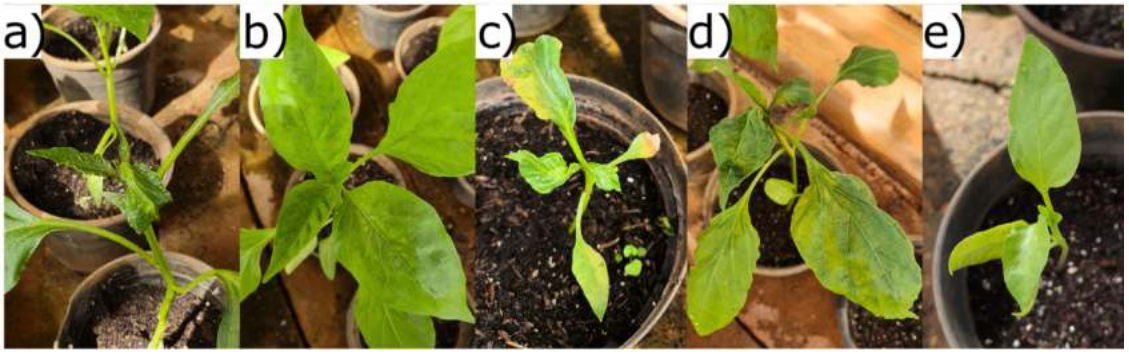
Whitlock MC (1996) The red queen beats the jack-off-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. The American Naturalist 148:S65–S77.

Yarwood CE (1979) Host passage effects with plant viruses. Advances in Virus Research pp 169–190. https://doi.org/ 10.1016/s0065-3527(08)60570-9

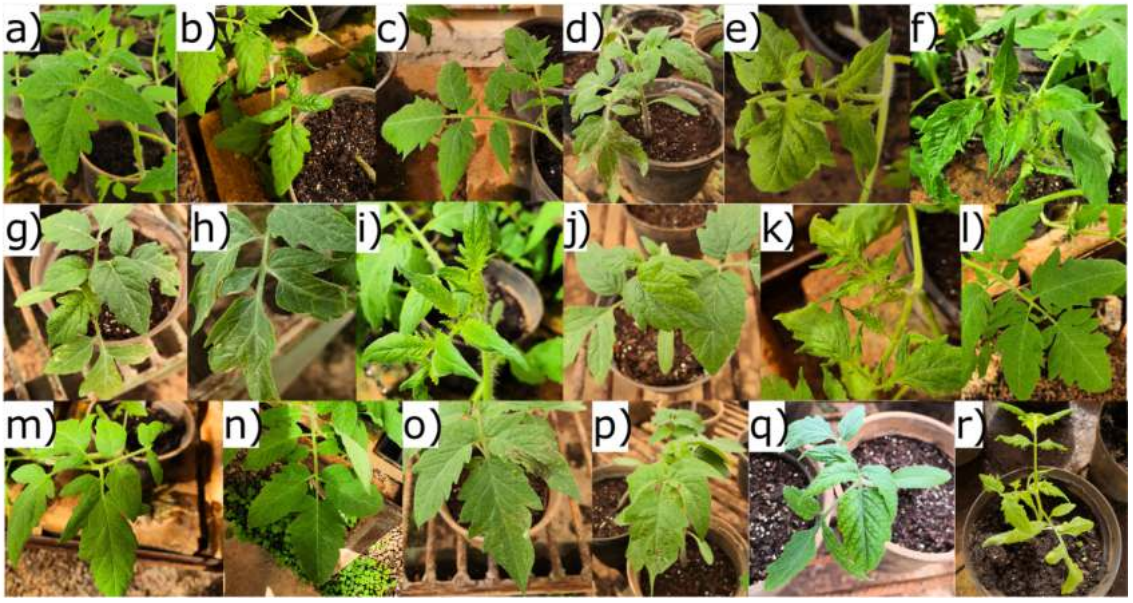
Zimnoch-Guzowska E, Yin Z, Chrzanowska M, Flis B (2013) sources and effectiveness of potato PVY resistance in ihar's breeding research. American Journal of Potato Research 90:21–27. https://doi.org/10.1007/s12230-012-9289-5



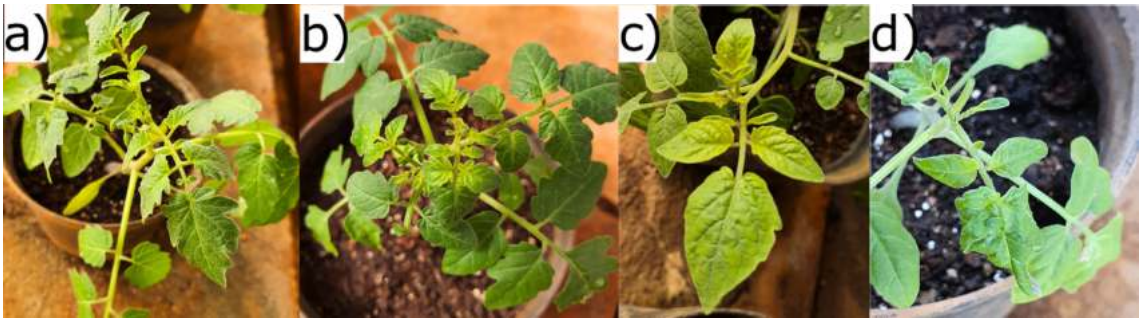
Sup. Fig 1. Dot-ELISA of pepper plants with 3, 5, 7, 9, 11 and 13 days post inoculation (d.p.i.) using two isolates, PVYCa and PVYSl, and two different leaves, the second (2nd) and the third (3rd) leaf from the top. The purple color development represents a positive reaction. Positive samples are marked with + symbol.



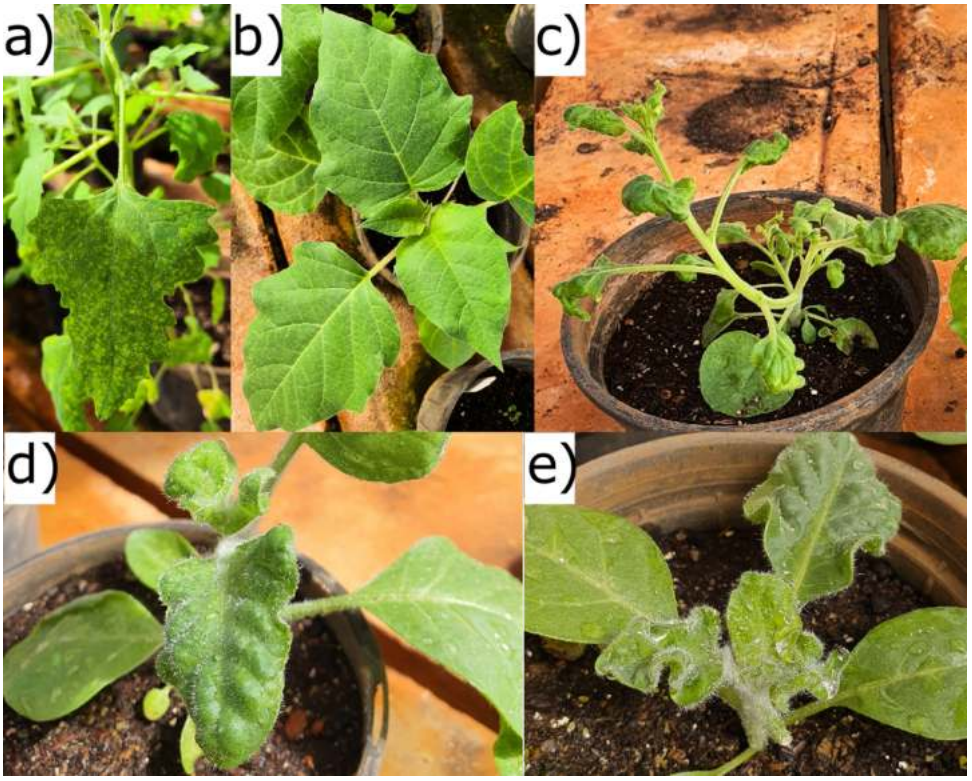
Sup Fig 2. Symptoms observed in sweet pepper cultivars mechanically inoculated with PVYCa: (a) foliar deformation and interveinal chlorosis in Camaro, (b) interveinal and veinal chlorosis in Dahra RX, (c) foliar deformation, stunting and necrotic spots in Ikeda, (d) foliar deformation, interveinal and veinal chlorosis in Raquel, and (e) stunting in Taurus.



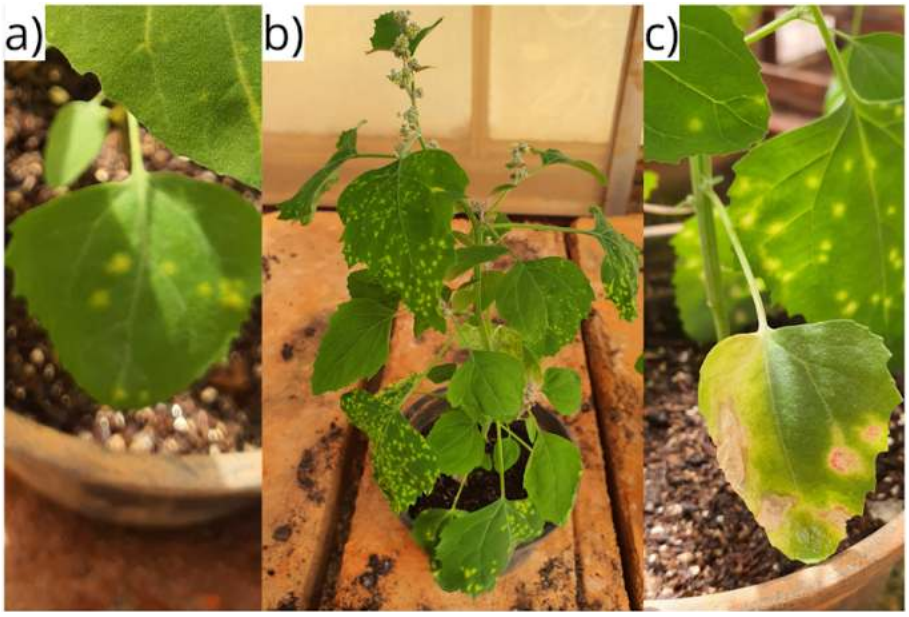
Sup Fig 3. Symptoms observed in tomato cultivars mechanically inoculated with PVYSl: (a) veinal chlorosis in Ângela Hiper, (b) foliar deformation in BRS Sena, (c) veinal chlorosis in Candieiro, (d) necrotic spots in Dominador, (e) chlorotic spots in Durino, (f) foliar deformation in Grazianni, (g) veinal chlorosis in Itaipava, (h) chlorotic spots in Matinella, (i) chlorosis in Milão, (j) interveinal chlorosis in Monza, (k) foliar deformation in Parma, (l) interveinal chlorosis in Protheus, (m) veinal chlorosis in Santa Clara, (n) veinal chlorosis in Santyno, (o) local necrosis in Serato, (p) local necrosis in Tyson, (q) interveinal chlorosis in Vento and (r) chlorosis in Viradoro. Only cultivars with symptoms are shown in the figure.



Sup Fig 4. Symptoms observed in wild tomato cultivars mechanically inoculated with PVYSl: (a) veinal chlorosis in IAC 237, (b) chlorosis in LA1584, (c) chlorosis in LA722 and (d) leafroll in PI126925. Only cultivars with symptoms are shown in the figure.



Sup Fig 5. Symptoms observed in indicator plants mechanically inoculated with PVYSt and PVYSt (a) chlorotic spots in *C. quinoa* (PVYSl), (b) veinal chlorosis in *D. metel* (PVYSt), (c) foliar deformation in *N. benthamiana* (PVYSl), (d) foliar deformation in *N. rustica* (PVYSl) and (e) blistering and foliar deformation in *N. sylvestris* (PVYSl).



Sup Fig 6. Evolution of symptoms in *C. amaranticolor* infected with PVYSl, showing (a) the first chlorotic spots on the inoculated leaves, followed by (b) systemic infection with chlorotic spots in young leaves and (c) the appearance of red halos around old lesion.

Chapter II

Detecting and sequencing the whole-genome of distinct potato virus Y isolates using a PCR-Nanopore approach

Abstract

Potato virus Y (PVY) is a relevant pathogen affecting a range of solanaceous crops, including potatoes, tomatoes, and peppers. This study aimed to develop and validate a robust sequencing approach for PVY genomes using Oxford Nanopore Technologies (ONT). For selecting the virus isolates to be sequenced, a total of 21 PVY-positive samples were collected from infected potato, pepper, and tomato plants. One isolate per host was selected to conduct ONT sequencing: PVYCa (pepper), PVYSt (potato), and PVYSl (tomato). For this purpose, four overlapping primer sets covering the complete viral genome were designed based on conserved regions identified through alignment of a large dataset (n = 445). We also used a barcorde kit, able to sequence up to 24 samples in the same flowcell. Although a similar amount of input amplicons was used for sequencing, the obtained read coverage was not uniform along the genome, yet we were able to produce sufficient reads to assemble the genome of all isolates. The number of reads varied according to the samples, but the expected sizes of ~1.8 and ~3 kb were consistently obtained, including long reads covering the entire genome. Illumina sequencing was used to validate the Nanopore assembly using the isolate PVYCa. By calculating the pairwise nucleotide distance using Tamura-Nei model and performing a phylogenetic analysis, our results demonstrated a high level of identity between both PVYCa genomes, validating the sequence quality obtained by the ONT approach. Furthermore, we developed PVY-specific primers to facilitate specific detection. These primers effectively distinguished PVY from other viruses, including closely related potyviruses. This study highlights the reliability of ONT for sequencing diverse PVY genomes, demonstrating its utility for high-throughput, cost-effective, and rapid viral genome analysis. The successful application of this methodology in sequencing multiple

PVY isolates will contribute to a deeper understanding of PVY diversity and host interactions, advancing on both diagnostic and evolutionary studies.

Key-words: High-throughput genome sequencing, Nanopore, Sequencing methodology, Virus detection, Virus epidemiology, Virus sequencing

Introduction

Potato virus Y (PVY) is a positive-sense, single-stranded RNA virus, classified in the family *Potyviridae*, genus *Potyvirus* and species *Potyvirus yituberosi*, with a genome size of approximately 9.7 Kb. It encodes a polyprotein that undergoes autoproteolysis (Inoue-Nagata et al. 2022). PVY can be transmitted in a non-persistent manner by at least 65 species of aphids (Lacomme et al. 2017; Rizk et al. 2020). PVY has a broad host range (Edwardson and Christie 1997; Jeffries 1998) that includes important solanaceous plants, *e.g.*, tomato (*Solanum lycopersicum*), potato (*S. tuberosum*), and pepper (*Capsicum annuum*) (Kerlan et al. 2008). Potato plants are greatly affected by PVY infection, with reported losses of up to 80% (Hane and Hamm 1999) and adversely affecting tuber quality (Le Romancer et al. 1994). Yield reduction of 0.1805 Tons/ha has been reported for each 1% increase in PVY incidence (Nolte et al. 2004). In contrast, studies of PVY in tomatoes are limited. These properties rendered to PVY a ranking position of the fifth most important plant virus (Scholthof et al. 2011). This position may not have changed in the last decade.

RNA viruses, such as PVY, have high mutation rates of 10^{-6} to 10^{-4} substitutions/nucleotide/replication (Peck and Lauring 2018) due to various mechanisms, such as the lack of 3' exonuclease proofreading activity of the RdRp (Steinhauer et al. 1992), genome size, type, and replication mode (Sanjuán and Domingo-Calap 2016). Notwithstanding, recombination is known to highly modulate PVY populations (Revers et al. 1996). Recombination in viruses mostly happens when two or more virus genomes combine through replicase-driven template switching, resulting in a chimeric genome that may exhibit unique genetic traits compared to their parental viruses. This process is particularly significant in potyviruses, for which the estimated recombination rate is 3.427×10^{-5} per nucleotide site per generation, comparable to the rate of mutation, highlighting the important role of recombination in generating viral diversity (Tromas et al. 2014). Hence, PVY has a high genetic variability, and it exists as a complex of strains, classified based on the symptoms' development in potatoes and tobacco. Initially, PVY was classified in three strains, the ordinary (PVY^{O}), common (PVYCA) and necrotic (PVY^{N}) (Jones 1990; Singh et al. 2008). Then, many recombinant strains were reported in the last 20 years (Le Romancer

et al. 1994; Chikh Ali et al. 2010; Karasev et al. 2011; Funke et al. 2017; Green et al. 2017, 2020; Davie et al. 2017; Rodriguez-Rodriguez et al. 2020), which corroborates with the high mutation rates.

A broad range of diagnostic tools are available to detect plant viruses, including enzyme-linked immunosorbent assay (ELISA), loop mediated isothermal amplification (LAMP) and the most commonly used polymerase chain reaction (PCR). Also important, genome sequencing is generally used to detect, identify and determine the virus properties. More recently, high throughput sequencing, HTS (or next generation sequencing) became popular for virus identification. Obtaining the complete genome sequence of a virus is still costly and labor-intensive when using the traditional methods of Sanger sequencing or HTS. With the purpose to facilitate and cheapen the sequencing process, Oxford Nanopore Technology (ONT) released an equipment and protocols based on the use of nanopores to determine the base sequences by analyzing the electrical current when the nucleic acid passes through these nanopores (Mikheyev and Tin 2014). ONT sequencing offers various benefits, such as the ability to sequence individual molecules, generate lengthy sequencing reads, achieve fast sequencing speeds, and monitor sequencing data in real-time (Laver et al. 2015; Deamer et al. 2016). Likewise, to decrease even more the time and price of sequencing per sample the use of barcodes can be applied, sequencing more than one sample in a single flow-cell. The ONT MinION, the Nanopore sequencing platform, has gained widespread acceptance, and it can be powered through a USB port on a personal computer. This platform is unique among other sequencing technologies because it allows for sequencing and real-time data analysis to be conducted directly on laboratory benches. Since its introduction to the public, ONT has been employed in several plant virus genome studies (Martins et al. 2021; Amoia et al. 2022; Dong et al. 2022), though it remains in its early stages and offer substantial potential for further refinment. Nanopore sequencing offers numerous applications in virus research, including viral detection and surveillance, genome assembly, the discovery of new variants and novel viruses, and the identification of chemical modifications and impose advantages over HTS, such as the capability to produce ultra-long reads, real-time monitoring and analysis, portability, and the ability to directly sequence RNA or DNA molecules (Ji et al. 2024).

As explained above, PVY is characterized by a diverse population structure composed of a large number of variant genomes, known as *quasispecies* (Más et al. 2010), which arises due to its high mutation rates. This mutant cloud is constantly changing in relative frequency during viral replication. Sequencing the genomes of many isolates, and capturing the diversity present in their genomes are an arduous task, whereas they are essential for population structure and evolution studies. For this reason, we wanted to use PVY as the model virus to evaluate the use of Nanopore sequence for genome sequence studies. Three PVY isolates were selected, one from potato, one from pepper and another one from tomato.

Here, we provide a step-by-step guide for a simple strategy to detect and obtain up to 24 PVY whole genome sequences in a single ONT flow-cell. Our methodology is efficient in enriching and amplifying the target sequences, as we designed primers that are highly specific for PVY genomes while also detecting all available variability within the species.

Materials and methods

Virus isolates collection

PVY isolates were collected from pepper (*Capsicum* annuum), potato (*Solanum tuberosum*) and tomato (*S. lycopersicum*) in commercial fields located near each other in the greenbelt of Brasília (16°04'23.1"S 47°21'32.9"W to PVYSt and 15°55'58.8"S 47°35'47.1"W to PVYCa and PVYSl), Distrito Federal, Brazil. We randomly collected plants with and without apparent symptoms (Sup. Table 1 and Sup. Fig. 1).

All samples (n = 18 from pepper, n = 52 from potato and n = 110 from tomato) were submitted to a serological test (dot-ELISA) with our PVY polyclonal antibody, according to Nagata et al. (1995) (a list of positively detected plants is detailed in Sup. Table 1). The leaves were weighed, homogenized in phosphate-buffered saline (PBS, pH 7.0), and spotted in two dilutions: 1:10 and 1:100. After color development using nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indolyl-phosphate (BCIP), we randomly chose one positive sample from each plant species to proceed to the next

steps. The virus isolates were identified as PVYCa (from pepper), PVYSt (from potato) and PVYSl (from tomato).

RNA extraction, cDNA synthesis, PCR and Sanger sequencing

Total RNA was extracted using the RNeasy plant mini kit (Qiagen, Hilden, Germany) following the manufacturer protocol. The cDNA was constructed using SuperScript III (ThermoFisher, California, USA) with primers M4T (5'-GTT TTC CCA GTC ACG AC(T_{15})-3') (Chen et al. 2001) and random hexamers. An incubation at 37 °C for 20 min using 2 Units of *Escherichia coli* RNase H (ThermoFisher) was done to remove the RNA strand of the RNA-cDNA hybrid.

To confirm the identity of the PVY isolates, their genomes were partially sequenced by Sanger sequencing. PCR was performed with *Taq* DNA recombinant polymerase (ThermoFisher) with M4 (5'-GTT TTC CCA GTC ACG AC-3') and Sprimer (5'-GGX AAY AAY AGY GGX CAZ CC-3') primers (Chen et al. 2001), producing an amplicon of ~1.7 kb. The PCR products were separated on a 1% agarose gel, the agarose gel fragments containing the target DNA were sliced and isolated from the agarose gel using Wizard SV Gel and PCR Clean-up System (Promega, Wisconsin, USA). The amplicons were sequenced using the Sanger method with Sprimer by Macrogen Inc. (South Korea).

Nanopore primer design

All complete genomes of PVY available in the GenBank database were downloaded (*n* = 634) in February, 2023. The dataset was analyzed, excluding all dubious and incomplete sequences, aligned and manually adjusted using Muscle (Edgar 2004). Only one representative haplotype was maintained by using DnaSP (Rozas et al. 2017) resulting in a dataset of 445 sequences. Highly conserved regions were searched in the alignment. These regions were selected and candidate primers were designed and evaluated using optimal primer conditions, such as melting temperature, folding and hybridization of strands, GC content and amplicon size using OlygoAnalyzer (Owczarzy et al. 2008).

Four sets of degenerated primers with overlapping regions were designed in these highly conserved regions, covering the entire PVY genome (Table 1). Each set of

primers was composed of three forward and three reverse primers (A to C) with minor differences, named Y1F to Y4F and Y1R to Y4R for forward and reverse primers, respectively. To avoid any eventual mismatch in the last base of the primer, hence capturing all the genome diversity, an inosine was added at the 3'-termini of all designed primers. For amplification of the 3' terminal end, the M4 primer was used in combination with the Y4F primer mix.

Amplification of target region with PCR

PCR amplification was performed using the cDNA with high-fidelity Q5 DNA polymerase (NEB, Massachusetts, USA) according to the manufacturer recommendations. The best temperature and reaction conditions were tested for obtaining the highest amplicon yield. For sequencing, the PCR products were purified using Wizard SV Gel and PCR Clean-up System (Promega, Wisconsin, USA) to remove excess primers and nucleotides.

ONT sequencing strategy

Amplified DNA fragments were first quantified by Qubit 3 fluorometer (Invitrogen, Massachusetts, USA), and mixed to obtain equimolar quantities for amplicon and for PVY isolate, according to the recommended protocol.

The barcode expansion kit (EXP-NBD 104, NEB) was used to sequence the three viruses at the same time, following the recommendation. The DNA repair and end-preparation were performed without the fragmentation step. After barcoding, the three amplicon pools were measured and diluted again to have the same equimolar quantity in each of the three samples, approximately 600 ng each. The sequencing was done using the Nanopore ligation sequencing kit SQK-LSK109 (NEB) and the prepared library was mixed together and loaded on a MinION with a R9.4.1 flow cell (FLO-MIN106). All other procedure steps of native barcoding genomic DNA Nanopore were followed as described in the recommended protocol.

Sequencing analysis and genome assembly

The quality of the reads was assessed both within and between samples using NanoPack2 (De Coster et al. 2018). Three PVY isolates (PVYCa, PVYSt, PVYSl) were

subjected to Nanopore sequencing at the same time and in the same flowcell. Sequencing data were analyzed to derive mean and median read lengths, mean and median read quality scores, number of reads, total bases, read length N50, standard deviation (SD) of read lengths, and read quality distribution (> Q10, > Q15, > Q20). Dorado (Oxford Nanopore Technologies) was employed for base calling. It is important to notice that the Q-score for Nanopore and Illumina sequencing are similar, but the cutoff values and achievable accuracies differ because Nanopore works with lower Q-scores, and higher error rates are compensated for by different downstream processing strategies, such as polishing and consensus generation.

Subsequently, minimap2 (Li 2018) was used to align reads to the reference PVY genome (X12456) and convert it to SAM file. SAM file was then converted and sorted to BAM using Samtools (Danecek et al. 2021) and used to assemble the consensus genome using Geneious Prime v. 2022.2 (Biomatters). Manual inspection was conducted to validate the assembled genomes, and BLASTn (Johnson et al. 2008) was utilized to compare the three genomes against the GenBank nucleotide sequence database.

Total RNA purification and Illumina sequencing

To validate the accuracy of the Nanopore genome sequence and assembly, we selected one isolate, PVYCa, to sequence with the conventional HTS method by using the Illumina platform. Pepper cultivar Ikeda, infected with the original PVYCa, was used to semi-purify the viral particles (Blawid et al. 2017). Total RNA was extracted using the AllPrep DNA/RNA Micro Kit (QIAGEN) and subsequently sequenced on the Illumina Novaseq platform by Macrogen Inc. (Seul, South Korea). Following Illumina sequencing, the reads underwent trimming using BBduk (https://sourceforge.net/projects/bbmap/), and the resulting contigs were *de novo* assembled using MEGAHIT (Li et al. 2015). The assembled contigs were aligned and subjected to diamond Blastx (Buchfink et al. 2015) against the nr database (downloaded on the 2024-06-24) to exclude non-PVY contigs. The longest contig was subjected to BLASTn against the nucleotide database to identify the closest isolate (EU563512) and then the reads were aligned to this sequence with BBMap

227 (https://sourceforge.net/projects/bbmap/) and the consensus sequence was generated
228 with Geneious Prime v. 2022.2.
229
230 **Genome analysis**
231 In order to compare the assembled consensi genomes, we computed the pairwise
232 distances using the better-fit model with the lowest AIC, Tamura-Nei model with
233 Gamma distributed rates, using MEGAX (Kumar et al. 2018). To this analysis, we
234 included both PVYCa assembled genomes (Illumina and Nanopore) along with PVYSl,
235 PVYSt, and the PVY reference genome (X12456). Pheatmap package (Kolde 2019) on
236 R (R Core Team 2022) was used to generate the heatmap. Furthermore, to confirm the
237 accuracy of the sequencing, we calculated the Pearson correlation coefficient between
238 the distances obtained from Illumina and Nanopore sequencing of PVYCa.
239 In a second approach, we added 49 representative PVY genomes from diverse
240 strains, countries and hosts, together with our sequenced isolates (n = 53). Furthermore,
241 we included three outgroup species, namely bidens mosaic virus (BiMV) (Dujovny et
242 al. 1998; Inoue-Nagata et al. 2006b), pepper severe mosaic virus (PSMV) (Ahn et al.
243 2006) and sunflower chlorotic mottle virus (SCMoV) (Dujovny et al. 2000), known to
244 be the closest relatives to PVY. Consequently, our dataset consisted of a total of 56
245 genomes. After construction of the alignment with Muscle (Edgar 2004), a
246 maximum-likelihood (ML) phylogenetic tree was inferred using iq-tree2 (Minh et al.
247 2020) with 10,000 bootstrap replicates. The phylogenetic tree was edited using iTol
248 (Letunic and Bork 2021).
249
250 **Design and evaluation of PVY-specific primers**
251 Two PVY-specific primers, a forward (YSF: 5'-ACT ATG ATT TTT CGT CGA GAA
252 CAA G-3') and a reverse (YSR: 5'-GGC GAG GTT CCA TTT TCA ATG C-3') primer,
253 were designed using the same alignment (n = 445) that was used for Nanopore primer
254 design. We searched for regions conserved in PVY genomes, using primer Blast
255 (Johnson et al. 2008).
256 In order to test the efficiency of the primers, they were tested with three PVY
257 isolates. Additionally, we included in our analysis other widespread viruses that infect
258 solanaceous plants: groundnut ringspot virus (GRSV - genus *Orthotospovirus*), pepper

yellow mosaic virus (PepYMV - *Potyvirus*), pepper mild mottle virus (PMMoV - *Tobamovirus*), and tomato mosaic virus (ToMV - *Tobamovirus*). Total RNA and the cDNA construction of infected leaves were performed essentially as described above. We also included a non-infected healthy plant in our analyses. The PCR was done using *Taq* DNA recombinant polymerase (ThermoFisher) in a 35-cycling reaction of 95 ℃ denaturing for 30 sec, 52 ℃ annealing for 30 sec and 72 ℃ extension for 1 min with a final extension of 10 min.

Results

Field collection and identification of PVY isolates

We wanted to establish a protocol for amplification and sequencing of PVY genomes within a wide range of diversity, hence isolates were collected from potato, pepper and tomato plants, expecting they were divergent though coexisting in the same agroecosysstem. Leaves of infected plants were collected in different fields and a dot-ELISA procedure using polyclonal PVY antibody was performed as the first detection test. We detected 17 positive samples in the potato crop (Infection Rate (IR) = 32.7%), 2 positive samples in pepper (IR = 11.1%) and 2 positive samples in tomatoes (IR = 1.8%) (Supplementary Table 1). Infected plants exhibited blistering, chlorosis, mosaic and necrosis in pepper; chlorosis and necrotic spots in potato; and no symptom in tomato plants. One positive sample was selected from each crop, total RNA was extracted, and used for cDNA construction with a random hexamer and an anchored oligodT primer. The 3' terminal region of the genome was amplified by PCR using the Sprimer, located in the NIb region, and the anchor primer M4, described as universal potyvirus primers (Chen et al. 2001). This fragment was Sanger sequenced and confirmed that all three viruses are isolates of PVY (not shown). They were named PVYCa (pepper isolate), PVYSt (potato isolate) and PVYSl (tomato isolate).

Nanopore primer design and evaluation

To design primers able to capture the diversity of the PVY genome, we used all full genome sequences available at the GenBank (n = 445) for complete genome alignment

and searched for conserved regions. The PVY genome was divided in four regions, three with ~3 kb and a 3'-end region with ~1.8 kb. Therefore, four primer sets were designed, each set being composed by three forward and three reverse primers (except the set 4, which uses M4 as reverse primer) (Fig 1a - Table 1). Some primers were degenerated (a list of all primers and their characteristics can be found in Table 1). An inosine was added to the 3' end of the primer to avoid misannealing due to unexpected divergency in this position.

The primers were tested with the three PVY isolates and the optimal Q5 DNA Polymerase PCR conditions were determined for each set of primers (Table 1). Using the optimal conditions, a PCR was done for each primer set and each isolate. Some conditions may include the use of GC enhancer due the presence of rich G-C regions and a difference in extension time for Set 4. All primers have the same melting temperature of 55 degrees. All sets of primers successfully amplified all genome regions for all the three isolates (Fig 1b).

The PCR amplification yield was not uniform for all amplicons, consequently before and after the barcoding ligation, the DNA was measured to ensure the input of equimolar quantity and then proceed to sequencing (Fig 1c).

Tab 1. Description of the sequencing primers and the PCR conditions for amplification of the whole-genome of PVY using Q5 DNA polymerase. Each color represents a set of primers used to sequence each fragment.

Primer	Sequence (5' → 3')	Size	%GC	GC En$	ET (min)#	MT (°C)*
Y1F-A	AAATTAAAACAACTCAATACAACATAAI	28	18			
Y1F-B	AAATTAAAACAACTCAATACAACAI	25	20			
Y1F-C	AAATTAAAACAACTCAATACAI	22	18	no	1:45	55
Y1R-A	AACGCCTAAAGATTCTACGAATI	23	35			
Y1R-B	AAACGCCTAAAGAKYSTACGI	22	33			
Y1R-C	GGCAAACGCCTAAARAKYSTAI	22	32			

Y2F-A	ATGGAAAAAAAYTATCTARRYCTCTTI	27	26		
Y2F-B	ATGGAAAAAAAYTATCTARRYCTCI	25	28		
Y2F-C	TTATGGAAAAAAAYTATCTARRYCI	25	24		
Y2R-A	GCYTTRTCRBACCARTCYTI	20	46		
Y2R-B	TTRTCRBACCARTCYTTYCTI	21	39		
Y2R-C	CCARTCYTTYCTRAARTANGCI	22	41	yes	
Y3F-A	CCACTGTTGGTATGGGCAI	19	53		
Y3F-B	CACCACTGTTGGTATGGGI	19	53		
Y3F-C	GGCACCACTGTTGGTATGI	19	53		
Y3R-A	ATGCACCARACCATWAGCCCAI	22	48		
Y3R-B	GCACCARACCATWAGCCCATI	21	50		
Y3R-C	ACCARACCATWAGCCCATTCAI	22	43		
Y4F-A	GTNGTDGAYAAYTCYCTYATGGTI	24	41		
Y4F-B	CBGTNGTDGAYAAYTCYCTYATGI	24	44	no	1:00
Y4F-C	GTDGAYAAYTCYCTYATGGTYGTI	24	41		
M4	GGNAAYAAYAGYGGNCARCC	20	55		

* Melting temperature

Extension time

$ Usage or GC enhancer from Q5 Polymerase

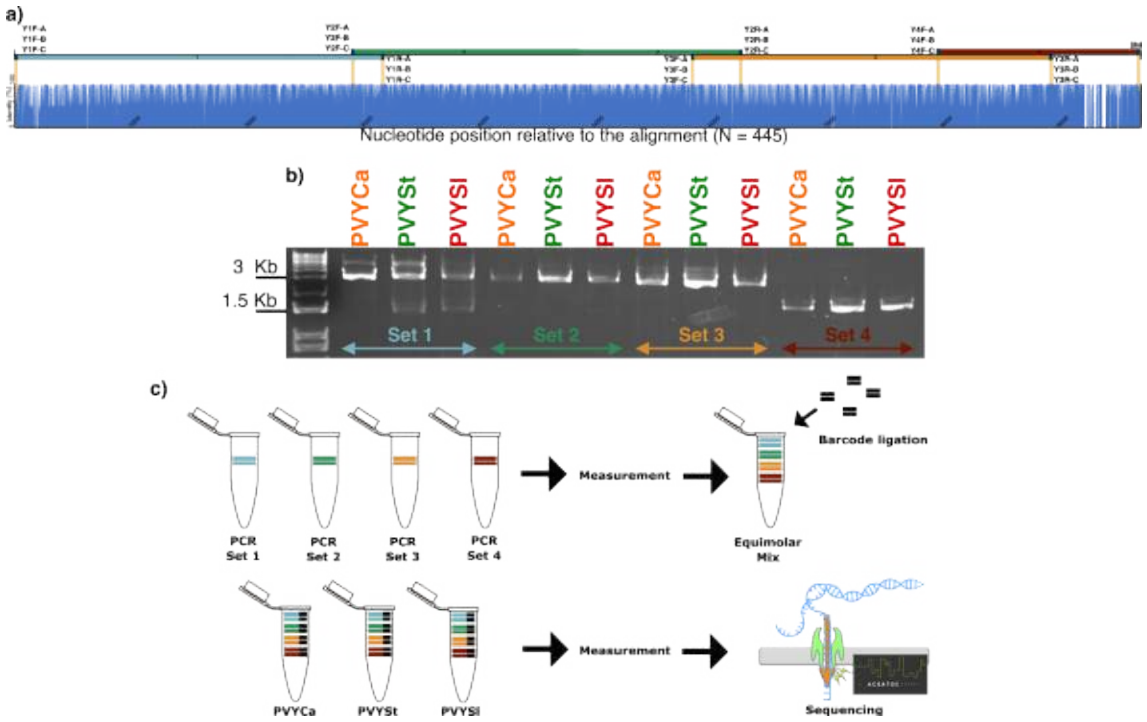


Fig 1. a) Position of primers designed along the PVY alignment (*n* = 445). Four sets of primers (indicated by different colors) were designed, each composed of 3 forward and 3 reverse primers with small differences, producing amplicons of ~3 kb. The exception is the set 4, for which the anchor primer was used for amplification, and produced an amplicon of ~1.8 kb. b) Agarose gel electrophoresis of the amplicons of the three isolates, PVYCa, PVYSt and PVYSl, using the 4 sets of primers (1, 2, 3, and 4). Each well has 1 µL of the PCR product. c) Schematic view of the sequencing strategy used to amplify the genome by Nanopore sequencing.

Nanopore sequencing results

This study evaluated the ONT sequencing performance based on the analysis of three PVY isolates. Key metrics such as read length, read quality, number of reads, total bases, and quality cutoffs were assessed to determine the sequencing efficacy and data quality for each sample. The sequencing procedure was conducted for a duration of 2 days, but the number of reads plateaued after approximately 16 hours.

We observed two predominant read lengths, approximately 1.8 kb and 3 kb, as expected (Fig. 2a-c). Both PVYCa and PVYSt sequences exhibited similar mean and median read lengths and qualities (Fig. 2a-b), whereas PVYSl showed shorter read lengths (Fig. 2c). All samples produced comparable base call quality scores (Fig. 2d).

Specifically, PVYCa had a mean read length of 2,561.5 bases, a median read length of 3,125.0 bases, a mean read quality of 11.0, and a median read quality of 12.1. PVYSt had a mean read length of 2,394.4 bases, a median read length of 3,112.0 bases, a mean read quality of 11.0, and a median read quality of 12.1. PVYSl, on the other hand, had a mean read length of 2,059.6 bases, a median read length of 1,871.0 bases, a mean read quality of 11.0, and a median read quality of 12.0.

Regarding the number of reads, read length N50, standard deviation (SD) of read lengths, and total bases, PVYSt yielded the highest values, followed by PVYCa and PVYSl. PVYCa produced 72,121 reads (Fig. 2e), with 66,698 assembled to the reference genome, a read length N50 of 3,157 bases (Fig. 2f), a read length SD of 1,021 bases, and a total base count of 184,736,289. PVYSt produced 94,479 reads (Fig. 2e), with 85,297 assembled to the reference genome, a read length N50 of 3,201 bases (Fig. 2f), a read length SD of 1,131.5 bases, and a total base count of 226,218,980. PVYSl produced 36,463 reads (Fig. 2e), with 27,668 assembled to the reference genome, a read length N50 of 3,127 bases (Fig. 2f), a read length SD of 1,248.7 bases, and a total base count of 75,100,253.

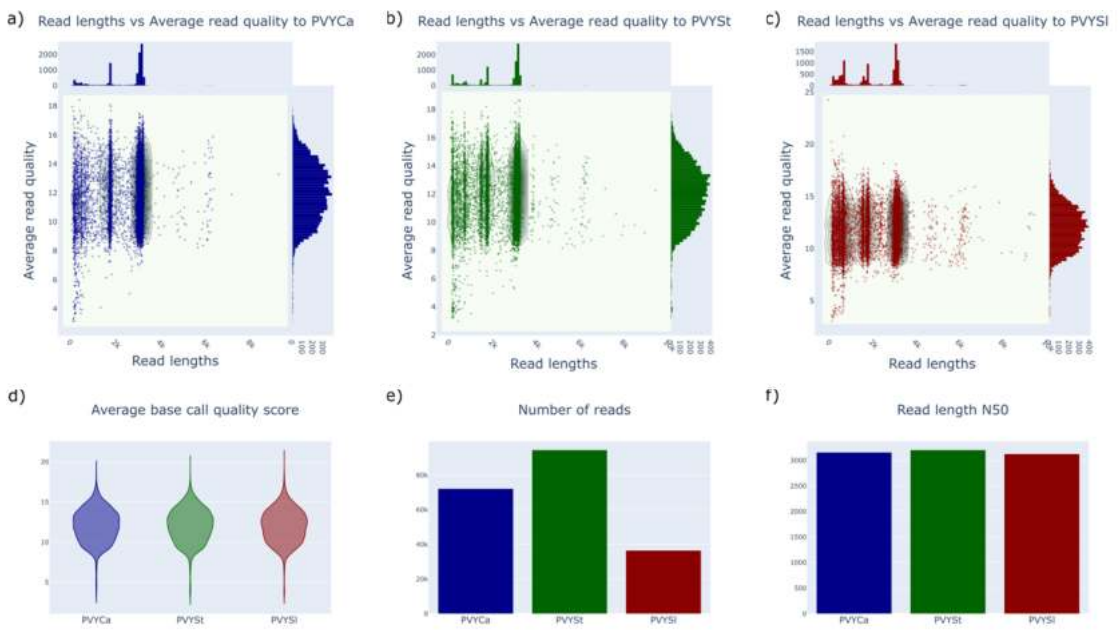


Fig 2. Scatter plot graph showing the distribution and clustering of the reads based on the length and quality for a) PVYCa (blue), b) PVYSt (green) and c) PVYSl (red).

Violin plot of the average base call quality score (d) for each library. Barplots showing the number of reads (e) and read length of N50 (f).

The longest read lengths and the distribution of reads above quality cutoffs (>Q10 (~90% of base-calling accuracy), >Q15 (~96.8% of accuracy), >Q20 (~99% of accuracy) are presented in Table 2. PVYCa had a longest read of 10,550 bases, with 59,824 reads (82.9%) above Q10, totaling 155.8 Mb, 4,527 reads (6.3%) above Q15, totaling 11.4 Mb, and 2 reads (0.0%) above Q20. PVYSt had a longest read of 12,601 bases, with 77,983 reads (82.5%) above Q10, totaling 190.9 Mb, 6,229 reads (6.6%) above Q15, totaling 15.1 Mb, and 4 reads (0.0%) above Q20. PVYSl had a longest read of 11,689 bases, with 29,943 reads (82.1%) above Q10, totaling 63.2 Mb, 2,299 reads (6.3%) above Q15, totaling 4.5 Mb, and 5 reads (0.0%) above Q20.

Among the reads longer than 9 kb, we identified 8 reads for PVYCa, 21 reads for PVYSt, and 21 reads for PVYSl. All the longest reads were compared to GenBank using BLASTn, showing high identity with PVY genomes (a list of the five longest reads with BLASTn results is available in Sup. Table 2). An exception was one PVYCa read of 9,666 bases, which mapped to *Xanthomonas euvesicatoria* (CP018467). This exception was likely due to residual contamination or a misclassification of a non-target sequence as part of the PVY dataset. Such anomalies highlight the importance of stringent quality control measures and careful analysis to ensure the accuracy of sequencing results and the reliability of data interpretation. Future efforts will focus on refining the sequencing protocol and enhancing the accuracy of read assignment to further minimize these issues.

Tab 2. Summary of Nanoplot results to sequenced reads of PVYCa, PVYSt and PVYSl using ONT Nanopore.

	Mean read lenght	**Mean read quality**	**Median read lenght**	**Median read quality**
PVYCa	2,561.5	11.0	3,125.0	12.1
PVYSt	2,394.4	11.0	3,112.0	12.1
PVYSl	2,059.6	11.0	1,871.0	12.0

	Number of reads	Read length N50	SD read length	Total bases
PVYCa	72,121.0	3,157.0	1,021.0	184,736,289.0
PVYSt	94,479.0	3,201.0	1,131.5	226,218,980.0
PVYSl	36,463.0	3,127.0	1,248.7	75,100,253.0
	Longest read	**>Q10***	**>Q15***	**>Q20***
PVYCa	10550	59824 (82.9%) 155.8Mb	4527 (6.3%) 11.4Mb	2 (0.0%) 0.0Mb
PVYSt	12601	77983 (82.5%) 190.9Mb	6229 (6.6%) 15.1Mb	4 (0.0%) 0.0Mb
PVYSl	11689	29943 (82.1%) 63.2Mb	2299 (6.3%) 4.5Mb	5 (0.0%) 0.0Mb

*Number, percentage and megabases of reads above quality cutoffs

The reads were mapped against the reference PVY genome, assessing coverage per base and genome coverage. Despite differences in coverage across the genome, we successfully reconstructed and assembled the entire genome for all isolates. Of the 184k bases produced for PVYCa sequencing, 169k were mapped against the reference genome, with coverage ranging from 79 to 8446 reads (Fig. 3a-b). For PVYSt, 197 of 226 kb were mapped, with coverage ranging from 123 to 8750 reads (Fig. 3c-d). For PVYSl, 64 of 75 kb were mapped, with coverage ranging from 366 to 8044 reads (Fig. 3e-f).

Ultimately, we reconstructed the consensus sequence for each isolate. The PVYCa consensus sequence was 9,699 bases long, with an ORF of 9,186 bases, a 5' UTR of 185 bases, and a 3' UTR of 328 bases. The PVYSt genome was 9,689 bases long, with a 5' UTR of 188 bases, an ORF of 9,173 bases, and a 3' UTR of 328 bases. The PVYSl genome was 9,699 bases long, with a 5' UTR of 185 bases, an ORF of 9,186 bases, and a 3' UTR of 328 bases. The 3’ UTR region excluded the polyadenylated sequence.

To determine the closest related genomes, a BLASTn analysis was performed against a reference database. The PVYCa and PVYSl sequences exhibited the highest identities (91.1% and 90.7%, respectively) with the Dutch PVY isolate from 1938 (EU563512) collected from potato plants. Additionally, PVYSt showed the highest

identity (98.3%) with a potato isolate from Russia obtained in 2021 (accession OR479975).

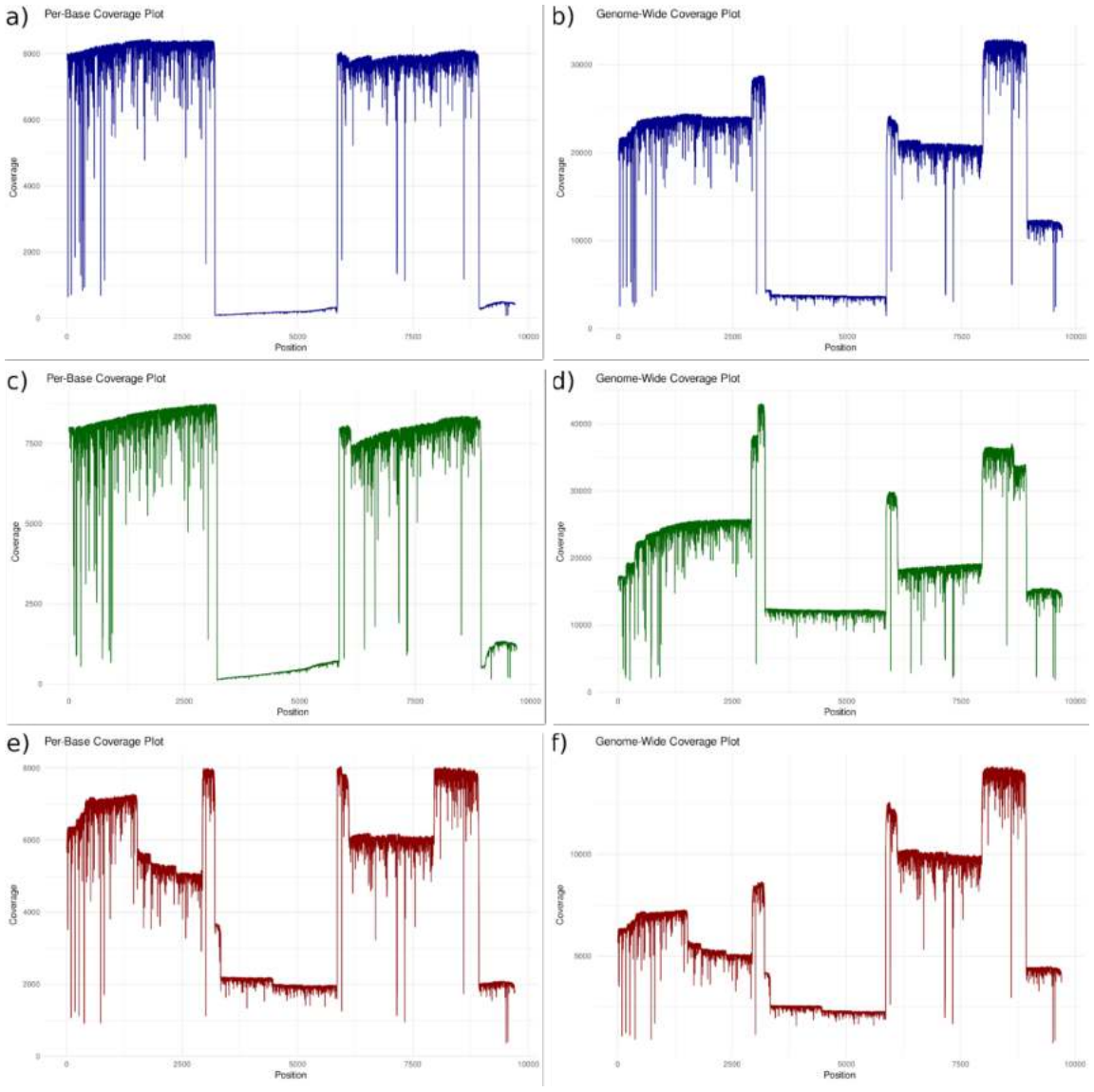


Fig 3. Number of Nanopore reads that covers each individual nucleotide position in the genome (per-base coverage, left side) to: (a) PVYCa, (c) PVYSt and (e) PVYSl; and number of sequencing reads covering each position across the entire genome (genome-wide coverage, right side) to (b) PVYCa, (d) PVYSt and (f) PVYSl.

In summary, we were able to determine the sequence of three PVY isolates, demonstrating consistent sequencing performance and data quality metrics such as read length, quality scores, and genome coverage. Despite variation in read lengths and coverage across isolates, complete genome reconstructions were achieved, validating

the efficacy of the sequencing approach. While we achieved satisfactory Q-scores, there remains room for improvement in future research to further enhance the overall sequencing accuracy. These findings underscore the reliability and utility of ONT sequencing in PVY infected populations.

Illumina sequencing results

While we were able to achieve complete genome coverage for all samples, we encountered challenges in obtaining an equal number of reads along the whole-genome. Starting from the same sample, PVYCa was the sample that presented the smaller genome coverage and number of reads. To evaluate the accuracy of the assembly, we employed HTS as a validation method for the genome with lower coverage obtained through Nanopore sequencing. A total of 57 million reads were generated through HTS, resulting in 8617 million bases. The quality assessment revealed a Q20 score of 97.6% and a Q30 score of 93.7%. After applying BBduk for trimming, we removed 49,000 reads (0.09%) or 766 million bases (8.67%), resulting in 57 million reads (7870 million bases) for contig assembly. The consensus sequence was constructed using Geneious assembler and has 9699 nt, a 5'UTR of 185 nt and ORF of 9186 and 3'UTR of 328 nt, exactly the same size and genome organization as the one constructed using Nanopore sequencing. It is important to mention that both PVYCa sequencings were done using the same sample, but to increase the number of viral particles to Illumina sequencing, a single mechanically passage was added using sweet pepper cv. Ikeda. About 30 plants were used to achieve the necessary weight of infected plants for semi-purification. On the other hand, Nanopore sequencing was done using the field collected sample.

Genome comparison and phylogenetics

We first calculated the pairwise distance between the assembled genomes with the PVY reference genome, shown in Fig 4.

Our results revealed that the identity between PVYCa and PVYSt was 99.77% ±0.007 (±1 SD). When compared to the reference PVY genome, both PVYCa sequences showed a identity of 99.82% ±0.006. In contrast, PVYSt was closer to the reference genome, with a identity of 99.86% ±0.005, suggesting it is genetically more similar to the reference sequence, which was isolated from potato. Interestingly, the

identity between both PVYCa sequences and PVYSl was the lowest at 99.92% ±0.003, indicating a high level of identity. However, PVYSt and PVYSl exhibited a identity of 99.77% ±0.007, similar to the divergence observed between PVYCa and PVYSt. The comparison between the PVY reference genome and PVYSl resulted in a identity of 99.82% ±0.006, similar to the distance between PVYCa and the reference genome. Furthermore, the comparison between PVYCa sequenced with Nanopore and with Illumina showed a low distance, with identity of 99.99% ±0.000.

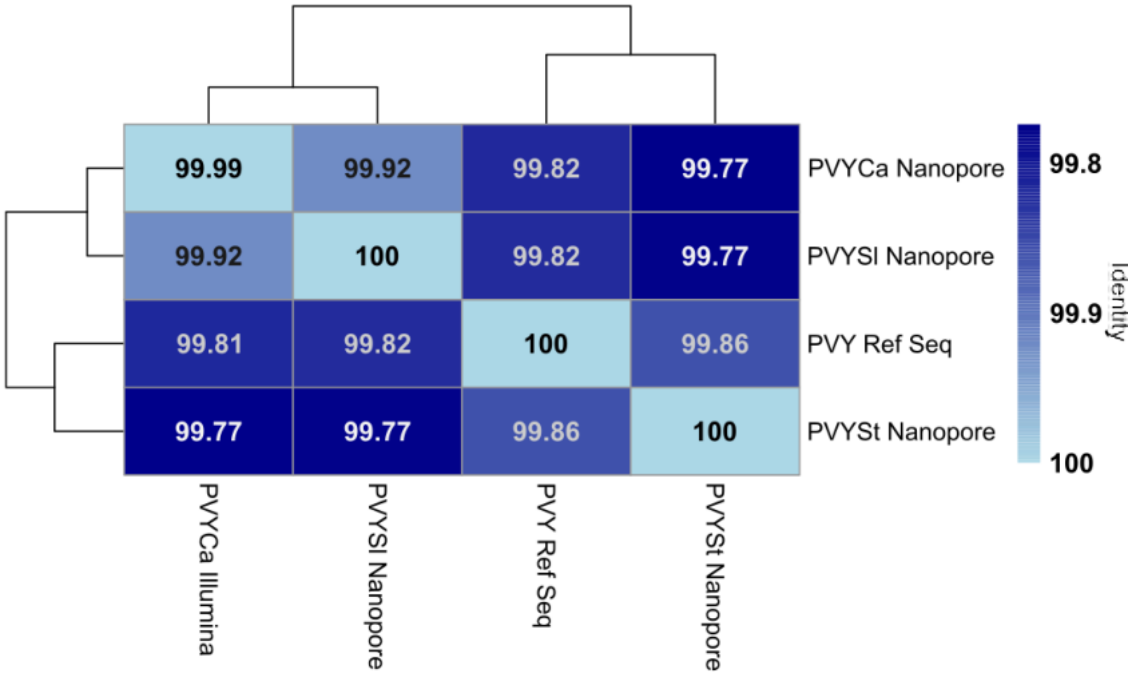


Fig 4. Pairwise distances (identity) between the four assembled genomes and the reference genome of PVY using the Tamura-Nei nucleotide substitution model with a Gamma distribution of sites.

The Pearson correlation results demonstrate a correlation coefficient of r = 0.99, which means there is an almost perfect positive correlation between the Illumina and Nanopore distances (Sup. Fig. 2). This indicates that the distances obtained from both methods are identical for these comparisons, reinforcing the claim of consistency and accuracy between the two sequencing technologies.

In summary, we demonstrated the relative accuracy of different sequencing technologies. The minimal genetic distance between PVYCa sequenced by Nanopore and Illumina underscores the reliability and consistency of both technologies in accurately detecting variations within the same viral sample, supporting their complementary use in comprehensive genomic analysis across different isolates.

For a further analysis, we utilized a representative dataset comprising 49 PVY isolates, in addition to BiMV, PSMV and SCMoV sequences, to reconstruct the ML-phylogenetic tree (Fig. 5). The phylogeny was quite consistent with the pairwise distance, as both PVYCa genomes were clustered in the same clade. PVYSl also appears to have a genetically close relationship with PVYCa, and with other isolates collected from tomato and pepper, primarily corresponding to strains C and O. On the other hand, PVYSt clustered with other isolates collected from potato and of strain N, revealing a separation influenced by the host.

Once again, the phylogeny highlights a close proximity between the two PVYCa genomes, underscoring the good sequencing capability of Nanopore.

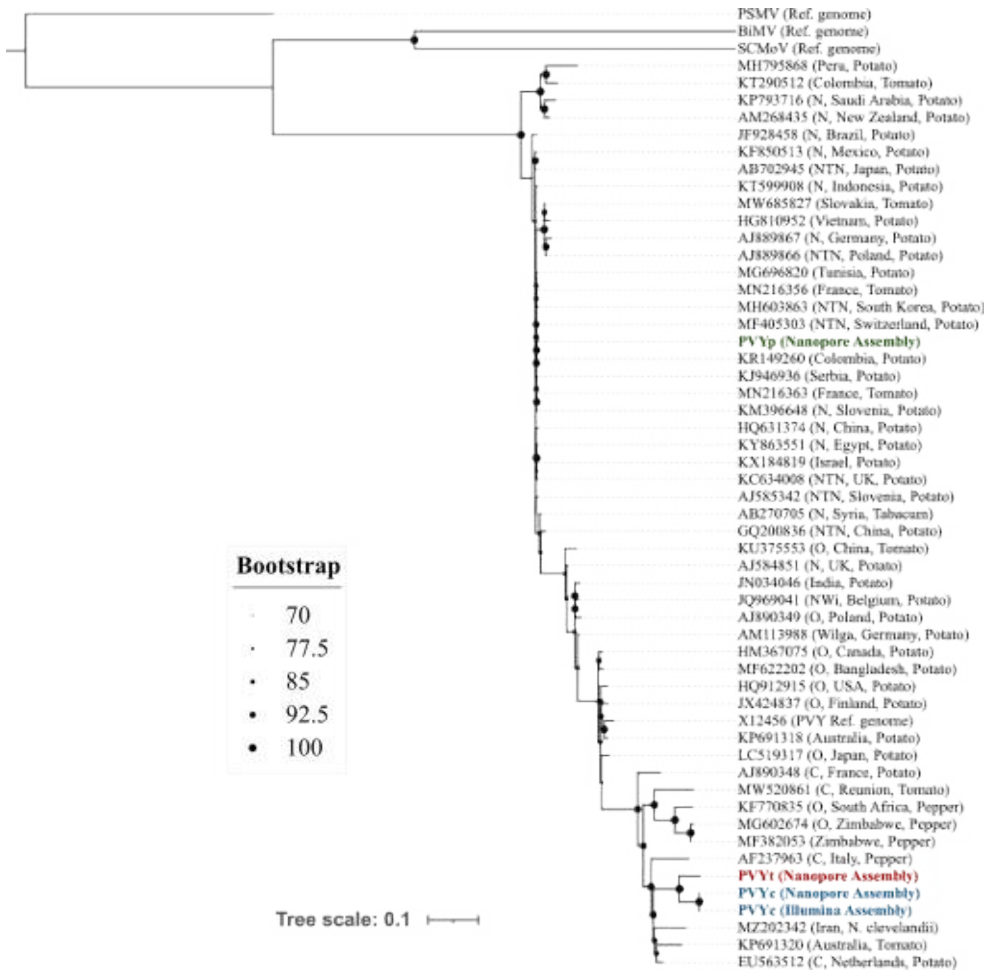


Fig 5. Maximum likelihood (ML) tree of 49 PVY isolates, the four assembled genomes (PVYCa, in blue, PVYSt in green and PVYSl in red) and three related viruses, BiMV, PSMV and SCMoV. Each isolate is represented by the GenBank accession, with the strain, country, and host of origin provided in parentheses, when available.

PVY-specificity primers

Capturing the whole variability in a highly variable virus, but without detecting other species, is a difficult task. The primers need to be specific to the virus but identify all possible variants within the populations arisen from evolution of this virus. For this purpose, we designed a pair or primers (UniYF and UniYR), able to detect any PVY isolate (Fig 6a). The pair of primers were verified by the BLASTn tool and the only hit was with PVY. We tested the pair of primers using PVYCa, -p and -t and common viruses such as GRSV, ToMV, PepYMV and PMMoV. We also included a non-infected plant and a negative control. It is important to note that PepYMV is a potyvirus and our

primers were not able to produce amplicons from this virus. The ideal PCR conditions with *Taq* DNA recombinant polymerase was using an initial denaturation at 95 °C for 1 min followed by 35 cycles of 95 °C for 30 s, 52 °C for 30 s, 72 °C for 1 min and a final extension with 72 °C for 10 min. An electrophoresis with agarose gel was used to visualize the PCR products. The primers were able to amplify only PVY samples (Fig 6b). Further analysis using other potyviruses is still necessary to validate the specificity of the designed primers.

a)

Primer	Sequence (5' -> 3')	Size	%GC
YSF	ACTATGATTTTTCGTCGAGAACAAG	25	36
YSR	GGCGAGGTTCCATTTTCAATGC	22	50

b)

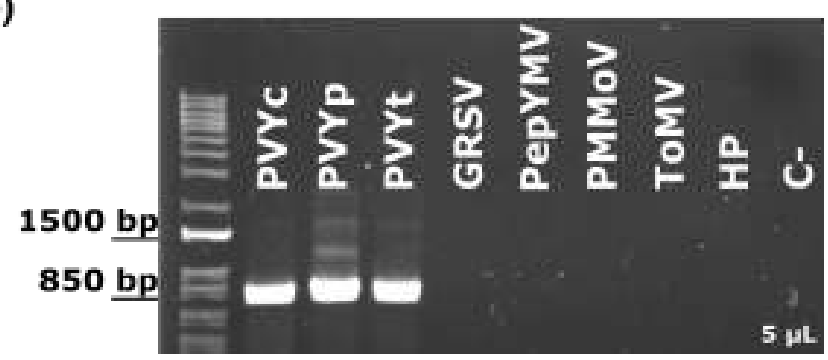


Fig 6. a) PVY-specific primers and PCR conditions using *Taq* DNA-recombinant polymerase. b) Agarose gel electrophoresis of the PCR amplified products using PVY-specific primers of the following templates: three PVY isolates, GRSV, PepYMV, PMMoV and ToMV. A healthy plant (HP) and a negative control (C-) were added to the analysis.

Discussion

In our investigation, we collected samples from symptomatic and asymptomatic plants, with a focus on develop a methodology to detect PVY in the most cultivated solanaceous plants. Despite the unbalanced plant species sampling, potatoes exhibited a higher infection rate (IR) compared to tomatoes, which is consistent with potatoes being

the most affected and studied host of PVY (Kreuze et al. 2020). We observed that symptom inspection is not a reliable method for confirming PVY infection, especially on tomatoes. It indicates that analysis of the samples by various detection tests is necessary for an accurate diagnosis, including genome sequencing.

Detecting and distinguishing different virus species in a sample is crucial for control strategies and genetic improvement programs (Du et al. 2006). However, detecting a single virus with high specificity can also be valuable. Various primers have been developed for PVY detection and classification (Moravec et al. 2003; Glais et al. 2005; Chikh Ali et al. 2010; Chikh-Ali et al. 2013). Our approach, using regular RT-PCR primers, differs by focusing on identifying PVY presence in samples using a conservative dataset. This methodology can be extended to other viruses with high genome variability and divergence. Although we sampled viruses commonly found in Brazilian tomato fields, only one potyvirus (PepYMV) was included, necessitating further validation with other potyviruses.

After collecting the plants in the field, we used dot-ELISA and Sanger sequencing prior to ONT sequencing to confirm the PVY infection. Serological tests are cost-effective and simple but prone to errors (Hühnlein et al. 2013), related to low sensitivity, and presence of molecules inducing false positive or false negative results. The RT-PCR method provides a sensitive, specific, and reliable diagnostic method (Malgosa et al. 2005; López et al. 2009), and thus it is considered one of the most widely used detection method. This method amplifies cDNA, which is then sequenced after amplicon purification. Although newer techniques allow for direct RNA and cDNA sequencing, PCR-amplified cDNA continues to be widely used RNA sequencing experiments (Bayega et al. 2018; Chen et al. 2021; Garalde et al. 2018). Our method could be advantageous for samples with low viral loads or highly divergent genomes, as it specifically amplifies the target virus, minimizing interference and background noise.

Traditional short-read sequencing technologies present important constraints, such as the difficulties in assembly of repetitive regions, which may cause structural variations due to the limitations of the short DNA fragments they analyze (Mak et al. 2016). This may result in fragmented genomes and potential biases in alignments (Huang et al. 2013). The efficiency of PCR amplification often decreases for long fragments, leading to smeared gel bands from unamplified truncated products. Illumina

can yield a broad overview of the sample composition but may suffer from low coverage or significant noise of non-specific reads. In contrast, our PCR-based approach reduces noise, and while non-specific reads were sequenced, they did not compromise sequencing confidence.

For our PCR approach, we divided the PVY genome into four segments. Similar amplification methods with a reduced number of genomes during the primer design and smaller amplicons have been used in other studies (Quick et al. 2017; Stubbs et al. 2020), typically following the Primal Scheme methodology (Quick et al. 2017). Multiplex RT-PCR can misidentify new genetic variants, especially rare and recombinant genotypes (Green et al. 2018). Our strategy, however, is more conservative and potentially covers all possible genome variations in an attempt to ensure no genomes are excluded. Despite differences in read numbers among the three sequenced samples, the final read quality was similar before and after base calling, indicating that read quantity does not necessarily correlate with better consensus assembly.

Although ONT was previously known for its high error rate (Rang et al. 2018), recent advances in base-calling algorithms have achieved consensus sequences with over 99.9% accuracy (Oxford Nanopore Technologies 2020; Chang et al. 2020). Our study, focusing on three PVY isolates, demonstrated that our methodology can accommodate up to 24 isolates for simultaneous identification and sequencing, facilitating whole-genome construction. We confirmed the identity of our isolates through BLASTn and phylogenetic analyses including closely related virus species as outgroups. This study also contributes to the limited datasets of PVY genomes from tomato and pepper, enhancing our understanding of host species' roles in PVY evolution. While HTS remains expensive and often inaccessible to small laboratories, we show that Nanopore sequencing is efficient, cost-effective when using barcode, and quicker, with simpler preparation (Petersen et al. 2019). We achieved long reads over 9 kb, some representing entire viral genomes. This capability enhances the accuracy of identifying complex repetitive or rearranged structures and facilitates the detection of a full spectrum of structural variations (Cretu Stancu et al. 2017; Gong et al. 2018). Although we only sequenced three isolates, our method can be applied to a broader range of viruses or organisms. Future work should focus on testing this approach on a larger number of PVY isolates.

Using pairwise distance and a phylogenetic approach, we were able to indirectly compare the assembled genomes, since both platforms use different bioinformatic pipelines. It is important to note that PVYCa and PVYSl are very similar. This genetic distance could be due the geographic barriers, since tomato and pepper plants are usually cultivated side by side, which can facilitate the movement of the virus between these two hosts. Differently, potato fields are often cultivated on large scale farms and away from other crops to avoid the movement of pests and diseases. But this does not appear to be the only cause, since phylogenetically, PVYCa and PVYSl tend to cluster with other isolates from pepper and tomato, and PVYSt is present among other isolates collected from potato. Thus, there appears to be an influence of the host on the evolutionary course of the virus, since the genetic distance between them is positively or negatively affected.

Diagnostic methods evolve for reliability, sensitivity and efficiency, meanwhile time and cost are two factors of great concerns. We present here two methods: (1) for a universal detection of PVY; and (2) for rapid sequencing the genome of PVY. The second method is particularly useful for small laboratories and for field studies, requiring minimal bioinformatics and computational skills, thereby reducing sequencing costs and training. Nanopore sequencing can be achieved with a reduced time and equipment costs (Lu et al. 2016; Petersen et al. 2019), offering versatility and simplicity. Studies have shown that Nanopore sequencing is more cost-effective than those provided by PacBio or Illumina platforms (Logsdon et al. 2020; Ranasinghe et al. 2022). Sequencing virus genomes is the basis for identifying genetic variation and study virus evolution. Overall, we experienced that the Nanopore technology, emerging from 2014, offered a powerful sequencing tool that required minimal preparation time and provided quick results. With the decrease in error rate in Nanopore sequencing (Oxford Nanopore Technologies 2020; Chang et al. 2020), this method is a valuable tool for understanding viral biology and evolution, with applications across various fields, including agricultural pest management.

Finally, we were able to detect and sequence three different PVY isolates, which show high genomic diversity among isolates. Moreover, both Illumina and Nanopore consensus assembly of PVYCa were highly similar, indicating the efficacy of our sequencing methodology. By sequencing a highly variable virus and finding results very

similar using Nanopore at a lower cost, we will be able to explore the technique for use in other virus-host systems. Although new tools are being generated to decrease the error rate of Nanopore sequencing, this method may have advantages when compared with Illumina if considering the fast result delivery (Garcia-Pedemonte et al. 2023). A direct comparison between Nanopore and Illumina is difficult to perform, but it is important to differentiate the bioinformatics skills required on both platforms, as Nanopore offers fewer steps, making the process simpler and easier.

As observed in Sup. Table 1, PVYSl was isolated from an asymptomatic tomato plant in the field, demonstrating that symptom inspection alone is insufficient for accurate PVY detection. Instead, a combination of various detection techniques, including PCR-based methods and Nanopore sequencing, provides a more reliable approach. Our sequencing methodology proved effective in capturing the PVY genome diversity, even with low viral loads or divergent genomes. This approach offers cost-effective, high-throughput sequencing with minimal preparation and bioinformatic skills, presenting a viable alternative to traditional methods like Illumina sequencing. Illumina still remains as the gold standard method of sequencing, but Nanopore sequencing may offer a reasonable performance and reliability.

References

Ahn HI, Yoon JY, Hong JS, et al (2006) The complete genome sequence of pepper severe mosaic virus and comparison with other potyviruses. Archives of Virology 151:2037–2045. https://doi.org/10.1007/s00705-006-0776-1

Amoia SS, Minafra A, Nicoloso V, et al (2022) A new jasmine virus C isolate identified by nanopore sequencing is associated to yellow mosaic symptoms of jasminum officinale in Italy. Plants (Basel) 11. https://doi.org/10.3390/PLANTS11030309

Bayega A, Wang YC, Oikonomopoulos S, et al (2018) Transcript profiling using long-read sequencing technologies. Methods in Molecular Biology 1783:121–147. https://doi.org/10.1007/978-1-4939-7834-2_6

Blawid R, Silva JMF, Nagata T (2017) Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. Annals of Applied Biology 170:301–314. https://doi.org/10.1111/aab.12345

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59–60. https://doi.org/10.1038/nmeth.3176

Chang JJM, Ip YCA, Bauman AG, Huang D (2020) MinION-in-ARMS: nanopore sequencing to expedite barcoding of specimen-rich macrofaunal samples from autonomous reef monitoring structures. Frontiers in Marine Science 7:448. https://doi.org/10.3389/FMARS.2020.00448/BIBTEX

Chen J, Chen J, Adams MJ (2001) A universal PCR primer to detect members of the *Potyviridae* and its use to examine the taxonomic status of several members of the family. Archives of Virology 146:757–766. https://doi.org/10.1007/s007050170144

Chen Y, Davidson NM, Wan YK, et al (2021) A systematic benchmark of Nanopore long read RNA sequencing for transcript level analysis in human cell lines. Nature Methods, 21, 1349-1363. https://doi.org/10.1101/2021.04.21.440736

Chikh-Ali M, Maoka T, Natsuaki KT, 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. https://doi.org/10.1016/j.jviromet.2009.12.010

Chikh-Ali M, Gray SM, Karasev AV. (2013) An improved multiplex IC-RT-PCR assay distinguishes nine strains of Potato virus Y. Plant Disease 97:1370–1374. https://doi.org/10.1094/PDIS-02-13-0161-SR

Cretu Stancu M, van Roosmalen MJ, Renkens I, et al (2017) Mapping and phasing of structural variation in patient genomes using nanopore sequencing. Nature Communications 8:1326. https://doi.org/10.1038/s41467-017-01343-4

Danecek P, Bonfield JK, Liddle J, et al (2021) Twelve years of SAMtools and BCFtools. Gigascience 10. https://doi.org/10.1093/gigascience/giab008

Davie K, Holmes R, Pickup J, Lacomme C (2017) Dynamics of PVY strains in field grown potato: impact of strain competition and ability to overcome host resistance mechanisms. Virus Research 241:95–104. https://doi.org/10.1016/J.VIRUSRES.2017.06.012

De Coster W, D'Hert S, Schultz DT, et al (2018) NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Deamer D, Akeson M, Branton D (2016) Three decades of nanopore sequencing. Nature Biotechnology 34:518–524. https://doi.org/10.1038/NBT.3423

Dong ZX, Lin CC, Chen YK, et al (2022) Identification of an emerging cucumber virus in Taiwan using Oxford nanopore sequencing technology. Plant Methods 18. https://doi.org/10.1186/S13007-022-00976-X

Du Z, Chen J, Hiruki C (2006) Optimization and Application of a multiplex RT-PCR system for simultaneous detection of five potato viruses using 18S rRNA as an internal control. Plant Disease 90:185–189. https://doi.org/10.1094/PD-90-0185

Dujovny G, Sasaya T, Koganesawa H, et al (2000) Molecular characterization of a new potyvirus infecting sunflower. Archives of Virology 145:2249–2258. https://doi.org/10.1007/s007050070018

Dujovny G, Usugi T, Shohara K, Lenardon S (1998) Characterization of a potyvirus infecting sunflower in Argentina. Plant Disease 82:470–474. https://doi.org/10.1094/PDIS.1998.82.5.470

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797. https://doi.org/10.1093/nar/gkh340

Edwardson JR, Christie RG (1997) Viruses infecting peppers and other solanaceous crops. Agricultural Experiment Station, University of Florida, Gainesville.

Funke CN, Nikolaeva OV, Green KJ, et al (2017) Strain-specific resistance to potato virus y (PVY) in potato and its effect on the relative abundance of PVY strains in commercial potato fields. Plant Disease 101:20–28. https://doi.org/10.1094/PDIS-06-16-0901-RE

Garalde DR, Snell EA, Jachimowicz D, et al (2018) Highly parallel direct RNA sequencing on an array of nanopores. Nature Methods 2018 15:3 15:201–206. https://doi.org/10.1038/nmeth.4577

Garcia-Pedemonte D, Carcereny A, Gregori J, et al (2023) Comparison of nanopore and synthesis-based next-generation sequencing platforms for SARS-CoV-2 variant monitoring in wastewater. International Journal of Molecular Science 24:17184. https://doi.org/10.3390/ijms242417184

Glais L, Tribodet M, Kerlan C (2005) Specific detection of the PVYN-W variant of Potato virus Y. Journal of Virology Methods 125:131–136. https://doi.org/10.1016/j.jviromet.2005.01.007

Gong L, Wong CH, Cheng WC, et al (2018) Picky comprehensively detects high-resolution structural variants in nanopore long reads. Nature Methods 15:455–460. https://doi.org/10.1038/s41592-018-0002-6

Green KJ, Brown CJ, Gray SM, Karasev AV (2017) Phylogenetic study of recombinant strains of Potato virus Y. Virology 507:40–52. https://doi.org/10.1016/J.VIROL.2017.03.018

Green KJ, Brown CJ, Karasev AV (2018) Genetic diversity of potato virus Y (PVY): sequence analyses reveal ten novel PVY recombinant structures. Archives of Virology 163:23–32. https://doi.org/10.1007/S00705-017-3568-X

Green KJ, Funke CN, Chojnacky J, et al (2020) Potato Virus Y (PVY) isolates from *Solanum betaceum* represent three novel recombinants within the PVYN strain group and are unable to systemically spread in potato. Phytopathology 110:1588–1596. https://doi.org/10.1094/PHYTO-04-20-0111-R

Hane DC, Hamm PB (1999) Effects of seedborne potato virus y infection in two potato cultivars expressing mild disease symptoms. Plant Disease 83:43–45. https://doi.org/10.1094/PDIS.1999.83.1.43

Huang L, Popic V, Batzoglou S (2013) Short read alignment with populations of genomes. Bioinformatics 29:i361–i370. https://doi.org/10.1093/BIOINFORMATICS/BTT215

Hühnlein A, Drechsler N, Steinbach P, et al (2013) Comparison of three methods for the detection of Potato virus Y in seed potato certification. Journal of Plant Diseases and Protection 120:57–69. https://doi.org/ https://doi.org/10.1007/BF03356455

Inoue-Nagata AK, Jordan R, Kreuze J, et al (2022) ICTV Virus Taxonomy Profile: *Potyviridae* 2022. Journal of General Virology 103:001738. https://doi.org/10.1099/jgv.0.001738

Inoue-Nagata AK, Oliveira PA, Dutra LS, Nagata T (2006) Bidens mosaic virus is a member of the potato virus Y species. Virus Genes 33:45–49. https://doi.org/10.1007/s11262-005-0037-5

Jeffries CJ (1998) FAO/IPGRI Technical guidelines for the safe movement of potato germplasm. Food and Agriculture Organization of the United Nations, Rome/International Plant Genetic Resources Institute, Rome, Italy 19.

Ji CM, Feng XY, Huang YW, Chen RA (2024) The applications of nanopore sequencing technology in animal and human virus research. Viruses 16:798. https://doi.org/10.3390/v16050798

Johnson M, Zaretskaya I, Raytselis Y, et al (2008) NCBI BLAST: a better web interface. Nucleic Acids Res 36. https://doi.org/10.1093/NAR/GKN201

Jones RAC (1990) Strain group specific and virus specific hypersensitive reactions to infection with potyviruses in potato cultivars. Annals of Applied Biology 117:93–105. https://doi.org/https://doi.org/10.1111/j.1744-7348.1990.tb04198.x

Karasev A V., Hu X, Brown CJ, et al (2011) Genetic diversity of the ordinary strain of Potato virus Y (PVY) and origin of recombinant PVY strains. Phytopathology 101:778–785. https://doi.org/10.1094/PHYTO-10-10-0284

Kerlan C, Moury B, Granoff A, Webster RG (2008) Encyclopedia of virology. Association of Applied Biologists 287–296.

Kolde R (2019) Pheatmap: pretty heatmaps version 1.0.12 from CRAN.

Kreuze JF, Souza-Dias JAC, Jeevalatha A, et al (2020) Viral diseases in potato. In: The Potato Crop. Springer International Publishing, Cham, pp 389–430.

Kumar S, Stecher G, Li M, et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology Evolution 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Lacomme C, Glais L, Bellstedt DU, et al (2017) Potato virus Y: biodiversity, pathogenicity, epidemiology and management. Springer International Publishing, Springer Cham. https://doi.org/10.1007/978-3-319-58860-5

Laver T, Harrison J, O'Neill PA, et al (2015) Assessing the performance of the Oxford Nanopore Technologies MinION. Biomolecular Detection and Quantification 3:1–8. https://doi.org/10.1016/J.BDQ.2015.02.001

Le Romancer M, Kerlan C, Nedellec M (1994) Biological characterization of various geographical isolates of potato virus Y inducing superficial necrosis on potato tubers. Plant Pathology 43:138–144. https://doi.org/10.1111/j.1365-3059.1994.tb00563.x

Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49:W293–W296. https://doi.org/10.1093/nar/gkab301

Li D, Liu C-M, Luo R, et al (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Logsdon GA, Vollger MR, Eichler EE (2020) Long-read human genome sequencing and its applications. Nature Reviews Genetics 21:597–614. https://doi.org/10.1038/s41576-020-0236-x

López MM, Llop P, Olmos A, et al (2009) Are molecular tools solving the challenges posed by detection of plant pathogenic bacteria and viruses? Current Issues Molecular Biology 11(1):13-46. https://doi.org/10.21775/cimb.011.013

Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinformatics 14:265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Mak ACY, Lai YYY, Lam ET, et al (2016) Genome-wide structural variation detection by genome mapping on nanochannel arrays. Genetics 202:351–362. https://doi.org/10.1534/GENETICS.115.183483

Malgosa A, Montiel R, Díaz N, et al (2005) Ancient DNA: a modern look at the infections of the past. In: Pandalai SG (ed) Recent research developments in microbiology, 1st ed. Trivandrum, pp 213–236.

Martins TP, Souza TA, da Silva PS, et al (2021) Nanopore sequencing of tomato mottle leaf distortion virus, a new bipartite begomovirus infecting tomato in Brazil. Archives of Virology 166:3217–3220. https://doi.org/10.1007/s00705-021-05220-w

Más A, López-Galíndez C, Cacho I, et al (2010) Unfinished stories on viral quasispecies and darwinian views of evolution. Journal of Molecular Biology 397:865–877. https://doi.org/10.1016/J.JMB.2010.02.005

Mikheyev AS, Tin MMY (2014) A first look at the Oxford Nanopore MinION sequencer. Molecular Ecology Resources 14:1097–1102. https://doi.org/10.1111/1755-0998.12324

Minh BQ, Schmidt HA, Chernomor O, et al (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 1;37(5):1530-1534 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Moravec T, Cerovská N, Boonham N (2003) The detection of recombinant, tuber necrosing isolates of Potato virus Y (PVY^{NTN}) using a three-primer PCR based in the coat protein gene. Journal of Virological Methods 109:63–68. https://doi.org/10.1016/S0166-0934(03)00047-8

Nolte P, Whitworth JL, Thornton MK, McIntosh CS (2004) Effect of seedborne Potato virus Y on performance of russet burbank, russet norkotah, and shepody potato. Plant Disease 88:248–252. https://doi.org/10.1094/PDIS.2004.88.3.248

Owczarzy R, Tataurov AV., Wu Y, et al (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Research 36:W163–W169. https://doi.org/10.1093/nar/gkn198

Oxford Nanopore Technologies (2020) New research algorithms yield accuracy gains for nanopore sequencing. In: https://nanoporetech.com/news/news-new-research-algorithms-yield-accuracy-gains-nanopore-sequencing

Peck KM, Lauring AS (2018) Complexities of Viral Mutation Rates. Journal of Virology 92. https://doi.org/10.1128/JVI.01031-17

Petersen LM, Martin IW, Moschetti WE, et al (2019) Third-generation sequencing in the clinical laboratory: exploring the advantages and challenges of nanopore sequencing. Journal of Clinical Microbiology 58. https://doi.org/10.1128/JCM.01315-19

Quick J, Grubaugh ND, Pullan ST, et al (2017) Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nature Protocols 12:1261–1266. https://doi.org/10.1038/nprot.2017.066

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ranasinghe D, Jayadas TTP, Jayathilaka D, et al (2022) Comparison of different sequencing techniques for identification of SARS-CoV-2 variants of concern with multiplex real-time PCR. PLoS One 17:e0265220. https://doi.org/10.1371/journal.pone.0265220

Rang FJ, Kloosterman WP, de Ridder J (2018) From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. Genome Biology 19. https://doi.org/10.1186/S13059-018-1462-9

Revers F, Le Gall O, Candresse T, et al (1996) Frequent occurrence of recombinant potyvirus isolates. Journal of General Virology 77:1953–1965. https://doi.org/10.1099/0022-1317-77-8-1953

Rizk MN, Ketta HA, Shabana YM (2020) Potential alternative hosts and transmissibility of potato virus Y. Journal of Plant Protection and Pathology 11:549–553. https://doi.org/10.21608/jppp.2020.133232

Rodriguez-Rodriguez M, Chikh-Ali M, Johnson SB, et al (2020) The recombinant Potato virus Y (PVY) Strain, PVY^{NTN}, identified in potato fields in Victoria,

Southeastern Australia. Plant Disease 104:3110–3114. https://doi.org/10.1094/pdis-05-20-0961-sc

Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, et al (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology Evolution 34:3299–3302. https://doi.org/10.1093/molbev/msx248

Sanjuán R, Domingo-Calap P (2016) Mechanisms of viral mutation. Cellular and Molecular Life Sciences 73:4433–4448. https://doi.org/10.1007/s00018-016-2299-6

Scholthof KBG, Adkins S, Czosnek H, et al (2011) Top 10 plant viruses in molecular plant pathology. Molecular Plant Pathology 12:938–954. https://doi.org/10.1111/j.1364-3703.2011.00752.x

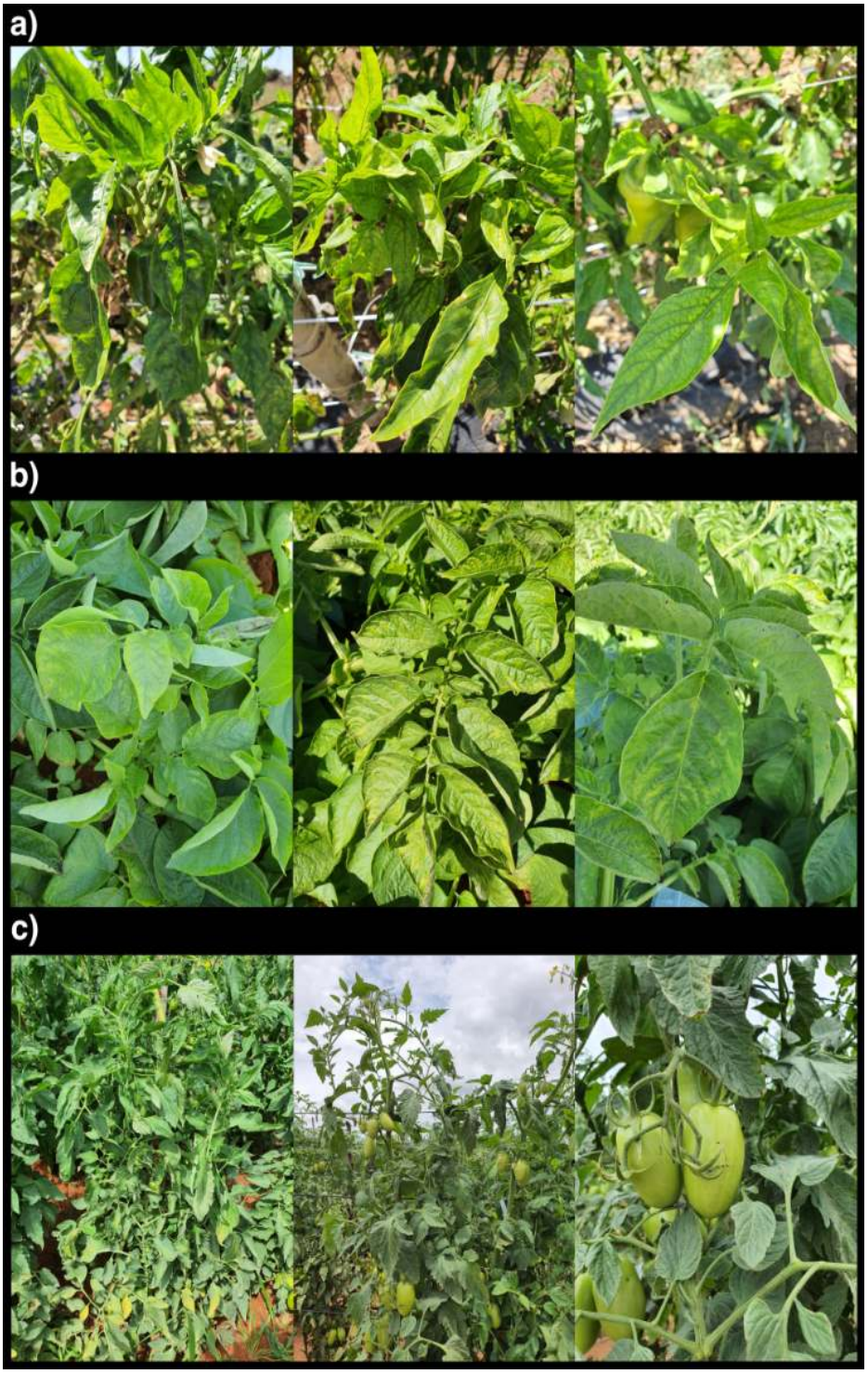
Singh RP, Valkonen JPT, Gray SM, et al (2008) Discussion paper: the naming of Potato virus Y strains infecting potato. Archives of Virology 153:1–13. https://doi.org/10.1007/s00705-007-1059-1

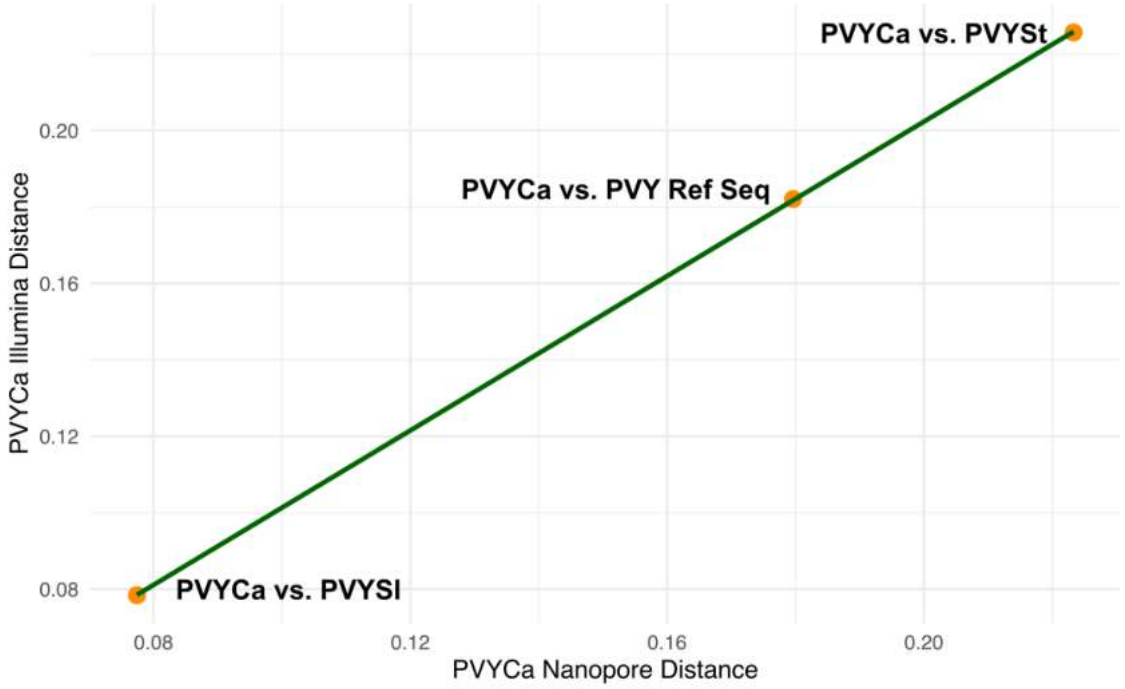
Steinhauer DA, Domingo E, Holland JJ (1992) Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene 122:281–288. https://doi.org/10.1016/0378-1119(92)90216-C

Stubbs SCB, Blacklaws BA, Yohan B, et al (2020) Assessment of a multiplex PCR and Nanopore-based method for dengue virus sequencing in Indonesia. Virology Journal 17:24. https://doi.org/10.1186/s12985-020-1294-6

Tromas N, Zwart MP, Maïté P, Elena SF (2014) Estimation of the in vivo recombination rate for a plant RNA virus. Journal of General Virology 95:724–732. https://doi.org/10.1099/vir.0.060822-0

Sup. Fig 1. Photograph of plant samples collected in the field: pepper plants (a) with blistering, mosaic and interveinal chlorosis; potato (b) plants showing chlorosis, mottling and necrotic spots; and tomato (c) plants with leafroll and necrotic spots.





Sup. Fig 2. Scatter plot comparing pairwise distances between PVYCa sequences obtained from Illumina and Nanopore sequencing technologies. Each point represents the distance between PVYCa and other PVY isolates (PVYSt, PVYSl, and the reference genome). The orange points indicate the pairwise distances, while the green line represents the linear regression fit (r = 0.99).

Sup. Table 1. Description of the samples.

Code	Plant species	Symptoms found	Collection date
1.1	Potato	Foliar chlorosis and necrotic spots	Aug 12, 2021
1.2	Potato	Foliar chlorosis	Aug 12, 2021
1.3	Potato	Foliar chlorosis	Aug 12, 2021
1.4	Potato	Foliar chlorosis	Aug 12, 2021
1.5	Potato	Foliar chlorosis	Aug 12, 2021
1.6	Potato	Foliar chlorosis	Aug 12, 2021
1.7	Potato	Foliar chlorosis	Aug 12, 2021
1.8	Potato	Foliar chlororis and necrosis	Aug 12, 2021
1.9	Potato	Foliar chlorosis	Aug 12, 2021
1.10	Potato	Nerval chlorosis	Aug 12, 2021
1.11	Potato	Necrotic spots	Aug 12, 2021
1.12	Potato	Foliar chlorosis	Aug 12, 2021
1.13	Potato	Foliar chlorosis and necrotic spots	Aug 12, 2021
1.14	**Potato**	**Foliar chlorosis and necrotic spots**	Aug 12, 2021
1.15	Potato	Foliar chlorosis and necrotic spots	Aug 12, 2021
1.16	Potato	Foliar chlorosis	Aug 12, 2021
1.17	Potato	Foliar chlorosis	Aug 12, 2021
1.18	Potato	Foliar chlorosis	Aug 12, 2021
1.19	Potato	Necrotic spots	Aug 12, 2021
1.20	Potato	Necrotic spots	Aug 12, 2021
1.21	Potato	Necrotic spots	Aug 12, 2021
1.22	Potato	Foliar chlorosis	Aug 12, 2021
1.23	Potato	Foliar chlorosis	Aug 12, 2021
1.24	Potato	Foliar chlorosis and necrotic spots	Aug 12, 2021
1.25	Potato	Necrotic spots	Aug 12, 2021
1.26	Potato	Necrotic spots and stunting	Aug 12, 2021
1.27	Potato	Necrosis	Aug 12, 2021
1.28	Potato	Mottle and stunting	Aug 12, 2021
1.29	Potato	Necrosis	Aug 12, 2021
1.30	Potato	Necrosis	Aug 12, 2021
1.31	Potato	Necrosis	Aug 12, 2021
1.32	Potato	Necrosis	Aug 12, 2021
1.33	Potato	Mottle and leaf distortion	Aug 12, 2021
1.34	Potato	Chlorosis	Aug 12, 2021
1.35	Potato	Necrotic spots	Aug 12, 2021
1.36	Potato	Veinal necrosis	Aug 12, 2021
1.37	Potato	no symptoms	Aug 12, 2021
1.38	Potato	Chlorosis	Aug 12, 2021
1.39	Potato	Chlorosis	Aug 12, 2021

1.40	Potato	Chlorosis	Aug 12, 2021
1.41	Potato	Necrotic spots and stunting	Aug 12, 2021
1.42	Potato	Necrotic spots and stunting	Aug 12, 2021
1.43	Potato	Leafroll, stunting and veinal chlorosis	Aug 12, 2021
1.44	Potato	Leafroll and necrotic spots	Aug 12, 2021
1.45	Potato	Necrotic spots	Aug 12, 2021
1.46	Potato	Necrotic spots	Aug 12, 2021
1.47	Potato	Chlorosis and crinkling	Aug 12, 2021
1.48	Potato	Necrotic spots	Aug 12, 2021
1.49	Potato	Necrotic spots and small leaves	Aug 12, 2021
1.50	Potato	Necrotic spots, stunting and small leaves	Aug 12, 2021
1.51	Potato	Necrotic spots and stunting	Aug 12, 2021
1.52	Potato	Necrotic spots	Aug 12, 2021
2.1	Tomato	Interveinal chlorosis	Aug 23, 2021
2.2	Tomato	Interveinal chlorosis and necrosis	Aug 23, 2021
2.3	Tomato	Interveinal chlorosis and necrosis	Aug 23, 2021
2.4	Tomato	Interveinal chlorosis	Aug 23, 2021
2.5	Tomato	Interveinal chlorosis and necrosis	Aug 23, 2021
2.6	Tomato	Interveinal chlorosis and small leaves	Aug 23, 2021
2.7	Tomato	Interveinal chlorosis	Aug 23, 2021
2.8	Tomato	Interveinal chlorosis	Aug 23, 2021
2.9	Tomato	Wrinkled leaves	Aug 23, 2021
2.10	Tomato	Interveinal chlorosis	Aug 23, 2021
2.11	Tomato	Interveinal chlorosis	Aug 23, 2021
2.12	Tomato	Stunting	Aug 23, 2021
2.13	Tomato	Leafroll and stunting	Aug 23, 2021
2.14	Tomato	Chlorosis, leaf deformation and necrotic spots	Aug 23, 2021
2.15	Tomato	Chlorosis, leafroll and stunting	Aug 23, 2021
2.16	Tomato	Leafroll and stunting	Aug 23, 2021
2.17	Tomato	Yellowing	Aug 23, 2021
2.18	Tomato	Chlorosis and yellowing	Aug 23, 2021
2.19	Tomato	Chlorosis	Aug 23, 2021
2.20	Tomato	Stunting	Aug 23, 2021
3.1	Pepper	Chlorosis and stunting	Aug 23, 2021
3.2	Pepper	Chlorosis, leafroll and stunting	Aug 23, 2021
3.3	Pepper	Leafroll	Aug 23, 2021
3.4	Pepper	Blistering and stunting	Aug 23, 2021
2.21	Tomato	Stunted growth	Aug 23, 2021
2.22	Tomato	Stunted growth	Aug 23, 2021
2.23	Tomato	Chlorosis, mottle, necrosis and yellowing	Aug 23, 2021
2.24	Tomato	Yellowing	Aug 23, 2021
2.25	Tomato	Interveinal chlorosis and necrotic spots	Aug 23, 2021

2.26	Tomato	Chlorosis	Aug 23, 2021
2.27	Tomato	Chlorotic and necrotic spots	Aug 23, 2021
3.5	Pepper	Blistering, foliar chlorosis and necrotic spots	Aug 23, 2021
3.6	Pepper	Blistering and foliar chlorosis	Aug 23, 2021
3.7	Pepper	Foliar chlororis and necrosis	Aug 23, 2021
3.8	Pepper	Blistering, foliar chlorosis and necrotic spots	Aug 23, 2021
3.9	Pepper	Interveinal chlorosis	Aug 23, 2021
3.10	Pepper	Blistering, foliar chlorosis and necrotic spots	Aug 23, 2021
3.11	Pepper	Mosaic, necrosis and veinal blistering	Aug 23, 2021
3.12	Pepper	Chlorotic spots and necrosis	Aug 23, 2021
3.13	**Pepper**	**Chlorosis, interveinal chlorosis and veinal blistering**	Aug 23, 2021
3.14	Pepper	Chlorotic spots and necrosis	Aug 23, 2021
3.15	Pepper	Chlorotic spots, small leaves and necrosis	Aug 23, 2021
3.16	Pepper	Chlorosis and interveial necrosis	Aug 23, 2021
3.17	Pepper	Chlorotic spots and necrosis	Aug 23, 2021
3.18	Pepper	Chlorosis, interveinal and necrotic spots	Aug 23, 2021
2.28	Tomato	Chlorosis, leafroll and necrotic spots	Aug 23, 2021
2.29	Tomato	Necrotic spots and stunting	Aug 23, 2021
2.30	Tomato	No symptoms	Aug 23, 2021
2.31	Tomato	No symptoms	Aug 23, 2021
2.32	Tomato	No symptoms	Aug 23, 2021
2.33	Tomato	No symptoms	Aug 23, 2021
2.34	Tomato	No symptoms	Aug 23, 2021
2.35	Tomato	No symptoms	Aug 23, 2021
2.36	Tomato	No symptoms	Aug 23, 2021
2.37	Tomato	No symptoms	Aug 23, 2021
2.38	Tomato	Necrosis	Aug 23, 2021
2.39	Tomato	Chlorosis	Aug 23, 2021
2.40	Tomato	Chlorotic spots	Aug 23, 2021
2.41	Tomato	Chlorosis	Aug 23, 2021
2.42	Tomato	Chlorotic spots	Aug 23, 2021
2.43	Tomato	Chlorosis	Aug 23, 2021
2.44	Tomato	Chlorosis	Aug 23, 2021
2.45	Tomato	No symptoms	Aug 23, 2021
2.46	**Tomato**	**No symptoms**	Aug 23, 2021
2.47	Tomato	Chlorosis	Aug 23, 2021
2.48	Tomato	Chlorosis	Aug 23, 2021
2.49	Tomato	No symptoms	Aug 23, 2021
2.50	Tomato	No symptoms	Aug 23, 2021
2.51	Tomato	No symptoms	Aug 23, 2021
2.52	Tomato	No symptoms	Aug 23, 2021

2.53	Tomato	No symptoms	Aug 23, 2021
2.54	Tomato	No symptoms	Aug 23, 2021
2.55	Tomato	No symptoms	Aug 23, 2021
2.56	Tomato	No symptoms	Aug 23, 2021
2.57	Tomato	No symptoms	Aug 23, 2021
2.58	Tomato	No symptoms	Aug 23, 2021
2.59	Tomato	No symptoms	Aug 23, 2021
2.60	Tomato	No symptoms	Aug 23, 2021
2.61	Tomato	No symptoms	Aug 23, 2021
2.62	Tomato	No symptoms	Aug 23, 2021
2.63	Tomato	No symptoms	Aug 23, 2021
2.64	Tomato	No symptoms	Aug 23, 2021
2.65	Tomato	No symptoms	Aug 23, 2021
2.66	Tomato	Chlorosis	Aug 23, 2021
2.67	Tomato	Chlorosis	Aug 23, 2021
2.68	Tomato	Chlorosis	Aug 23, 2021
2.69	Tomato	Interveinal chlorosis	Aug 23, 2021
2.70	Tomato	Chlorosis	Aug 23, 2021
2.71	Tomato	Chlorosis	Aug 23, 2021
2.72	Tomato	Chlorosis	Aug 23, 2021
2.73	Tomato	Chlorosis	Aug 23, 2021
2.74	Tomato	Leafroll	Aug 23, 2021
2.75	Tomato	Interveinal chlorosis	Aug 23, 2021
2.76	Tomato	Chlorosis and leafroll	Aug 23, 2021
2.77	Tomato	Chlorosis	Aug 23, 2021
2.78	Tomato	No symptoms	Aug 23, 2021
2.79	Tomato	Chlorosis	Aug 23, 2021
2.80	Tomato	No symptoms	Aug 23, 2021
2.81	Tomato	No symptoms	Aug 23, 2021
2.82	Tomato	No symptoms	Aug 23, 2021
2.83	Tomato	No symptoms	Aug 23, 2021
2.84	Tomato	No symptoms	Aug 23, 2021
2.85	Tomato	No symptoms	Aug 23, 2021
2.86	Tomato	No symptoms	Aug 23, 2021
2.87	Tomato	No symptoms	Aug 23, 2021
2.88	Tomato	No symptoms	Aug 23, 2021
2.89	Tomato	No symptoms	Aug 23, 2021
2.90	Tomato	No symptoms	Aug 23, 2021
2.91	Tomato	No symptoms	Aug 23, 2021
2.92	Tomato	No symptoms	Aug 23, 2021
2.93	Tomato	No symptoms	Aug 23, 2021
2.94	Tomato	No symptoms	Aug 23, 2021

2.95	Tomato	No symptoms	Aug 23, 2021
2.96	Tomato	No symptoms	Aug 23, 2021
2.97	Tomato	No symptoms	Aug 23, 2021
2.98	Tomato	No symptoms	Aug 23, 2021
2.99	Tomato	No symptoms	Aug 23, 2021
2.100	Tomato	No symptoms	Aug 23, 2021
2.101	Tomato	No symptoms	Aug 23, 2021
2.102	Tomato	No symptoms	Aug 23, 2021
2.103	Tomato	No symptoms	Aug 23, 2021
2.104	Tomato	No symptoms	Aug 23, 2021
2.105	Tomato	No symptoms	Aug 23, 2021
2.106	Tomato	No symptoms	Aug 23, 2021
2.107	Tomato	No symptoms	Aug 23, 2021
2.108	Tomato	No symptoms	Aug 23, 2021
2.109	Tomato	No symptoms	Aug 23, 2021
2.110	Tomato	No symptoms	Aug 23, 2021

Colored lines in red represent positive samples for PVY using dot-ELISA.
Bolded lines represent samples selected for sequencing.

Sup. Table 2. The longest reads for each virus isolate and the most closely related virus based on Blast analysis. All sequences shared high identity with PVY sequences.

Virus	Number	Read length	Identity (%)	BLASTn result
PVYCa	1	10,550	88.00	MT200665 [PVY]
PVYCa	2	9,484	91.18	EU563512 [PVY]
PVYCa	3	9,364	89.39	MT200665 [PVY]
PVYCa	4	9,341	88.66	MT200665 [PVY]
PVYCa	5	9,199	91.92	OM056939 [PVY]
PVYSt	1	12,601	94.28	OR479975 [PVY]
PVYSt	2	12,535	95.17	KX756672 [PVY]
PVYSt	3	12,343	95.00	OR480043 [PVY]
PVYSt	4	10,430	88.15	EU563512 [PVY]
PVYSt	5	9,741	95.28	KX756672 [PVY]
PVYSl	1	11,689	86.22	EU563512 [PVY]
PVYSl	2	10,186	86.82	MT200665 [PVY]
PVYSl	3	9,899	90.39	MT200665 [PVY]
PVYSl	4	9,835	89.88	MT200665 [PVY]

PVYSI	5	9,535	94.32	OR479975 [PVY]

954

Chapter III

Experimental evolution of host range for two isolates of *Potyvirus yituberosi*

Abstract

Potato virus Y (PVY) is a highly diverse and adaptable plant pathogen with a significant economic impact on solanaceous crops. This study investigates the evolutionary dynamics and host-specific adaptation of two PVY isolates (PVYNb and PVYSt) across three plant species (benthamiana, potato and tomato) through a series of experimental infections and serial passages. Transmission efficiency, viral load, and host-specific adaptation were assessed over 10 mechanical passages using RT-qPCR and high-throughput sequencing. Results demonstrated significant differences in infection efficiency between the two viral strains and across the three host species. PVYNb exhibited higher overall infection efficiency, particularly in benthamiana, whereas PVYSt showed limited infectivity, especially in tomato. Host-dependent variations were observed, with *N. benthamiana* acting as a source host supporting high viral replication, while tomato frequently acting as a sink, hindering sustained infection. Serial passage experiments revealed fluctuating viral loads, with significant interactions between viral isolate, host species, and passage number influencing virus accumulation. Infectivity tests of evolved lineages indicated that viruses passaged in *N. benthamiana* were generally more infectious, whereas those evolved in tomato or mixed hosts showed reduced infectivity. Genome analysis revealed higher population variation in PVYNb that tends to specialize in specific hosts. In contrast, PVYSt has a generalist behavior with lower frequency of fixed SNPs. This study highlights the significant role of host species in shaping PVY adaptation, with implications for understanding virus evolution and developing effective management strategies for PVY in diverse agricultural systems.

Keywords: Emerging virus; *Potyvirus*; Source-sink dynamics; Virulence; Virus evolution

Introduction

Viruses of the *Potyviridae* family are amongst the most prevalent plant pathogens. Specifically, potato virus Y (PVY; species *Potyvirus yituberosi*, genus *Potyvirus*, family *Potyviridae*) was once ranked as the fifth most important plant virus (Scholthof et al. 2011). PVY has a positive-sense single-stranded RNA genome about 9.7 kb long, linked at the 5' end to a viral protein (VPg) and featuring a poly(A) tail at the 3' end (Shukla et al. 1994). It translates into a single polyprotein of roughly 3062 amino acids (Inoue-Nagata et al. 2022), which is further processed into 10 mature peptides. An additional peptide is translated from a small open reading-frame that results from a +2 read-through within the P3 cistron. The PVY genome is replicated by its own RNA-dependent RNA polymerase, NIb. Due to the low replication fidelity of NIb, potyviruses are known for their high mutation rates, with reports of 2.6×10^{-6} in turnip mosaic virus (de la Iglesia et al. 2012) and 2.9×10^{-5} in tobacco etch virus (Sanjuán et al. 2009) substitutions per site per replication event. Such high mutation rates result in the generation of a viral quasispecies population structure of closely related viral genomes that undergo constant genetic variation, competition, and selection of the most fit variants in specific environments (Domingo et al. 2012). The large size of viral populations facilitates competitive as well as cooperative interactions between genetic variants, resulting in a dynamic quasi-equilibrium distribution. The high diversity of PVY allows it to be classified into strains based on biological properties, symptoms in potato and tobacco hosts, and phylogeny.

Given its high genomic diversity and evolutionary potential, PVY is expected to easily adapt to new hosts and readapt when returning to previous hosts. However, it remains unclear why some PVY strains appear more adapted to certain plant species and what genomic alterations occur due to host changes. Plant species are likely one of the major drivers of virus evolution by exerting strong selective pressures. In turn, the virus exerts selective pressure on the host, leading to a continuous cycle of reciprocal coevolutionary adaptations, commonly referred to as the Red Queen hypothesis (Whitlock 1996). During coevolution, viruses acquire the ability to encode proteins and regulate various functions within the minimal length of RNA sequences (Belshaw et al.

2007), producing multifunctional proteins that play roles in viral infection, from genome replication to interaction with the host plant and vector transmission.

An important factor in viral biology is the range of species a virus can infect. Some plant viruses specialize in infecting a few host species, while many are generalists, capable of infecting multiple species across different taxonomic groups. PVY is classified as a generalist virus with a polyphagous vector, allowing infection in diverse hosts and persistence in the environment (Edwardson and Christie 1997; Jeffries 1998). PVY is a globally significant plant virus, affecting at least 495 species across 72 genera in 31 families (Edwardson and Christie 1997). It infects economically relevant solanaceous crops such as potato (*Solanum tuberosum*), tomato (*S. lycopersicum*), pepper (*Capsicum* spp.),and tobacco (*Nicotiana tabacum*) and is spread by at least 70 aphid species in a non-persistent manner (Kerlan et al. 2008).

Emerging viruses face significant ecological challenges. With low initial abundance and limited within-host fitness, their persistence depends on how and how often they are transmitted, which can be affected by the proximity of alternative host populations in space or time (Gandon et al. 2013). If transmission is insufficient, the virus population cannot sustain its growth, leading to extinction before it can genetically adapt to the new host (Morse 1995; Antia et al. 2003; Gandon 2004). Therefore, the transmission rate is a crucial component of fitness at the between-host scale. The initial persistence of a viral population is determined by its fitness at both within- and between-host scales. A new host species acts as a “source” if within-host growth compensates for the population bottleneck during transmission, and as a “sink” if the growth rate or transmission is too low for the population to sustain itself (Dennehy et al. 2006, 2007). When only a single host is available, the virus becomes a specialist, increasing replicative fitness in the new host but decreasing it in the original one, *i.e.*, antagonistic pleiotropy (Elena et al. 2009). However, antagonistic pleiotropy is not universal. For instance, tomato spotted wilt virus can adapt to new hosts and expand its host range through positive pleiotropy (Ruark-Seward et al. 2020). Some interactions and adaptations do not incur any cost to generalist viruses or generate fitness trade-offs between hosts (Bedhomme et al. 2012). Instead, ecological fitting, due to the phenotypic plasticity of the viral quasispecies, occurs when viruses colonize new niches without undergoing adaptive evolution (Peláez et al. 2021). Generalist viruses, due to

their broader host range, are more likely to cross species boundaries and infect new hosts (Woolhouse and Gowtage-Sequeria 2005).

Experimental evolution studies, in which the same virus isolate or genotype is repeatedly passed through different hosts (either various species within the host range or sequential hosts in the infection cycle), typically reveal a pattern of specialization (Elena 2017). Virus lineages evolved in one host tend to perform better in that host compared to lineages evolved in other hosts, though this often comes at the cost of reduced fitness in alternative hosts (Wallis et al. 2007; Agudelo-Romero et al. 2008; Bedhomme et al. 2012b; Hillung et al. 2014). Despite some studies performed on a few potyviruses, no work has yet been done using PVY. It remains unclear if fitness trade-offs across host species may arise in PVY, and what the limits of adaptation in different hosts are from a molecular and phenotypical perspective. To address these questions, we experimentally infected three host species under five distinct conditions with two PVY strains, originally isolated from different hosts, and performed an evolution experiment that spanned ten sequential passages. We measured fluctuations in viral titer using RT-qPCR and sequenced the genomes by high-throughput sequencing (HTS) at different passages. Additionally, we assessed the impact of both viruses on symptomatology and plant height, correlating genome modifications with virus evolution and virulence across generations.

Materials and methods

Plants and growth environment

In this study, we used three plant species in the passage experiments: *Nicotiana benthamiana*, *S. lycopersicum* cv. Marmande, and *S. tuberosum* cv. Kennebec. Plants were maintained in a growth chamber, with a light period of 16 h at 24 ℃ (LED tubes at PAR 90 - 100 µmol m^{-2} s^{-1}), a dark period of 8 h at 20 ℃, and 40% relative humidity.

Individual plants were transplanted into pots, with each pot containing two plants, except for the potato tubers, which were cultivated in separate pots. The soil substrate comprised a mixture of DSM WNR1 R73454 substrate (Kekkilä Professional, Vantaa, Finland), grade 3 vermiculite, and 3-6 mm perlite in a ratio of 2:1:1.

Prior to the experiment, the batch of potato tubers were tested by RT-PCR to ensure the absence of PVY infection. Infection of the tuber by other viruses is not expect, since all tubers used are certified as free from viral infections. The potato tubers were cut into two or three sections, each containing at least one eye, and submerged in a 2-ppm gibberellic acid solution for approximately one hour before planting. To standardize the experimental conditions, only one potato stem was retained for each plant before the inoculation process, with additional stems removed. This approach aimed to minimize variability and ensure the uniformity of the experimental setup.

Isolates, inoculation and collection

Throughout the experiments, we used two isolates of PVY: PVYSt from the N-Wi strain, which was collected in potato field and propagated in potato plants, and PVYNb that belongs to the O strain, which was maintained continuously in *N. benthamiana* through several generations. The strains were defined by visualizing the formation of clades in a maximum-likelihood (ML) phylogenetic tree constructed using iqtree2 (Minh et al. 2020), with 10,000 bootstraps. This analysis utilized a dataset of 447 representative PVY haplotypes downloaded from GenBank (download on the 25-12-2023), along with the consensus sequences of PVYNb and PVYSt (Sup. Fig. 1), determined by HTS, described below.

For mechanical inoculation, we utilized a phosphate inoculation buffer, pH 7, with 3% polyethylene glycol (PEG), and 1:100 diluted of 100mg Carborundum. For inoculation, 20 μL inoculum were deposited per leaf on two leaves per plant. Inoculation was done manually.

Ten days post-inoculation (dpi), the three superior leaves of the plants were harvested, excluding the inoculated leaves. Subsequently, these plant tissues were rapidly frozen in liquid N_2, powdered, and homogenized. All collected samples were preserved at −80 ℃ to maintain their molecular integrity and ensure the preservation of viral particles for subsequent analyses.

Primer design and RNA amplification

In our study, we employed two approaches depending on the objectives. The first aimed to quantify viral RNA using quantitative reverse transcription polymerase chain reaction

(RT-qPCR), while the second focused on detecting the virus using standard RT-PCR. We designed two sets of primers for these purposes.

The RT-PCR primer set targeted a PVY-specific region with the following sequences: Forward 5'-ACTATGATTTTTCGTCGAGAACAA-3' (Universal PVY Primer Forward, UYF) and Reverse 5'-CGCGAGGTTCCATTTTCAATGC-3' (Universal PVY Primer Reverse, UYR), as described in Chapter II. Total RNA extractions were performed using the NZY Plant/Fungi RNA Isolation Kit (Tech MB45601, NZYtech). RT-PCR was carried out using NZYSupreme One-step RT-qPCR Probe Master Mix 2x (NZYtech) under the following conditions: 50 ℃ for 20 min, 95 ℃ for 5 min, followed by 40 cycles of 95 ℃ for 5 s, 60 ℃ for 40 s.

On the other hand, the RT-qPCR primer set targeted a conserved region of the *CP* gene of the virus (qYF, 5'-CAATCACAGTTTGATACGTGG-3' and qYR 5'-GGCGAGGTTCCATTTTCAATGC-3') and a common housekeeping gene, the glyceraldehyde-3-phosphate dehydrogenase (GAPDH), highly conserved among plant taxa (Martin and Cerff 1986) (qGAPDHF 5'-CTGTAACCCCAYTCGTTGTC-3' and qGAPDHR 5'-GTKGTKTCMAMWGAYTTTGTKGG-3').

To generate the standard curves, a series of cDNA dilutions of PVY ranging from 50 ng/µL to 0.005 ng/µL was prepared. Each dilution was tested in triplicate. The standard curves were used to calculate the qPCR reaction efficiency and the accuracy of the quantification, utilizing the linear regression equations derived from the C_T values *versus* the logarithm of the initial RNA concentration. The amplification efficiencies (%) were calculated based on the slope (s) of the standard curves using the expression efficiency = $100\times(10^{-1/s} - 1)$. The efficiency of PVY primers for the amplification of the portion of PVYNb and PVYSt genome was 94% (R^2 = 99.9%) and 104% (R^2 = 99.1%), respectively. For the GAPDH primers, the amplification efficiency was 89% for benthamiana (R^2 = 99.3%) and tomato (R^2 = 98.4%), and 91% (R^2 = 99.9%) for potato (Sup. Fig. 2). Thus, the RT-qPCR method was validated for adequate quantification of PVY RNA in the plant samples.

RNA extraction for this set was conducted using Sigma STRN250 Spectrum Plant Total RNA Kit with DNAse treatment (Invitrogen TURBO DNA-free Kit AM1907). The samples were checked for concentration and quality using Nanodrop and normalized to 50 ng/µL. RT-qPCR was performed using qPCRBIO SyGreen 1-Step Go

Hi-Rox (PCR BIOSYSTEMS) with at least three replicates for each sample. The conditions included 45 ℃ for 10 min, 95 ℃ for 2 min, followed by 40 cycles of 95 ℃ for 5 s and 60 ℃ for 30 s. All RT-qPCR results were filtered based on quality, using C_T cut-off values of 5 and 35, and with a maximum deviation between replicates of 0.3. The data were analyzed using qRAT (Flatschacher et al. 2022). Viral loads were then obtained using the $\Delta\Delta C_T$ method (Schmittgen and Livak 2008).

Evolution experiment

Twenty evolution experiments were simultaneously initiated, each with a total of ten serial passages; half were started with PVYNb and the other half with PVYSt. Five treatments were tested differing in their host plant composition as follows: (Sl) viruses only inoculated to tomato plants; (St) only to potato plants; or (Nb) only to benthamiana plants; (Correlated temporal fluctuations, CTF) viruses alternating inoculations among the three host species in the tomato-potato-benthamiana sequence; and (MIX) at each passage, viruses were inoculated into a mixture of the three plant species at equal proportions. Two independent evolution lineages per treatment were generated (L1 and L2). At each passage, the host population size was 16 plants. A full experiment design can be seen in the Fig. 1, totaling 20 experimental lines, 10 for each PVY isolate.

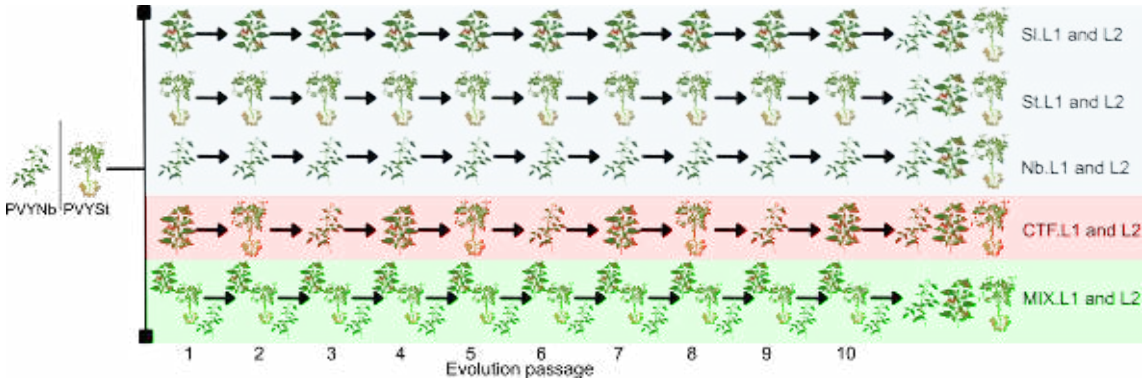


Fig 1. Schematic illustration of the passage experiment for analysis of the evolution of PVY according to the host. The founder population of two isolates, PVYNb from benthamiana and PVYSt from potato, were used to inoculate the 5 plant sets, in 2 replicates, totaling 10 Lines (Sl.L1, Sl.L2; St.L1, St.L2; Nb.L1, Nb.L2; CTF.L1, CTF.L2; MIX.L1, MIX.L2). The inoculation scheme for the passages in the same plant species are presented in the blue background, while the sequential scheme with

switching hosts is shown in red. In green, inoculations were performed in a mixture of different plants. Each plant in the scheme represents 16 inoculated plants, with MIX being the exception in which each plant represents 6 inoculated plant. The arrow represents mechanical inoculation from the previous passage. Using the last positive sample, 10 plants of each species were inoculated with the immediately preceding positive infection from the negative infection passage. A list of the last positive line is available at Sup. Table 1.

All evolving lineages underwent simultaneous inoculation on the same day, and the plant symptoms were daily monitored until 10 dpi. Then, the 16 plants were pooled and representative samples were collected. The tissue was powdered in liquid N_2, a portion used for inoculation of the next passage, and another portion was used for RNA extraction. Following quantification via relative RT-qPCR, only the positive lineages were continued. For those negative lineages, the inoculation process was repeated again using tissue from the previous positive passage to minimize potential inoculation error. Importantly, no previously negative lineage yielded a positive result after the second trial, ensuring the reliability of the experimental outcomes.

In each passage, six plants from each species served as mock controls, inoculated only with phosphate buffer. In addition, six plants of each species were employed as negative controls.

For the evaluation of disease phenotypic effects, the plant height was measured from the base of the plant to the apical meristem. Measurements were taken one day before inoculation and one day before collection (9 dpi).

At the end of passages and using the last positive passage available, all lineages were subjected to inoculation in the three different hosts, 10 plants of each. At this point, we applied an individual RT-PCR in order to detect the number of positive samples. Negative and mock controls were used during the inoculation and detection steps.

Transmission rate experiment

In the passage experiment, we recognized the potential for a loss of quantification accuracy due to pooling all 16 plants during collection and further inoculation. To

address this issue, we devised a transmission rate quantification experiment. In this setup, we inoculated 50 benthamianas, 50 potatoes, and 50 tomatoes using PVYNb or PVYSt. Each plant was individually collected to quantify the number of positive plants.

For this analysis, we employed an individual standard RT-PCR, and the samples were subjected to 1% agarose gel electrophoresis with SYBR green. Among the positive samples, we randomly selected three samples, with the exception of tomatoes infected with PVYSt, for which only two positive plants were obtained. Subsequently, each selected positive sample was used to inoculate another set of 50 plants for each species.

This process allowed us to assess the likelihood of a virus passing through the same host. Utilizing the same approach of RT-PCR and gel electrophoresis, we systematically analyzed the infection rates and dynamics within each host species. This individualized sampling strategy aimed to provide a more accurate and detailed understanding of virus transmission patterns among the different plant species.

HTS and sequence analyses

To determine the genome changes of the virus during the passage experiment, we employed Illumina HTS on three time points, in (1) the initial PVYNb and PVYSt inoculum source, in (2) the fourth passage (n = 12), and in (3) the latest available passage of each line (n = 17) (the list of the last available passage is available in Sup Table 1). Total RNA was extracted from fresh or dried leaf tissue using Sigma STRN250 Spectrum Plant Total RNA Kit (Invitrogen TURBO DNA-free Kit AM1907), then they were treated with DNase. The RNA concentration and the ratio absorbance at 260/230 and 260/280 nm quality was checked using Nanodrop before being sent for sequencing at Macrogen Inc. (Seoul, South Korea). GenTegra RNA GTR5001-S screwcap microtubes ensured secure sample storage and transportation during sequencing. Paired-end reads was prepared using TruSeq Stranded Total RNA Library Plant Kit library kit with TrueSeq Stranded Total RNA Reference Guide protocol in Illumina plataform.

To generate a consensus sequence for each sample (n = 31), reads were trimmed with BBDuk (https://sourceforge.net/projects/bbmap/), assembled with MEGAHIT (Martin and Cerff 1986; Li et al. 2015) and subjected to diamond blastx (Buchfink et al. 2015) searches against the non-redundant NCBI database (download on the

2024-06-24). The longest contigs identified as PVY for each sample were then subjected to BLASTn searches against the nucleotide database to identify their closest isolate (HM367076 and MW685829 for PVYNb and PVYSt, respectively). Then, reads were aligned against their respective reference sequence with BBMap v39.01 (https://sourceforge.net/projects/bbmap/) and a consensus sequence for each isolate was generated with Geneious Prime build 2022-03-15. Reads were aligned to the corresponding consensus sequence of PVY with BBmap with the vslow option. Base recalibration was then performed with GATK v4.0.5.1 (Van der Auwera and O'Connor 2020).

The diversity of each sequenced line was calculated as the sum of the Shannon entropy of each polymorphic site divided by the genome length using the aligned reads against the genome, resulting in a normalized quantity that varies between 0 (no polymorphic site) and 2 (all sites have an equal proportion of A, C, U and G). The genetic distance between population was calculated from the allele frequency difference (AFD) (Berner 2019). We also constructed the ML-tree using iqtree2 with 10,000 bootstrap replications with all consensus generate sequences.

Results

Test of transmission efficiency across host species

Initially, we examined the potential constraints imposed by the hosts on the adaptation of the PVY isolates and evaluated which of the three plant species may act as source or sink for the virus, by evaluating the probabilities of successfully infecting each of the three selected hosts with each of the two viral isolates. Fifty plants per host species were inoculated with the initial PVYNb and PVYSt isolates and analyzed individually by RT-PCR as illustrated in the experimental design shown in Fig. 2.

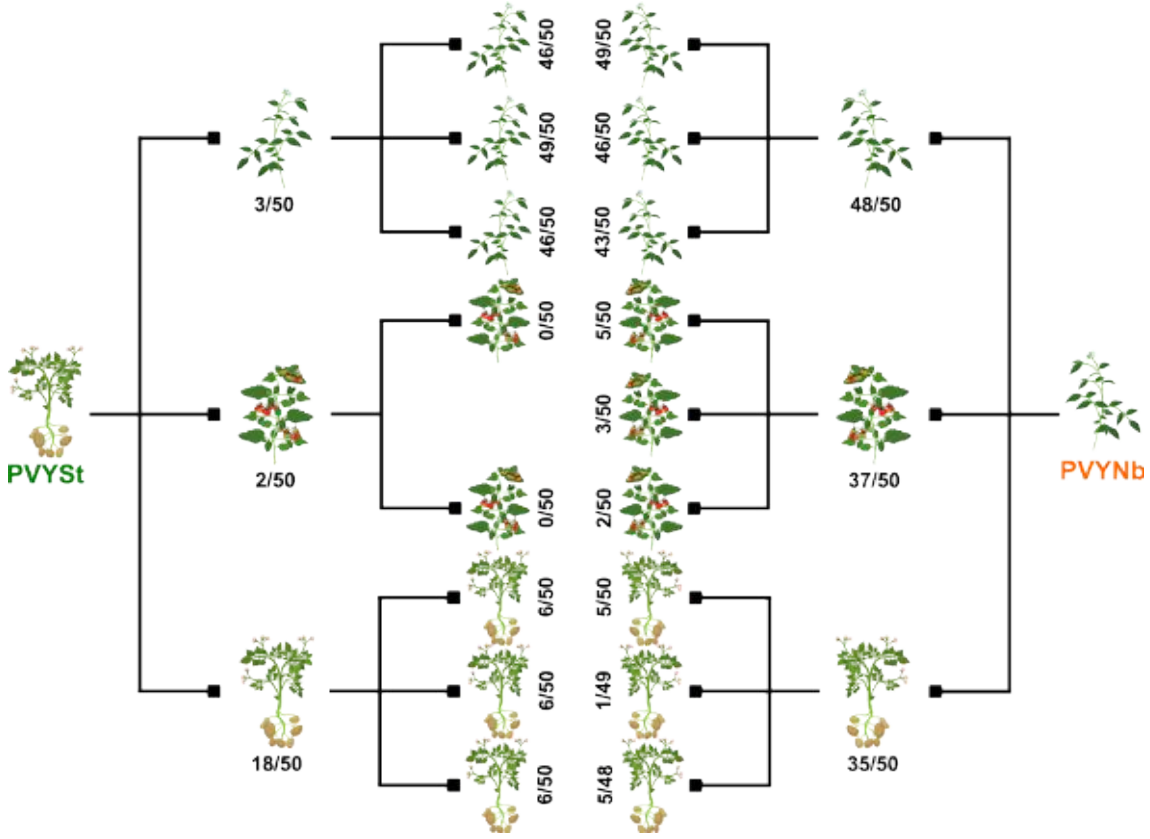


Fig 2. Infection rate evaluation according to the host, in which PVYSt and PVYNb were inoculated in 50 plants of each species and then reinoculated in the same host. The numbers below or at the side of each plant represent the infected and inoculated plants, respectively. The infection status of each plant was individually evaluated using RT-PCR.

Data shown in Fig. 3 were fitted to a logistic-regression using a generalized linear model (GLM) with a Binomial distribution probability and logit link function. Firstly, this analysis confirmed the high significant differences between the two viral isolates (χ^2 = 154.204, 1 d.f., P < 0.001), with PVYNb transmission efficiency across hosts being 0.844 ±0.037 (±1 SE), while it was 8.2-fold lower (0.103 ±0.030) for PVYSt. Differences in transmissibility among plant species was also observed (χ^2 = 9.821, 2 d.f., P = 0.007), with benthamiana (0.553 ±0.116) and potato (0.534 ±0.053) showing similar average PVY transmission efficiencies while tomato was about one-half less susceptible to infection (0.256 ±0.075). More interestingly, a significant interaction between viral isolate and host species was found (χ^2 = 30.608, 2 d.f., P < 0.001),

325 confirming that the transmission efficiency actually depended on the combination of
326 viral genotype and host species: in the first round of inoculation, maximum
327 transmissibility of PVYNb was shown in benthamiana (0.960 ±0.028) and minimum in
328 potato (0.700 ±0.065) while maximum transmissibility for PVYSt was observed in
329 potato (0.360 ±0.068) and minimum in tomato (0.040 ±0.028).

330 Following this first experiment, we randomly selected three positive plants
331 (except for PVYSt in tomatoes with only two positive samples), prepared new
332 independent inocula and reinoculated 50 plants of the same host. Differences among the
333 two viral isolates remained significant in this second infection (χ^2 = 4.001, 1 d.f., P =
334 0.045), although in this case the average transmission efficiency for the PVYNb-derived
335 samples was 0.287 ±0.037 but null for the PVYSt-derived ones. The transmission
336 efficiency results were strongly variable across host species (χ^2 = 707.822, 2 d.f., P <
337 0.001), being 0.931 ±0.015 for benthamiana, 0.094 ±0.017 for potato and null for
338 tomato. Finally, as observed in the first transmission trial, the outcome of this second
339 one also depended on the interaction between the origin of the inoculum and the host
340 species being inoculated (χ^2 = 12.770, 2 d.f., P = 0.002). In this second case,
341 PVYNb-derived inocula from benthamiana had a transmission efficiency of 0.920
342 ±0.022 in benthamiana, while inocula from tomato and potato were much less
343 transmissible when inoculated again in the same host (0.067 ±0.020 and 0.073 ±0.021,
344 respectively). Likewise, PVYSt-derived inocula from benthamiana also had a high
345 transmission efficiency in benthamiana (0.940 ±0.022), inocula from potato had a lower
346 transmissibility in potato (0.120 ±0.027) while inocula from tomato failed to be
347 transmitted to other tomatoes.

348 Based on the transmission efficiencies obtained at the first and second events, we
349 can now evaluate the sign and magnitude of the observed changes in transmission
350 efficiencies (Fig. 3). In the case of PVYNb, a host-species dependent reduction in
351 transmission efficiency has been observed (χ^2 = 6.946, 2 d.f., P = 0.031). While no
352 significant reduction was observed for benthamiana plants (0.960 ±0.028 *vs* 0.920
353 ±0.022; sequential Bonferroni *post hoc* test, P = 0.779), significant reductions were
354 observed for tomato (0.740 ±0.062 *vs* 0.067 ±0.020; P < 0.001) and potato (0.700
355 ±0.065 *vs* 0.073 ±0.021; P < 0.001). A host-dependent change in the transmission
356 efficiency between sequential inoculation experiments was observed for PVYSt (χ^2 =

123.735, 2 d.f., $P < 0.001$). In the case of benthamiana plants, transmission efficiency largely improved in the second transmission event compared to the first one (0.060 ±0.034 *vs* 0.940 ±0.019; $P < 0.001$). In sharp contrast, no change in efficiency was observed for tomato (0.040 ±0.028 *vs* 0.000 ±0.000; $P = 0.447$) and a significant reduction in potato (0.360 ±0.068 *vs* 0.120 ±0.027; $P = 0.006$) was found.

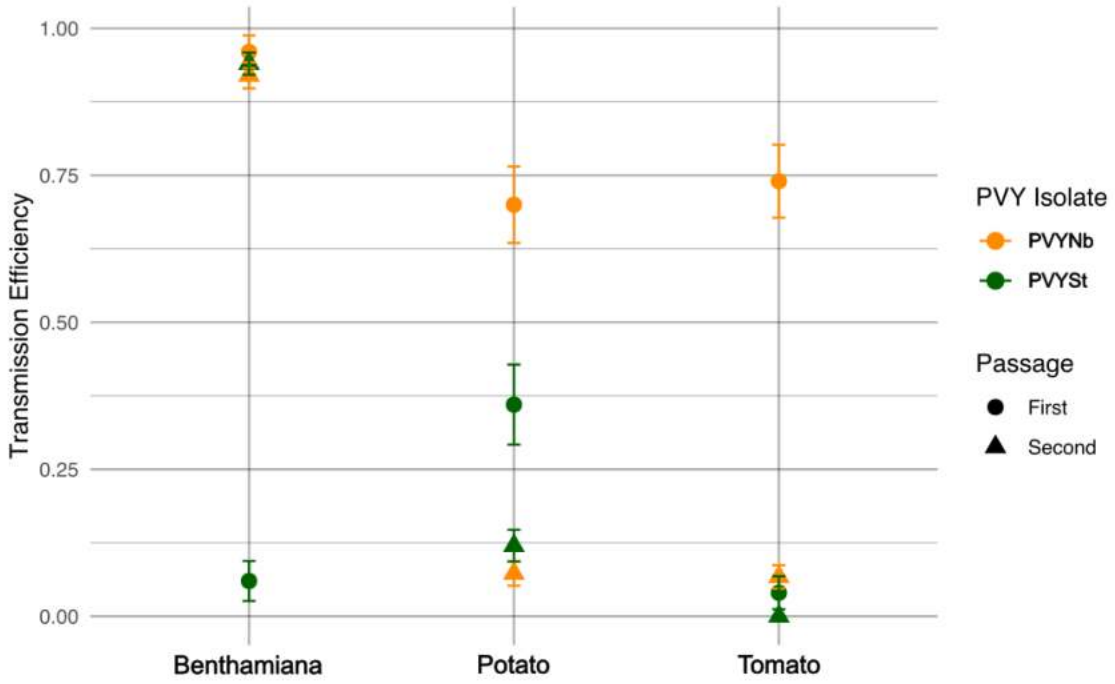


Fig 3. Transmission efficiencies of PVYNb and PVYSt in benthamiana, tomato and potato plants based on the results of the infection rate experiment during two sequential passages using the same host. The data were fitted to a logistic-regression using a generalized linear model (GLM) with a Binominal distribution probability and logit link function.

In conclusion, our study shows that the transmission efficiency of PVY isolates is highly dependent on both the viral genotype and host species. PVYNb exhibited higher transmissibility across all hosts compared to PVYSt, with benthamiana being the most susceptible. Sequential inoculations revealed that PVYNb transmission efficiency decreased significantly in potato and tomato, while remaining stable in benthamiana. In contrast, PVYSt improved its transmission in benthamiana but struggled in tomato and

potato. These results suggest that different plant species can act as either facilitators or barriers to viral adaptation and transmission.

Variation of viral loads along the passage through the plants

Our aim was to understand the host effect in the variation on the genome of PVY through serial passages. Closely related plants were used as test hosts, all in the family *Solanaceae*. In the passage experiment (Fig. 1), the isolates PVYSt and PVYNb were inoculated in five plant groups, tomato, benthamiana, potato, sequential switching hosts and mixed plants, through ten passages by mechanical inoculation. Viral load was taken as a proxy to within-host fitness. Fig. 4 shows the evolution of viral loads, measured by relative qPCR, for both PVY isolates under each of the five experimental host treatments. Data were fitted to a GLM with a Gamma probability function and a log-link function; viral isolate, experimental treatment and passage were included in the model as orthogonal factors and lineage was nested within the interaction of viral isolate and experimental treatment.

Firstly, a net effect of passage was observed (χ^2 = 3984.677, 9 d.f., $P < 0.001$) due to fluctuations and an overall tendency towards decreasing values in most of the conditions (Fig. 4). Secondly, net differences exist between both PVY isolates (χ^2 = 1691.114, 1 d.f., $P < 0.001$), with PVYNb, on average, accumulating orders of magnitude more than PVYSt (292.354 ±15.505 *vs*. 0.023 ±0.002, respectively).

Thirdly, differences exist between the five host treatments (χ^2 = 5682.249, 4 d.f., $P < 0.001$). Overall, the most permissive host for virus replication was benthamiana (Nb) (average accumulation 2.334 ±0.032 10^6) distantly followed by tomatoes (Sl) (161.143 ±18.076) and potatoes (St) (12.985 ±0.820). Viral loads estimated for the two mixed-hosts treatments (CTF and MIX) were low and hardly distinguishable from noise.

Interestingly, a significant interaction between PVY isolate and experimental conditions was observed (χ^2 = 2532.462, 4 d.f., $P < 0.001$), indicating that the accumulation of each isolate actually depended on the particular host in which it was measured. Furthermore, this effect also depends on passage number (χ^2 = 1498.356, 9 d.f., $P < 0.001$). For both isolates, benthamiana showed the highest accumulations, confirming its role as source host (2.750 ±0.049 10^6 for PVYNb and 1.981 ±0.041 10^6

408 for PVYSt). However, in the case of PVYNb CTF showed the second largest
409 accumulation (2.139 ±0.107 10^3), while for PVYSt in potato (St) ranked second (1.028
410 ±0.003 10^3). The lowest accumulation of PVYNb was observed in St (0.164 ±0.019)
411 while for PVYSt it was in CTF, and MIX values were indistinguishable from noise. In
412 tomatoes (Sl), PVYNb was unable to infect plants in the second passage after an initial
413 infection, being restricted to the first passage (Fig. 4). PVYSt also showed difficulty in
414 infecting tomatoes, but PVYSt/Sl.L1 was maintained in tomatoes for five serial
415 passages (Fig. 4). Interestingly, PVYSt/Sl.L1 showed an increase in viral RNA during
416 the passages, but after the peak of virus abundance, it failed to infect the plants in the
417 next passage. This result is in line with the evidences that tomato is a less permissive
418 sink host for the propagation of the two used PVY isolates. In potatoes (St), both PVY
419 isolates could infect systemically and be passed through at least five passages. The viral
420 RNA contents variation during passages did not follow a consistent pattern. PVYNb
421 lines infected potatoes for five passages, showing a similar infection pattern except for
422 the fourth passage, which decreased dramatically and resulted in absence of infection in
423 the next passage. PVYSt/St.L1 could infect potato plants for six passages and
424 PVYSt/St.L2 for ten passages. PVYSt/St.L2 exhibited a significant increase in virus
425 RNA after the eighth passage, suggesting adaptation to the host. In benthamiana (Nb),
426 both PVYNb/Nb.L1 and PVYNb/Nb.L2 and PVYSt/Nb.L1 and PVYSt/Nb.L2 exhibited
427 high viral RNA contents compared to other lines (Fig. 4). The viral RNA amount was
428 high from the first passage and remained high throughout all passages. The four
429 experimental lineages reached 10 passages. Both PVYNb lineages had a close detection
430 pattern and constant virus amount with minimal variation. In contrast, PVYSt lineages
431 displayed more variation but seemed to have adapted to the host in the last two
432 passages. For the host switching treatment (CTF), starting with tomatoes and then
433 passing to potatoes and benthamiana, PVYSt/CTF.L1 could not infect tomato plants,
434 while in PVYSt/CTF.L2 only the first passage contained infected plants, but failed to
435 infect potato plants in the next passage. PVYNb had an initial low infection ability in
436 tomatoes but it increased in subsequent passages with potatoes and benthamiana,
437 decreasing again when returning to tomatoes. PVYNb/CTF.L1 reached the fourth
438 passage but could not infect potatoes after being inoculated in tomatoes, while
439 PVYNb/CTF.L2 stopped at the eighth passage, unable to pass from potatoes to

benthamiana. In mixed host treatment (MIX), PVYSt/MIX.L1 was terminated at the second passage with a low viral load, and PVYSt/MIX.L2 caused no infection. PVYNb was detected until the third passage, with similar virus load in both lineages. Lineage PVYNb/MIX.L2 reached the ninth passage with significant variation between passages. It was unclear which host contributed to virus replication as plants were pooled during RNA extraction and virus detection.

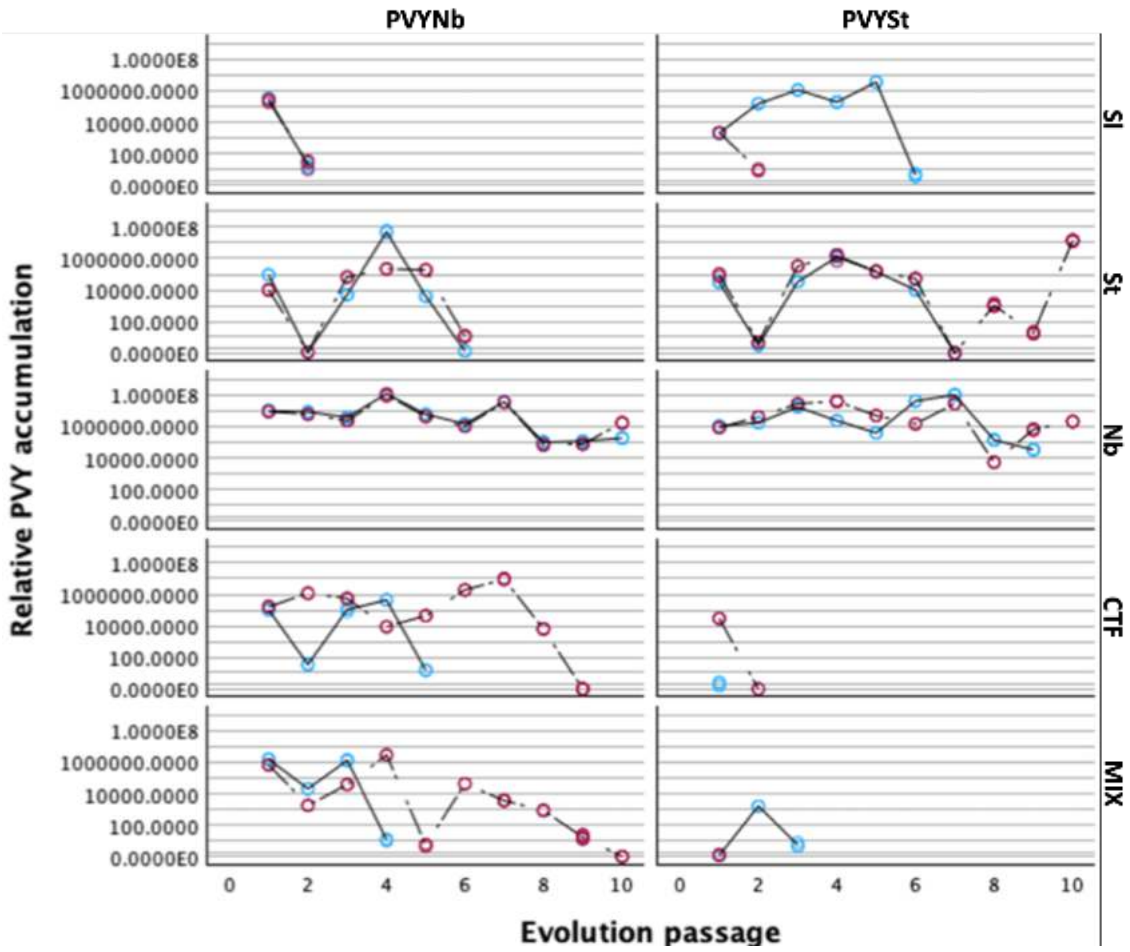


Fig 4. Relative quantification of the PVY RNA at each passage ($\Delta\Delta C_T$ method; see Materials and Methods section). Data are organized by PVY isolate (columns) and passage treatment (rows). Independent evolutionary lineages are indicated by different colors.

The observed effects of passages on virus titer did not necessarily represent an overall trend of increase or decrease in virus accumulation, but simply uncorrelated

455 significant differences among passages. Indeed, for lineages evolved in host
456 environments Sl (partial correlation coefficient controlling for viral strain and lineage: r
457 = 0.289, 34 d.f., P = 0.088), Nb (r = −0.121, 130 d.f., P = 0.167) and CTF (r = 0.156, 53
458 d.f., P = 0.254), there was no significant correlation between the number of the passage
459 and the virus titer. A weak positive yet significant correlation (r = 0.208, 108 d.f., P =
460 0.029) was found for viral lineages evolved in St, while a significant negative
461 correlation was found for lineages evolved in MIX (r = −0.360, 59 d.f., P = 0.004).
462 Based on this evolution experiment, we concluded that host species significantly
463 influence the pace of PVY adaptation. The permissiveness of different hosts varied
464 widely, with benthamiana plants demonstrating their role as source host supporting high
465 viral replication and efficient transmission across passages, while tomato plants acting
466 as sink hosts, often failing to support the virus beyond the initial passages. Host
467 switching revealed that initial low infection rates in less permissive hosts could improve
468 in more permissive hosts, and mixed host lines showed varied infection outcomes.
469
470 **Infectivity of evolved lineages depends on both the evolved environment and the**
471 **test host**
472 Using the last positive sample of PVYNb as inoculum, 10 plants of each host were
473 inoculated and their infection status determined by RT-PCR detection (Sup. Table 1).
474 Fitting these infectivity data to a logistic-regression, the analysis confirmed a significant
475 effect of the host compositions during serial passages (χ^2 = 35.722, 4 d.f., P < 0.001),
476 with viruses passaged in benthamiana being, on average, the most infectious, followed
477 by those evolved in potato. No net significant effect was observed for the host in which
478 the infectivity of the evolved lineages was tested (χ^2 = 0.000, 2 d.f., P = 1.000),
479 although a significant interaction existed between the evolved host and the test host (χ^2
480 = 33.915, 8 d.f., P < 0.001). PVYNb/Sl.L1 could not infect any hosts, while
481 PVYNb/Sl.L2 infectivity of benthamiana plants was 0.167 ±0.215 (LaPlace estimator of
482 the binomial parameter with 95% adjusted Wald CI). After five passages in potatoes,
483 PVYNb/St.L1 could not infect any hosts, but PVYNb/St.L2 infectivity in benthamiana
484 was 0.750 ±0.237, 0.833 ±0.215 in tomato and 0.250 ±0.237 in potato plants. After ten
485 passages in benthamiana, PVYNb/Nb.L1 infectivity in benthamiana was 0.750 ±0.237
486 and 0.917 ±0.142 tomato plants but no potatoes could be infected. PVYNb/Nb.L2

infectivity in benthamiana and tomatoes was 0.917 ±0.142 but dropped down to 0.167 ±0.215 in potatoes. After four passages, switching hosts and ending in tomatoes, PVYNb/CTF.L1 infectivity in benthamiana was zero, 0.500 ±0.263 in tomato and 0.333 ±0.252 in potato plants. PVYNb/CTF.L2, after eight passages, ending in potato plants, could not infect any hosts. After three passages, PVYNb/MIX.L1 infectivity in benthamiana was 0.833 ±0.215 and 0.750 ±0.237 in tomato plants but no potatoes, and PVYNb/MIX.L2, after nine passages, could not infect any hosts.

Following the same approach with the PVYSt isolate, we also found significant effects of the host composition during serial passages in the infectivity of the evolved viruses (χ^2 = 58.871, 4 d.f., P < 0.001), with again lineages evolved in benthamiana plants being the most infectious. However, for this viral isolate, significant differences among the three test hosts were observed (χ^2 = 17.708, 2 d.f., P < 0.001), with tomato showing more infected plants than the other two hosts. A significant interaction between evolution conditions and test host was also observed (χ^2 = 35.799, 8 d.f., P < 0.001). PVYSt/Sl.L1 could not establish infection in any host, while PVYSt/Sl.L2, after one passage, infectivity in benthamiana was 0.250 ±0.237 and 0.750 ±0.237 in tomato plants but null in potatoes. PVYSt/St.L1 reached six passages and could not infect benthamiana but had an infectivity of 0.417 ±0.261 in tomatoes and of 0.250 ±0.237 in potatoes. PVYSt/St.L2 reached ten passages but had an infectivity of 0.167 ±0.215 in potato plant not the other two hosts. In benthamiana, PVYSt/Nb.L1 and PVYSt/Nb.L2 reached ten passages. Lineage PVYSt/Nb.L1 infectivity in benthamiana was 0.833 ±0.215, 0.917 ±0.142 in tomato and 0.167 ±0.215 in potato plants, while lineage PVYSt/Nb.L2 infectivity in benthamiana was 0.917 ±0.142, 0.917 ±0.142 in tomato and 0.250 ±0.237 in potatoes plants. In switching hosts, only PVYSt/CTF.L2 achieved one passage and had infectivity 0.250 ±0.237 in tomato and in potato plants. Lineage PVYSt/MIX.L1 achieved the second passage and showed infectivity 0.167 ±0.215 in benthamiana and in potato plants.

All together, these observations suggest that benthamiana is the most permissive host for PVY strains evolution, allowing sustained virus replication across multiple passages. Potato served as an intermediate host with variable viral replication patterns, while tomato was the least permissive of the three tested hosts, often failing to support continued virus passages. Notably, PVYNb struggled to infect tomatoes beyond the

initial passage, while PVYSt showed limited but more sustained replication in certain lines. The mixed host lines indicated the complexity of host-virus interactions and potential adaptation mechanisms. These findings highlight the importance of host species in virus replication dynamics.

Changes in virulence and symptomatology

Next, we sought to evaluate the virulence of the PVY evolving lineages. As a first measure of virulence on each host species, we evaluated the effect of viral infection in plant growth relative to the mean growth of mock inoculated plants, $V = 1 - \frac{\Delta L_{infected}}{\langle \Delta L_{mock} \rangle}$, where L is the plant height measured at the time of inoculation and 9 dpi. $V < 0$ indicate a reduction in growth while $V > 0$ indicates infection enabled growth compared to noninfected plants. Virulence was measured after each of the 10 serial passages. Data are shown in Fig. 5. Data were fitted to a complex GLM with a Normal probability distribution and identity link function. Viral isolate, experimental passage, host treatment, and test host were used as orthogonal main factors, while experimental lineage was nested within the interaction between viral isolate and host treatment. Focusing in the main factors, overall significant differences were found between the two viral isolates (χ^2 = 4.680, 1 d.f., P = 0.031), being the mean virulence for PVYNb (0.040 ±0.029; ±1 SE) 8.6-fold larger than for PVYSt (0.005 ±0.084) and positive in both cases, suggesting plant elongation was a common symptom of infection. No main effects were associated for the host environment in which lineages evolved (χ^2 = 7.488, 4 d.f., P = 0.112) nor for the host in which virulence was tested (χ^2 = 3.542, 2 d.f., P = 0.170). However, significant differences among viral lineages evolved in a particular host environment and the host species in which virulence was evaluated have been found (χ^2 = 8.441, 2 d.f., P = 0.015), confirming that virulence indeed dependent on the interaction between viral strain, host environment and test host. For example, the largest reduction in growth induced by PVYNb infection was observed for lineages evolved in potatoes and tested in the same host (−0.133 ±0.078), while the smallest virulence was observed for lineages evolved in potato but tested in tomato (0.018 ±0.104). In contrast, in the case of PVYSt infections, the largest effect was observed for lineages evolved in tomatoes and tested in potato (−0.699 ±0.178) and the smallest for lineages evolved in

the mixed host populations and tested in the most permissive host, benthamiana (0.011 ±0.166).

Next, we decided to evaluate the possible effect of the source host on the evolution of symptoms in the most permissive host, *i.e.*, benthamiana. To do so, we monitored the presence or absence of symptoms in inoculated plants over 9 dpi. Mean time to the appearance of first visible symptoms was calculated using the Kaplan-Meier regression of the number of infected plants to days after inoculation. Fig. 6 shows the evolution of this mean time along the passage experiment. Remarkably, only the lineages evolved in benthamiana (Nb) plants, and that of MIX.L2 evolved in the mixed population were able to generate visible symptoms along all the passages. Indeed, for these lineages, a significant negative correlation exists between mean time to symptoms and passage number (partial correlation coefficient controlling for viral isolate and lineage: $r = -0.601$, 34 d.f., $P < 0.001$), indicating that symptoms tend to appear faster in benthamiana plants as the virus was evolving in this plant species. In other instances, symptoms appeared only sporadically (*e.g.*, lineage PVYNb/CTF.L2 recovered from benthamiana plants or early passages of PVYSt/MIX.L1), making additional statistical analyses unreliable.

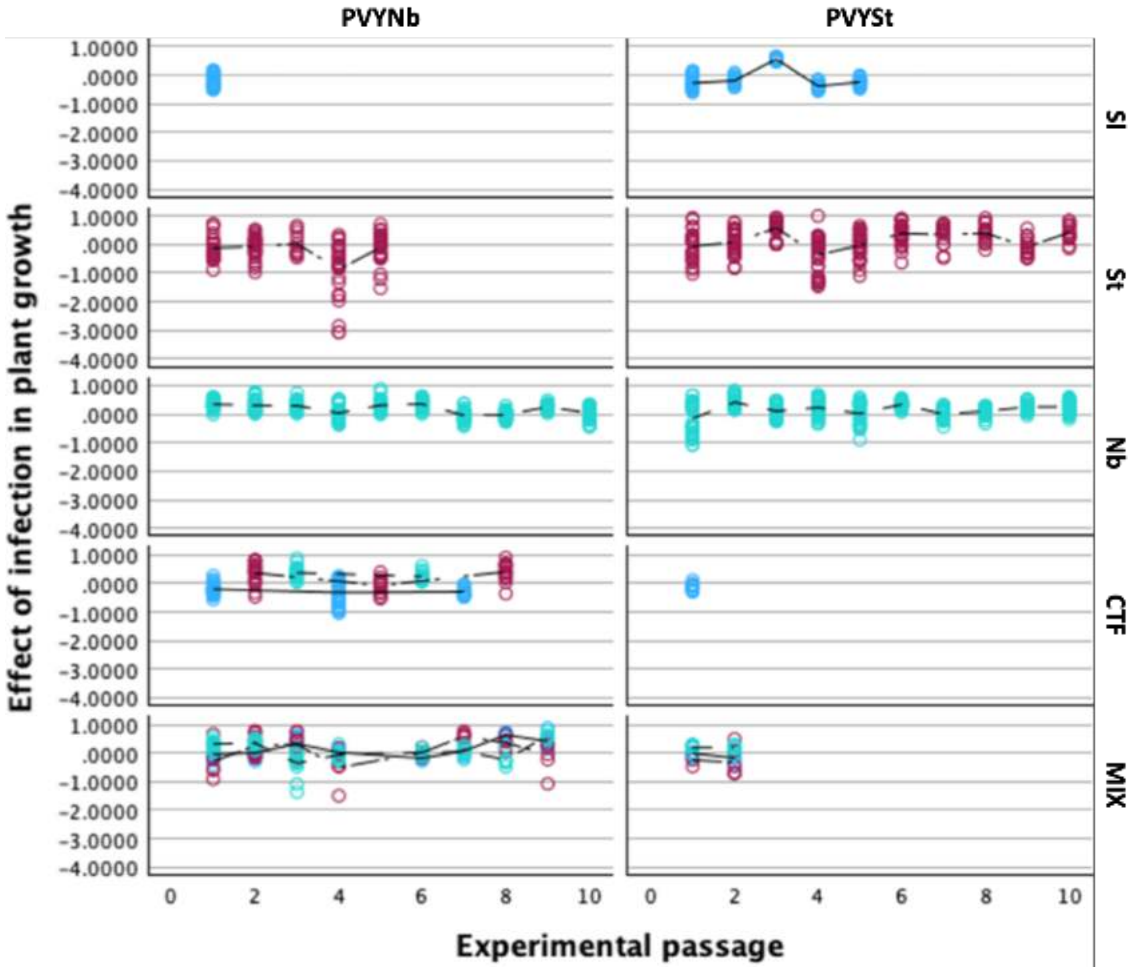


Fig 5. Evolution of virulence (relative effect of infection in plant growth). Data are organized by PVY isolate (columns) and passage treatment (rows). Plant species in which virulence was evaluated are indicated by colors: blue: tomato, red: potato, green: benthamiana. Viral lineages evolved in CTF were tested on the plant species corresponding to each passage. Viral lineages evolved in MIX were tested in all three plant species.

Taken together, these findings highlight the multifaceted nature of viral infection on plant growth, emphasizing the importance of considering various factors and their interactions in understanding the impact of viral infections on plant phenotypes.

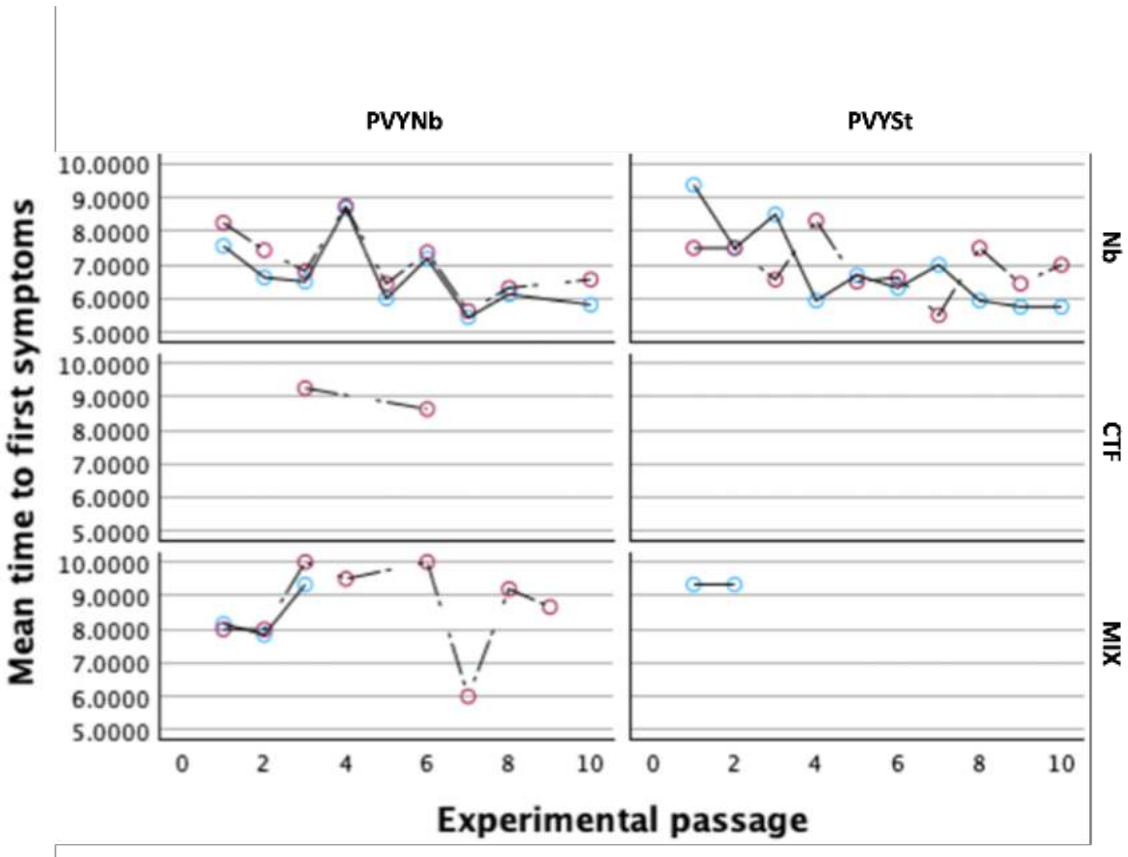


Fig 6. Evolution of the mean time for the appearance of the first symptoms in *N. benthamiana* plants inoculated with the different evolving lineages (indicated by colors), divided by treatment (Nb, CTF and MIX).

Genome alterations

We observed clear alterations in virus accumulation, infection rate and symptom induction throughout the passages, and now we were interested in understanding the genome modifications observed in selected time points. Sequencing was based on HTS from the original isolates, then those of the 4th passage and the latest viable passage for all the 20 lines. Firstly, the coverage along the genome differed between samples (Sup. Fig. 3). Five samples did not meet our threshold of 50× average coverage across the genome and were excluded from further analysis due to potential biases. The excluded samples were PVYNb/St.L1 and PVYNb/St.L2 (both 5th passage), PVYNb/CTF.L2 (8th passage), PVYSt/Sl.L1 (4th passage), and PVYSt/Sl.L1 (5th passage). Their low

coverage could lead to reduced accuracy in variant calling and lower confidence in quantitative analyses, thereby introducing uncertainty.

Using the reliable dataset, we calculated genetic differentiation between populations using allele frequency difference (AFD) (Fig. 7). The AFD measures the genetic difference between populations and how these changes are related to viral fitness. In our case, AFD allows to compare both PVY isolates under similar environmental conditions, helping to track if the population is undergoing genetic drift or natural selection.

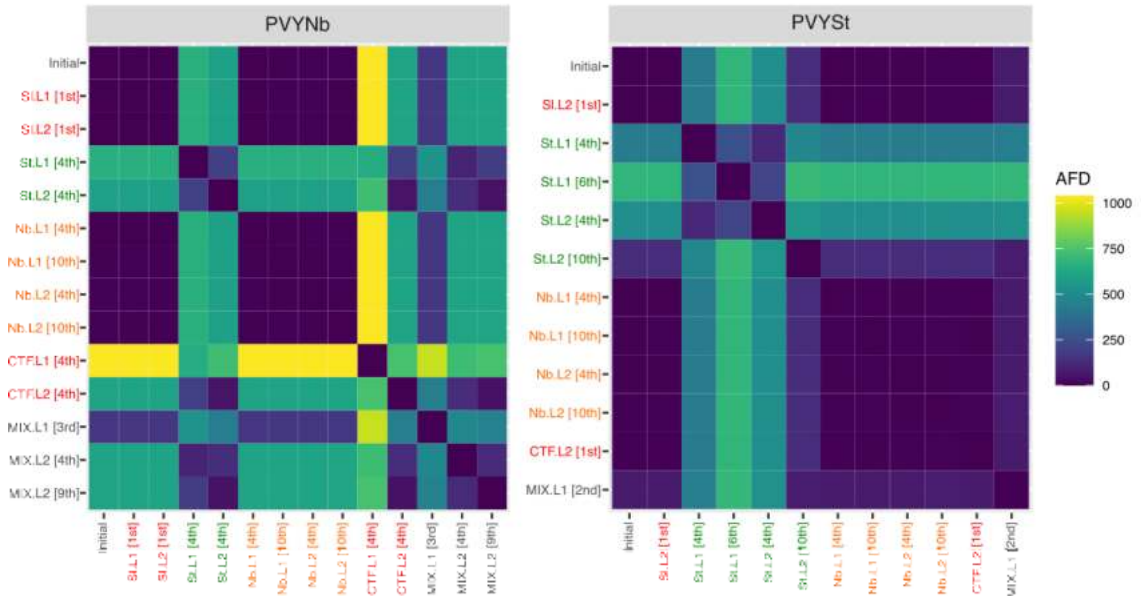


Fig 7. Allele frequency difference (AFD) analysis for genetic differentiation between PVY populations (treatments and lineage of the selected passage) based on sequences generated by Illumina sequencing of total RNA. The colors applied on sample column (y-axis) indicate the plant species: orange for benthamiana, green for potato and red for tomato.

Although PVYNb can infect a broad range of hosts, its population variation was higher (AFD = 1040) compared to PVYSt (AFD = 701), with PVYNb also exhibiting greater variation between samples (mean AFD ±1 SD: 388 ±335) than PVYSt (235 ±255).

PVYNb AFD can be divided into three clusters: (i) the first cluster, which includes the initial population, was closely related to tomato populations Sl.L1 and Sl.L2 at the 1st passage, as well as benthamiana populations Nb.L1 and Nb/L2 at the 4th

and 10th passages. For both tomato (Sl) and benthamiana (Nb), L1 and L2 were very closely related, with conserved AFD between passages; (ii) the second cluster comprises potato (St) and mixed lines (MIX), which were closely related and presented fewer differences compared to all samples. An exception was MIX.L1 at the 3rd passage, which was closer to the first cluster; (iii) the third cluster is represented by the CTF alternating line at the 4th passage (tomato), which showed a very distinct pattern, with the highest AFD for L1. Although L2 was similar to St and MIX, L1 presented significant genetic differences, suggesting limited gene flow, potential barriers to gene exchange, or a distinct evolutionary history.

In contrast, the PVYSt populations are more closely related but they could be divided into two clusters: (i) the first cluster includes the initial PVYSt, tomato Sl/L2, benthamiana populations Nb.L1 and Nb.L2 (4th to 10th passages), CTF.L2 (1st passage), and mixed population MIX.L1 (2nd passage). Additionally, the potato St.L2 at the 10th passage became much closer to the initial population than at the 4th passage; (ii) the second cluster includes the potato St.L1 at the 4th and 6th passages and St.L2 at the 4th passage. This suggest that the passage through potato plants filter some haplotypes, and that these plants may exert a selective pressure that reduces genetic diversity by filtering some haplotypes. In contrast, this filtering effect was not observed after passage through benthamiana and tomato plants, for which the virus populations tended to maintain greater genetic diversity and remained more closely related to the initial population.

The number of SNPs differed between populations, impacting the number of fixed synonymous and non-synonymous SNPs (Table 1). Comparing to the initial population (PVYNb time 0), PVYNb/Sl.L1 and Sl.L2 retained 18% of the SNPs in the 1st passage, while St and MIX lines presented none. SNP percentages in Nb.L1 decreased from 36% at the 4th passage to 9% at the 10th, while Nb.L2 remained stable at 27%. CTF.L1 preserved 9% of SNPs. For PVYSt, 5% of SNPs were preserved in Sl.L2 at the 1st passage, while St.L1 retained 22% from the 4th to 6th passage, and St.L2 decreased from 22% to 11%. SNP percentages in Nb.L1 decreased from 11% to 5%, and in Nb.L2 from 16% to 11% from the 4th to 10th passage. MIX.L1 had 11% at the 2nd passage, and CTF.L2 had zero at the 1st passage. This variability in SNP underscored the dynamic nature of viral evolution, reflecting how specific host environments influenced the genetic stability and adaptation of the virus.

The observed differences in the number of fixed SNPs between PVYNb and PVYSt highlighted significant distinctions in the evolutionary dynamics of both PVY isolates (Table 1). PVYNb populations exhibited a generally higher number of fixed SNPs compared to PVYSt, reflecting greater genetic variability and potentially more extensive adaptation within these populations. For example, St.L2 and CTF.L1 and CTF.L2 accumulated substantial numbers of fixed SNPs, with St.L2 reaching 414 fixed SNPs, including a significant proportion of non-synonymous changes. This suggests that PVYNb underwent considerable selective pressures or mutational events, particularly during passages through potato plants.

In contrast, PVYSt lines showed lower numbers of fixed SNPs, with several lines, including Sl, Nb and CTF, retaining no fixed SNPs throughout their passages. The minimal accumulation of fixed SNPs in PVYSt might indicate a more stable or conserved genetic profile, potentially due to less intense selective pressures or a more uniform host environment. However, St.L1 displayed a marked increase in fixed SNPs from the 4th to 6th passage, accumulating 192 fixed SNPs, which included a substantial number of synonymous changes. This suggests that while PVYSt generally exhibited less genetic divergence, specific conditions or passages could still drive significant genetic variation.

Overall, the greater variability in PVYNb suggests it may be more prone to genetic changes under different selective pressures, potentially allowing it to adapt more rapidly to new hosts or environmental conditions. In contrast, PVYSt relatively stable SNP profile could indicate greater selection pressure in its host environment, leading to the elimination of less adapted variants and consequently a lower genotypic diversity. This suggests that while PVYNb might be more flexible in its evolution, PVYSt evolutionary strategy might be more focused on optimizing fitness within a specific host, leading to a more conserved genetic makeup.

Genetic diversity, measured by Shannon Entropy (SH) (Table 1), fluctuated depending on the host and passage. Both initial populations had low SH, which can indicate genetic stability at the initial inoculum. For PVYNb, SH decreased in tomato Sl.L1 and Sl.L2 and benthamiana Nb.L1 and Nb.L2 at the 4th passage but slightly increased at the 10th passage. Other populations showed higher SH compared to the

initial population, with St.L1 (4th passage) and MIX.L1 (3rd passage) having the highest SH.

Table 1. Population genetic analyses of all last positive samples divided by treatment (T) and lineage (L1 or L2) and passage number.

Sample	Passage	% SNPs*	Fixed SNPs	Synonymous fixed SNPs	Nonosynonymous fixed SNPs	SH (×1000)
PVYNb	0	-	-	-	-	0.363
PVYNb/Sl.L1	1	0.1818	0	0	0	0.116
PVYNb/Sl.L2	1	0.1818	0	0	0	0.116
PVYNb/St.L1	4	0.0000	23	14	9	37.602
PVYNb/St.L2	4	0.0000	414	300	102	1.750
PVYNb/Nb.L1	4	0.3636	0	0	0	0.271
PVYNb/Nb.L1	10	0.0909	0	0	0	0.480
PVYNb/Nb.L2	4	0.2727	0	0	0	0.620
PVYNb/Nb.L2	10	0.2727	0	0	0	1.026
PVYNb/CTF.L1	4	0.0909	382	282	98	5.405
PVYNb/CTF.L2	4	0.0000	233	158	67	2.833
PVYNb/MIX.L1	3	0.0000	0	0	0	47.110
PVYNb/MIX.L2	4	0.0000	96	61	35	2.093
PVYNb/MIX.L2	9	0.0000	24	11	13	5.796

PVYSt	0	-	-	-	-	0.361
PVYSt/T1.L2	1	0.0556	0	0	0	0.063
PVYSt/St.L1	4	0.2222	0	0	0	62.550
PVYSt/St.L1	6	0.2222	192	147	32	3.677
PVYSt/St.L2	4	0.2222	0	0	0	56.470
PVYSt/St.L2	10	0.1111	1	0	0	1.222
PVYSt/Nb.L1	4	0.1111	0	0	0	0.201
PVYSt/Nb.L1	10	0.0556	0	0	0	0.750
PVYSt/Nb.L2	4	0.1667	0	0	0	0.404
PVYSt/Nb.L2	10	0.1111	0	0	0	1.240
PVYSt/CTF.L2	1	0.0000	0	0	0	0.112
PVYSt/MIX.L1	2	0.1111	2	0	0	12.812

*Number of SNPs from the initial PVY that are present in the population

SH = Shannon Entropy

The colors applied on sample column indicate the plant species: orange for benthamiana, green for potato and red for tomato.

Notable increases in SH observed in PVYNb/St.L1 (0.037) and PVYSt St.L1/L2 at the 4th passage might reflect the isolate adaptative response to potato plants, driving the generation of new variants that can better exploit the host environment. Curiously, the SH decreased in later passages in potato plants. The PVYSt/St.L1 at the 6th and St.L2 at the 10th passage showed a reduction in the genetic diversity, potentially indicating fixation of specific mutations, population homogenization or stabilization of the virus within the host environment. In contrast, the SH slightly increased in benthamiana populations from 4th to 10th passages to both isolates. This suggest a slight increase of genetic variability, indicating that over the time, viral populations might be

experiencing less bottleneck effects or more balanced selective pressures, allowing a broader range of genetic variants to coexist.

Although the number of SNPs in PVYNb CTF.L1/L2 was high, the SH did not necessarily increase proportionally, indicating that these SNPs may be clustered in specific regions of the genome rather than being spread evenly. This suggests that the observed genetic changes might be concentrated in particular genomic areas, which can influence measures of genetic diversity differently than simply counting SNPs.

In summary, comparing PVYNb with PVYSt, based on the AFD analysis, the results suggest that the PVYNb population may be more diverse and more specialized to certain hosts, whereas PVYSt appears more versatile and generalist. Host species seems to influence the evolutionary process, with a lower need to fix SNPs when infecting benthamiana, a permissive host. However, adaptation is required when transitioning to a different host environment. When infecting the same host, the virus tends to reduce the number of fixed SNPs. The exception was PVYSt/St.L1 from the 4th to 6th passage, from which the number of SNPs increased. Both synonymous and non-synonymous SNPs were present, with synonymous SNPs being more frequent. A host species change appeared to create a bottleneck effect, which is dependent on the PVY isolate. PVYSt, being more versatile, experiences high gene flow and low selective pressure, maintaining higher genetic diversity with fewer fixed SNPs. Conversely, populations struggling to adapt to new host environments tend to have a higher number of SNPs.

As a final step, a ML-tree was constructed using all consensus genomes, as depicted in Sup. Fig 4. Notably, sequences that were previously excluded due to not meeting the 50× average coverage threshold were included in this analysis. Despite the lower coverage, the consensus sequences still reflect the dominant viral population, allowing us to construct a phylogeny that may help understanding the relationships among the PVY populations. The tree shows well-structured group formations with highly intriguing clades. PVYNb and PVYSt isolates tended to cluster with isolates derived from the same initial virus. Additionally, the host played a significant role in shaping the phylogenetic structure, as seen in benthamiana isolates from both viruses, which showed to be closely related to the initial population and to each other. Indeed, the initial PVYNb clustered with all the benthamiana treatments (Nb.L1/L2 4th and 10th passage) and tomato isolates in the first generation, a pattern also observed for PVYSt.

This demonstrated multiple passages in a permissive host like benthamiana, and also in tomato, the most non-permissive host, produced similar effects. It may explain why tomato plants in general behaved as the dead-end host of PVYNb and PVYSt, since these isolates were not adapted to tomato plants, and new variants were not easily produced.

Another notable aspect is the tendency of potato isolates to remain grouped together, often forming a distinct clade separate from other isolates. This observation highlights the potato as a host that can rapidly induce and fix genome alterations within a shorter time frame. These alterations result in isolates that cluster differently from the initial population. For example, while PVYSt St.L2 in the 4th passage appears significantly distinct from the original population, by the 10th generation, this difference diminishes. This shift underscores the strong bottleneck effects during passage in the potato host, where isolates with new genomic characteristics can emerge. This clustering pattern reflects the influence of the host and passage number on the phylogenetic relationships among virus populations, particularly given the high number of SNPs observed, many of which were fixed during the passage experiment.

Discussion

Understanding the ecology, evolution, and population biology of viruses, their hosts and vectors, and the communities they inhabit is fundamental for thoroughly analyzing the conditions for disease emergence in new plant hosts (Morse 1995; Jones 2009; Lefeuvre et al. 2019). Emerging plant viruses, such as tomato torrado virus, tomato brown rugose fruit virus, cassava brown streak virus, rice stripe virus or cucumber green mottle mosaic virus, have caused significant damage over the past decades. The growing impact of these viruses on agricultural crops and wild plant populations underscores the urgent need to comprehend the ecological dynamics behind plant viral emergence (Jones 2009; Lefeuvre et al. 2019).

Viral emergence often occurs in multiple stages. Initially, virus fitness is typically lower on a new host compared to an original host due to the lack of adaptation. The virus must establish and maintain a population in the new host and then further dissemination (Morse 1995). To study the early phase of a plant viral emergence, we

have used PVY as a model, in its native host, potato, and included two new hosts, tomato and benthamiana. Our research focused on how PVY establishes persistent populations on these hosts.

The implications of source-sink dynamics for viral emergence depend on a pathogen's pattern of exposure to a given host. Broadly, a pathogen may encounter a novel host in two ways: homogeneously or heterogeneously. In the homogeneous case, a pathogen population initially contacts the novel host and is then isolated from the original host. If the novel host is of low quality, leading to the pathogen's absolute fitness being well below one, the pathogen is likely to go extinct before it can adapt (Holt and Gaines 1992; Gomulkiewicz and Holt 1995; Morse 1995; Antia et al. 2003). In the heterogeneous case, the pathogen intermittently encounters both hosts across different times or locations. This heterogeneous exposure can be coarse-grained or fine-grained in space, time, or both, and may facilitate initial persistence on the novel host. Our study examined various situations, including an extreme case of coarse-grained temporal heterogeneity, where the virus alternated between host environments with each passage, mimicking 100% dispersal between source and sink. This scenario is akin to a pathogen facing seasonal or annual host availability changes, like an agricultural pest infecting rotated crops. We hypothesized that intermittent exposure to the original host could act as a "rescue" mechanism for emerging PVY populations on novel hosts, potentially enhancing their chances of adaptation and persistence. This could provide insights into how viruses might exploit familiar host environments to overcome challenges in new hosts, offering information for predicting and managing viral outbreaks in agricultural settings. Understanding these dynamics is crucial for developing strategies to mitigate the impact of emerging plant viruses. Future research will explore other heterogeneous host use patterns, such as spatiotemporal variations in a metacommunity or fine-grained variation in well-mixed host communities.

The persistence of the virus with alternating host exposure suggests that a temporally heterogeneous host regime might provide emerging viruses time to adapt to novel hosts, facilitating this shift by affecting viral population size. However, whether this leads to host expansion or a host shift remains to be seen. The extensive literature on evolution in heterogeneous environments will not be covered in depth here, as our

focus is on the initial ecological challenges before adaptation. It is noteworthy that pathogens may persist on novel hosts despite low fitness, allowing new genes and gene combinations to arise through mutation and recombination. The success of new genes or combinations depends on their effects (Kawecki 2000). Holt and Gomulkiewicz (1997) suggested adaptation in a sink require a mutant capable of persisting on the novel host without immigration from the original host. This stringent "absolute fitness criterion" is rarely met when transmission is low and the novel host is a strong sink (*i.e.*, PVYSt/CTF.L1 and L2 from the 4th to 5th passage). However, this assumes a negative correlation of mutant fitness on the two hosts and unidirectional immigrant flow. Our research shows that alternating host exposure can select for mutations enhancing PVY growth on both hosts, as evidenced in PVYNb/CTF Lines. Positive correlations in selection responses increase emergence likelihood (Gandon 2004), and bidirectional transmission may enhance adaptation probability to the novel host (Kawecki and Holt 2002).

The primary aim of emerging disease research is to pinpoint the crucial ecological factors that drive the emergence of new plant viral diseases. By identifying these factors, we can improve our ability to predict which host populations are at greatest risk of future infections by emerging plant viruses (Lefeuvre et al. 2019). Our research has demonstrated that the persistence of a virus in a homogeneous environment can be predicted based on its growth and transmission rates. However, these metrics are less effective for predicting outcomes in more complex host exposure scenarios, which are more representative of real-world emergence events.

Plant virus populations are genetically heterogeneous, meaning each combination of host and virus is unique. The genetic diversity in RNA virus populations is governed by interactions between host and viral factors (Schneider and Roossinck 2001), and the population structure of plant viruses varies across different hosts (Huang et al. 2015). In our study, we found that benthamiana is a permissive host for PVY, while tomato acts as a sink for the PVY population. But it is important to highlight the source of the virus, while PVYSt was collected in potato field, PVYNb was maintained in laboratory in benthamiana.

Our results suggest that PVYNb is a more dynamic virus in terms of its ability to adapt to different host environments, likely due to its higher genetic variability. This

could make PVYNb more versatile but also potentially more prone to developing virulence or resistance to host defenses. In contrast, PVYSt, while generally more stable, can still exhibit significant diversity under certain conditions. This stability might make it less likely to evolve quickly, which could be advantageous in a consistent environment but might limit its adaptability to new or changing conditions. These results may indicate that in more restrictive hosts, such as potato, viral populations evolve more slowly, in contrast to that has observed for strain YC5 of *Potyvirus rapae* in *Arabidopsis thaliana* (Navarro et al. 2022). The bottleneck effect during host change was more pronounced in the PVYNb population, indicating that PVYNb has adapted more specifically to hosts, such as benthamiana, potentially limiting its adaptability to other host environments. When transmitted to different plant species, the ratio of synonymous to non-synonymous substitutions tends to increase, indicating that both types of substitutions are crucial for virus-host interactions (Huang et al. 2015). This increase in the ratio of synonymous to non-synonymous substitutions was more pronounced in the PVYNb population, highlighting its greater specialization to specific hosts. This suggests that PVYNb evolution involves more precise adaptation to host interactions. In response to virus replication and movement, the host can activate various defense mechanisms, such as innate immunity, autophagy, and gene silencing. Permissive hosts provide essential components needed for the virus to replicate within the cell (Kushner et al. 2003; Panavas et al. 2005) and in the absence of these factors, virus accumulation can reduce due the deficiency in replication or movement. The SNPs decrease in later lines, i.e. PVYNb/Nb.L1 and L2, can represent an adaptation of the population to the host environment. On the other hand, populations with high genetic diversity can activate plants defense factors that can repress the virus, *i.e.*, PVYNb/CTF.L1that was unavailable to infect the next generation.

As mentioned, tomato acted as a sink crop in our experiment, but this does not imply that tomato cannot be infected by PVY. Recent outbreaks of PVY in tomato crops in Brazil demonstrate that this virus remains present and capable of causing agronomic damage (Lucena et al. 2024). It is important to acknowledge that our study did not include PVY isolates obtained from tomato crops, and different isolates might exhibit varying behavior when infecting tomato plants. Specifically, isolates adapted to tomatoes might show different infectivity patterns compared to those from other

sources. Additionally, the host from which the virus was originally isolated plays a crucial role in determining the virus's behavior and adaptability. Future research using a broader range of isolates, including those adapted to different hosts, will be essential to fully understand the variability in virus-host interactions. Our findings highlight that PVYNb and PVYSt exhibit distinct transmission dynamics across different hosts, and the emergence of different viral populations, as observed through allele frequency differences in potato treatments St.L1 and L2, further supports this variability. This is consistent with previous studies showing that different PVY isolates can lead to diverse responses in various hosts, including the ability to complete the replicative cycle in resistant cultivars (Morais et al. 2024).

In conclusion, our study underscores the complexity and variability of plant virus emergence and adaptation across different hosts. By examining the behavior of PVY in different host species, we revealed critical insights into the dynamics of viral persistence and genetic differentiation. The observed specialization of PVYNb in specific host environments and the relatively stable genetic diversity of PVYSt across different hosts underscore how host conditions can significantly influence viral evolution. These findings illustrate that the host environment plays a crucial role in shaping the genetic and adaptive responses of plant viruses. The implications of source-sink dynamics, particularly in heterogeneous host exposure scenarios, suggest that viruses can persist and adapt to novel hosts through intermittent exposure to native hosts, potentially facilitating host expansion or shifts. These findings contribute to a broader understanding of the ecological and evolutionary factors driving the emergence of new plant viral diseases. Future research should focus on diverse host exposure patterns and their effects on viral adaptation, ultimately aiming to predict and mitigate the risks of emerging plant viruses in agricultural and wild populations. Understanding these dynamics is crucial for developing effective strategies to manage and prevent viral epidemics, ensuring the stability and productivity of global agriculture.

It is important to note that the experimental approach was focused on the detection of the early stages of interaction of virus and the host. A prolonged incubation time would produce additional factors leading to complex interpretations. On the other hand, it could also enable a buildup of fitter virus populations. This issue must be addressed in future studies.

One of the questions we wanted to answer was whether PVY evolved in a specific host could easily adapt to a closely related but distinct host, such as potato and tomato. Our findings indicate that PVY exhibits distinct evolutionary patterns depending on the host, with PVYNb adapting more rapidly and showing higher genetic diversity, while PVYSt demonstrated greater stability, underscoring the complex interplay between host environments and viral evolution.

References

Agudelo-Romero P, de la Iglesia F, Elena SF (2008) The pleiotropic cost of host-specialization in Tobacco etch potyvirus. Infection, Genetics and Evolution 8:806–814. https://doi.org/10.1016/j.meegid.2008.07.010

Antia R, Regoes RR, Koella JC, Bergstrom CT (2003) The role of evolution in the emergence of infectious diseases. Nature 426:658–661. https://doi.org/10.1038/nature02104

Bedhomme S, Lafforgue G, Elena SF (2012) Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. Molecular Biology and Evolution 29:1481–1492. https://doi.org/10.1093/molbev/msr314

Belshaw R, Pybus OG, Rambaut A (2007) The evolution of genome compression and genomic novelty in RNA viruses. Genome Research 17:1496–1504. https://doi.org/10.1101/gr.6305707

Berner D (2019) Allele Frequency Difference AFD: an intuitive alternative to FST for quantifying genetic population differentiation. Genes (Basel) 10:308. https://doi.org/10.3390/genes10040308

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12:59–60. https://doi.org/10.1038/nmeth.3176

de la Iglesia F, Martínez F, Hillung J, et al (2012) Luria-delbrück estimation of turnip mosaic virus mutation rate *in vivo*. Journal of Virology 86:3386–3388. https://doi.org/10.1128/JVI.06909-11

Dennehy JJ, Friedenberg NA, Holt RD, Turner PE (2006) Viral ecology and the maintenance of novel host use. The American Naturalist 167:429–439. https://doi.org/10.1086/499381

Dennehy JJ, Friedenberg NA, Yang YW, Turner PE (2007) Virus population extinction via ecological traps. Ecology Letters 10:230–240. https://doi.org/10.1111/j.1461-0248.2006.01013.x

Domingo E, Sheldon J, Perales C (2012) Viral quasispecies evolution. Microbiology and Molecular Biology Reviews 76:159–216. https://doi.org/10.1128/MMBR.05023-11

Edwardson JR, Christie RG (1997) Viruses infecting peppers and other solanaceous crops. Agricultural Experiment Station, University of Florida, Gainesville.

Elena SF (2017) Local adaptation of plant viruses: lessons from experimental evolution. Molecular Ecology 26:1711–1719. https://doi.org/10.1111/mec.13836

Elena SF, Agudelo-Romero P, Lalic J (2009) The evolution of viruses in multi-host fitness landscapes. The Open Virol Journal 3:1–6. https://doi.org/10.2174/1874357900903010001

Flatschacher D, Speckbacher V, Zeilinger S (2022) qRAT: an R-based stand-alone application for relative expression analysis of RT-qPCR data. BMC Bioinformatics 23:286. https://doi.org/10.1186/s12859-022-04823-7

Gandon S (2004) Evolution of multihost parasites. Evolution (NY) 58:455–469.

Gandon S, Hochberg ME, Holt RD, Day T (2013) What limits the evolutionary emergence of pathogens? Philosophical Transactions of the Royal Society B: Biological Sciences 368:20120086. https://doi.org/10.1098/rstb.2012.0086

Gomulkiewicz R, Holt RD (1995) When does Evolution by Natural Selection Prevent Extinction? Evolution (NY) 49:201. https://doi.org/10.2307/2410305

Hillung J, Cuevas JM, Valverde S, Elena SF (2014) experimental evolution of an emerging plant virus in host genotypes that differ in their susceptibility to infection. Evolution 68:2467–2480. https://doi.org/10.1111/evo.12458

Holt RD, Gaines MS (1992) Analysis of adaptation in heterogeneous landscapes: implications for the evolution of fundamental niches. Evolutionary Ecology 6:433–447. https://doi.org/10.1007/bf02270702

Holt RD, Gomulkiewicz R (1997) How does immigration influence local adaptation? A Reexamination of a Familiar Paradigm. The American Naturalist 149:563–572.

Huang L, Li Z, Wu J, et al (2015) Analysis of genetic variation and diversity of rice stripe virus populations through high-throughput sequencing. Frontiers in Plant Science 6. https://doi.org/10.3389/fpls.2015.00176

Inoue-Nagata AK, Jordan R, Kreuze J, et al (2022) ICTV Virus Taxonomy Profile: Potyviridae 2022. Journal of General Virology 103:001738. https://doi.org/10.1099/jgv.0.001738

Jeffries CJ (1998) FAO/IPGRI Technical guidelines for the safe movement of potato germplasm. Food and Agriculture Organization of the United Nations, Rome/International Plant Genetic Resources Institute, Rome, Italy 19.

Jones RAC (2009) Plant virus emergence and evolution: Origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. Virus Research 141:113–130. https://doi.org/10.1016/j.virusres.2008.07.028

Kawecki TJ (2000) Adaptation to marginal habitats: contrasting influence of the dispersal rate on the fate of alleles with small and large effects. Proceedings of the Royal Society of London B Biology Science 267:1315–1320. https://doi.org/10.1098/rspb.2000.1144

Kawecki TJ, Holt RD (2002) Evolutionary Consequences of Asymmetric Dispersal Rates. The American Naturalist 160:333–347. https://doi.org/10.1086/341519

Kerlan C, Moury B, Granoff A, Webster RG (2008) Encyclopedia of virology. Association of Applied Biologists 287–296.

Kushner DB, Lindenbach BD, Grdzelishvili VZ, et al (2003) Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. Proceedings of the National Academy of Sciences 100:15764–15769. https://doi.org/10.1073/pnas.2536857100

Lefeuvre P, Martin DP, Elena SF, et al (2019) Evolution and ecology of plant viruses. Nature Reviews Microbiology 17:632–644. https://doi.org/10.1038/s41579-019-0232-3

Li D, Liu C-M, Luo R, et al (2015) MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Lucena VS, Nakasu EYT, Pereira JL, et al (2024) Emergence of potato virus Y outbreaks in tomatoes in Brazil, the disease and spread. bioRxiv. https://doi.org/10.1101/2024.05.17.594728

MARTIN W, CERFF R (1986) Prokaryotic features of a nucleus-encoded enzyme. cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*). European Journal of Biochemistry 159:323–331. https://doi.org/10.1111/j.1432-1033.1986.tb09871.x

Minh BQ, Schmidt HA, Chernomor O, et al (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Morais IJ, Silva DYM, Camargo BM, et al (2024) Unraveling the dynamics of host specificity and resistance responses to potato virus Y, and implications for crop management. bioRxiv.

Morse SS (1995) Factors in the Emergence of Infectious Diseases. Emerging Infectous Diseases 1:7–15. https://doi.org/10.3201/eid0101.950102

Navarro R, Ambrós S, Butković A, et al (2022) Defects in plant immunity modulate the rates and patterns of RNA virus evolution. Virus Evolution 8. https://doi.org/10.1093/ve/veac059

Panavas T, Serviene E, Brasher J, Nagy PD (2005) Yeast genome-wide screen reveals dissimilar sets of host genes affecting replication of RNA viruses. Proceedings of the National Academy of Sciences 102:7326–7331. https://doi.org/10.1073/pnas.0502604102

Peláez A, McLeish MJ, Paswan RR, et al (2021) Ecological fitting is the forerunner to diversification in a plant virus with broad host range. Journal of Evolutionary Biology 34:1917–1931. https://doi.org/10.1111/jeb.13672

Ruark-Seward CL, Bonville B, Kennedy G, Rasmussen DA (2020) Evolutionary dynamics of Tomato spotted wilt virus within and between alternate plant hosts and thrips. Scientific Reports 10:. https://doi.org/10.1038/s41598-020-72691-3

Sanjuán R, Agudelo-Romero P, Elena SF (2009) Upper-limit mutation rate estimation for a plant RNA virus. Biology Letters 5:394–396. https://doi.org/10.1098/rsbl.2008.0762

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nature Protocols 3:1101–1108. https://doi.org/10.1038/nprot.2008.73

Schneider WL, Roossinck MJ (2001) Genetic diversity in RNA virus quasispecies is controlled by host-virus interactions. Journal of Virology 75:6566–6571. https://doi.org/10.1128/JVI.75.14.6566-6571.2001

Scholthof KBG, Adkins S, Czosnek H, et al (2011) Top 10 plant viruses in molecular plant pathology. Molecular Plant Pathology 12:938–954. https://doi.org/10.1111/j.1364-3703.2011.00752.x

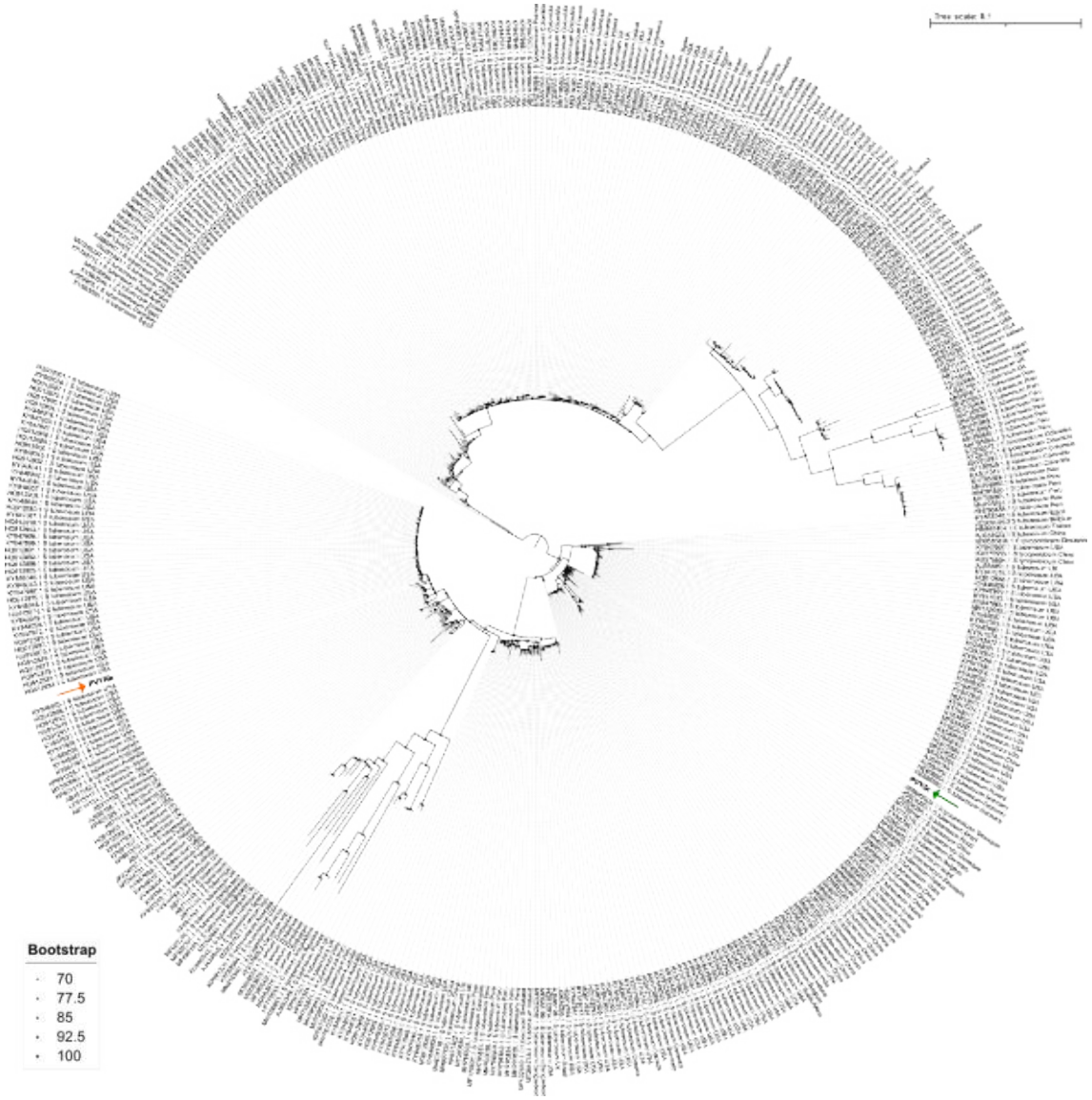
Shukla DD, Ward CW, Brunt AA (1994) The *Potyviridae*. Cambridge: CAB International.

Van der Auwera GA, O'Connor BD (2020) Genomics in the Cloud: using Docker, GATK, and WDL in Terra. O'Reilly Media, Inc.

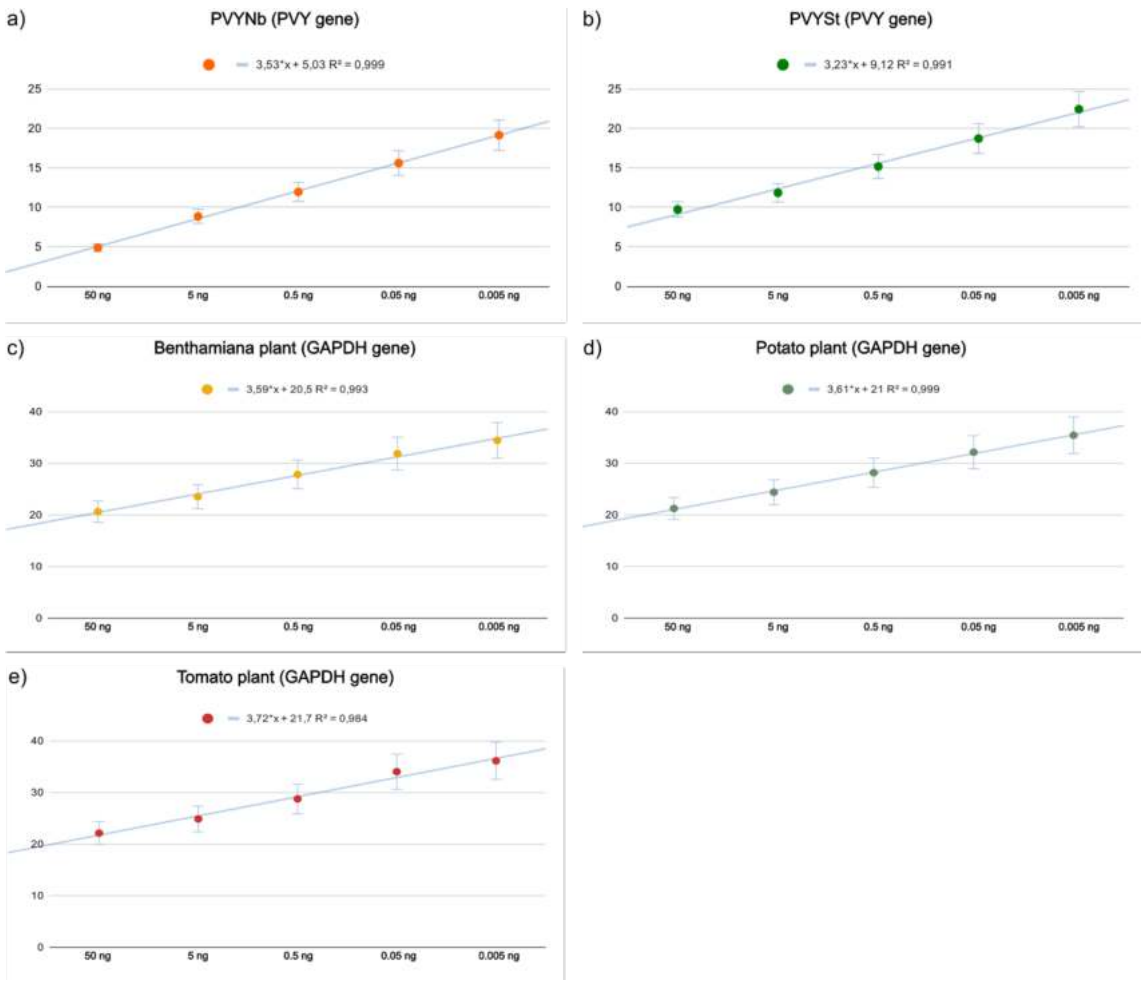
Wallis CM, Stone AL, Sherman DJ, et al (2007) Adaptation of plum pox virus to a herbaceous host (*Pisum sativum*) following serial passages. Journal of General Virology 88:2839–2845. https://doi.org/10.1099/vir.0.82814-0

Whitlock MC (1996b) The Red Queen beats the jack-of-all-trades: the limitations on the evolution of phenotypic plasticity and niche breadth. The American Naturalist 148:S65–S77. https://doi.org/10.1086/285902

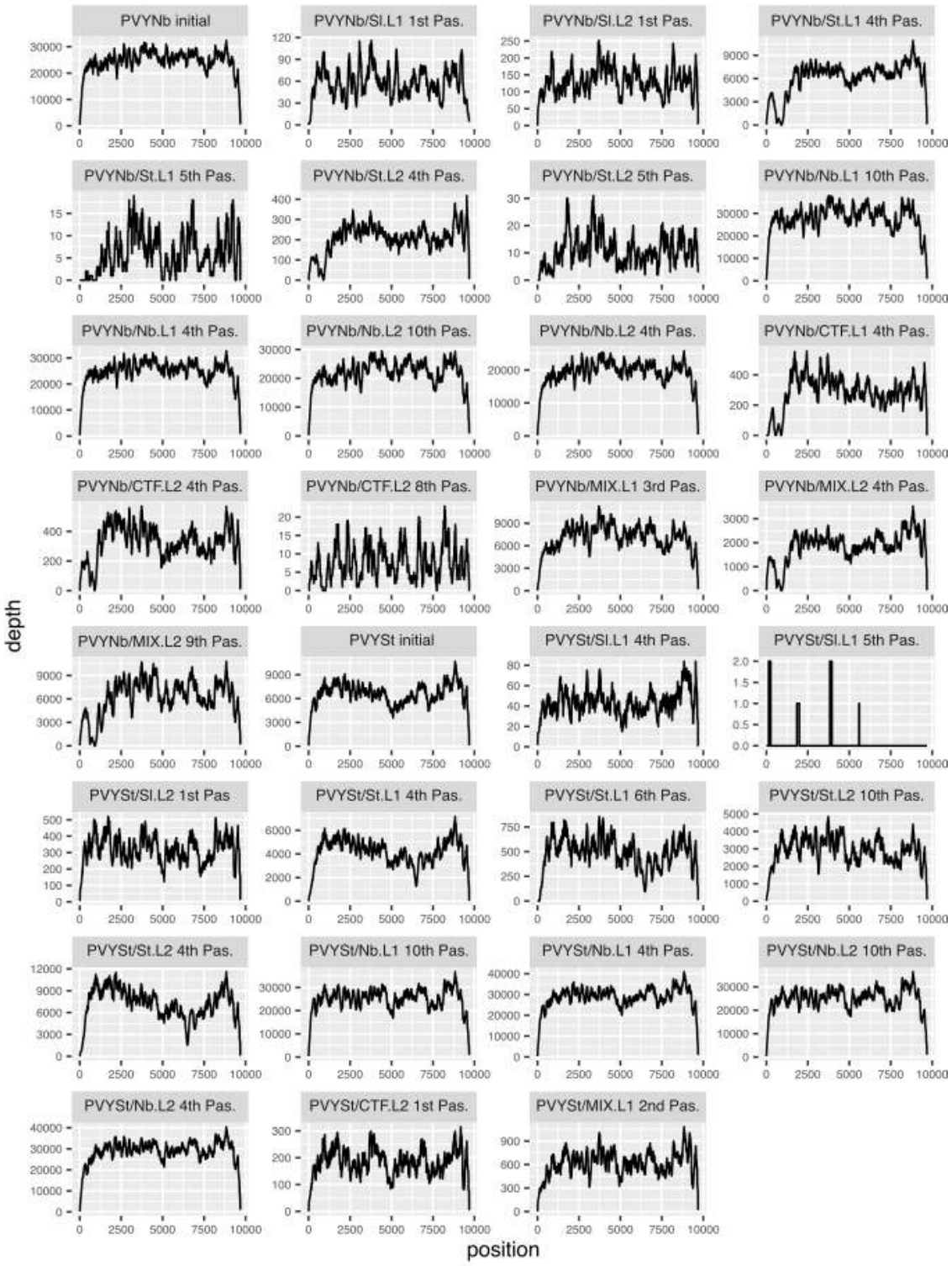
Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. Emerging Infectious Diseases 11:1842–1847. https://doi.org/10.3201/eid1112.050997



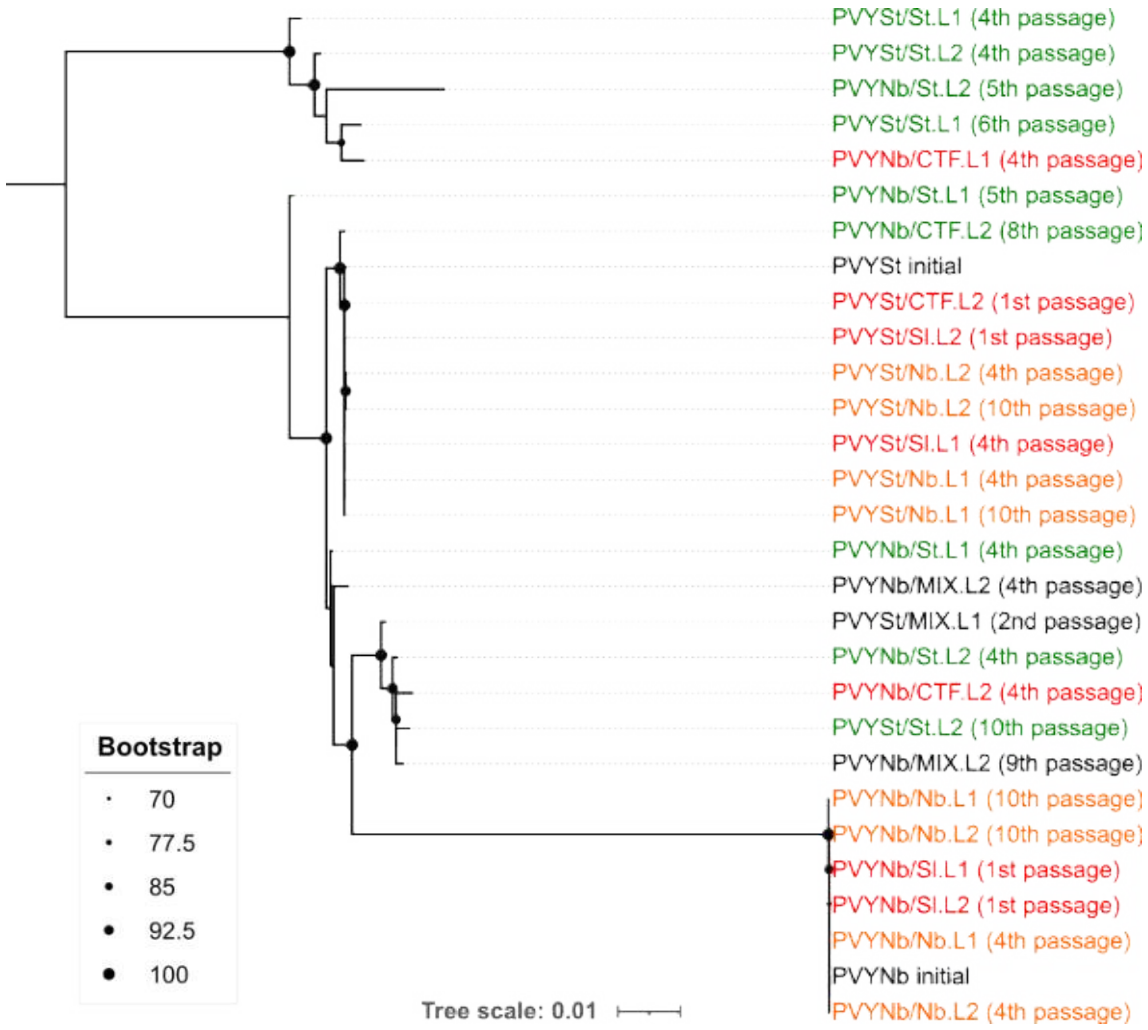
Sup Fig 1. ML-tree constructed by iqtree2 using 447 PVY isolates retrieved from GenBank together with PVYNb and PVYSt consensus sequence with 10,000 bootstrap replications. An arrow indicates the PVYNb (orange) and PVYSt (green) isolates.



Sup Fig 2. Efficiency of qPCR primers targeting PVYNb (a) and PVYSt (b) and GAPDH to benthamiana (c), potato (d) and tomato (e) plants. The regression equation and the R^2 are shown above the curve. The Eff was calculated using five dilutions ranging from 50 to 0.005 ng of initial RNA. The optimal qRT-PCR conditions were previously defined.



Sup Fig 3. Reads coverage along the genome. Sequencing was performed from total RNA in the Illumina platform.



Sup Fig 4. ML-tree constructed by iqtree2 using all Illumina consensus genomes with 10,000 bootstrap replications. The colors applied on sample labels indicate the plant species: orange for benthamiana, green for potato and red for tomato. The initial PVY sources and the mixed MIX treatment are shown in black.

Sup Table 1. Final inoculation results of the passage experiment, in which the last positive sample of each line was used to inoculate ten plants of each species followed by individual PVY detection by RT-PCR.

Virus	Line	Passage	Host	Total # of plants	Infected plants
PVYNb	Sl.L1	1	Benthamiana	10	0
PVYNb	Sl.L1	1	Potato	10	0
PVYNb	Sl.L1	1	Tomato	10	0
PVYNb	Sl.L2	1	Benthamiana	10	1
PVYNb	Sl.L2	1	Potato	10	0
PVYNb	Sl.L2	1	Tomato	10	0
PVYNb	St.L1	5	Benthamiana	10	0
PVYNb	St.L1	5	Potato	10	0
PVYNb	St.L1	5	Tomato	10	0
PVYNb	St.L2	5	Benthamiana	10	8
PVYNb	St.L2	5	Potato	10	2
PVYNb	St.L2	5	Tomato	10	9
PVYNb	Nb.L1	10	Benthamiana	10	8
PVYNb	Nb.L1	10	Potato	10	0
PVYNb	Nb.L1	10	Tomato	10	10
PVYNb	Nb.L2	10	Benthamiana	10	10
PVYNb	Nb.L2	10	Potato	10	1
PVYNb	Nb.L2	10	Tomato	10	10
PVYNb	CTF.L1	4	Benthamiana	10	0
PVYNb	CTF.L1	4	Potato	10	3

PVYNb	CTF.L1	4	Tomato	10	5
PVYNb	CTF.L2	8	Benthamiana	10	0
PVYNb	CTF.L2	8	Potato	10	0
PVYNb	CTF.L2	8	Tomato	10	0
PVYNb	MIX.L1	3	Benthamiana	10	9
PVYNb	MIX.L1	3	Potato	10	0
PVYNb	MIX.L1	3	Tomato	10	8
PVYNb	MIX.L2	9	Benthamiana	10	0
PVYNb	MIX.L2	9	Potato	10	0
PVYNb	MIX.L2	9	Tomato	10	0
PVYSt	Sl.L1	5	Benthamiana	10	0
PVYSt	SL.L1	5	Potato	10	0
PVYSt	Sl.L1	5	Tomato	10	0
PVYNb	Sl.L2	1	Benthamiana	10	2
PVYSt	Sl.L2	1	Potato	10	0
PVYSt	Sl.L2	1	Tomato	10	8
PVYSt	St.L1	6	Benthamiana	10	0
PVYNb	St.L1	6	Potato	10	2
PVYSt	St.L1	6	Tomato	10	4
PVYSt	St.L2	10	Benthamiana	10	0
PVYSt	St.L2	10	Potato	10	1
PVYNb	St.L2	10	Tomato	10	0
PVYSt	Nb.L1	10	Benthamiana	10	9

PVYSt	Nb.L1	10	Potato	10	1
PVYSt	Nb.L1	10	Tomato	10	10
PVYSt	Nb.L2	10	Benthamiana	10	10
PVYSt	Nb.L2	10	Potato	10	2
PVYSt	Nb.L2	10	Tomato	10	10
PVYSt	CTF.L2	1	Benthamiana	10	0
PVYNb	CTF.L2	1	Potato	10	2
PVYSt	CTF.L2	1	Tomato	10	2
PVYSt	MIX.L1	2	Benthamiana	10	1
PVYSt	MIX.L1	2	Potato	10	0
PVYNb	MIX.L1	2	Tomato	10	1

PVYNb treatments are shown in orange and PVYSt in green.

Chapter IV

In silico evidences for the presence of defective viral genomes (DVGs) in potato virus Y-infected plants

Abstract

Defective viral genomes (DVGs) are frequently found in RNA virus populations due to the error-prone nature of the viral replicases, in addition to other factors. DVGs are generated during the replication of the wild-type (WT) viral genome because the replicase detaches from the template it is copying and, in some cases, reattaches to a different region of the genome. DVGs can be classified into six types: 3' copy-back (cb) or snap-back (sb), 5' cb/sb, deletion forward, deletion reverse, insertion forward, and insertion reverse. To date, no DVGs have been described for members of the *Potyviridae* family. This study investigates the diversity and dynamics of DVGs generation in potato virus Y (species *Potyvirus yituberosi*) populations. Two datasets were analyzed: the first involving PVY strains (N, O and N-Wi) in potato plants, passage modes, types of transmission, and plant organ, and the second involving PVYNb (isolated from benthamiana) and PVYSt (isolated from potato) in different host plants. High-throughput sequencing data were analyzed to detect and categorize DVGs, using DVGFinder and custom filtering approaches. Principal Component Analysis (PCA) was employed to investigate clustering patterns of DVG types across samples. Furthermore, we explored the diversity of DVG formation and calculated population diversity using Shannon Entropy (SH). DVGs were consistently detected across all PVY samples, with strain-specific variations. Strain O exhibited the highest number of unique DVGs, while strains N and N-Wi showed lower but notable counts. DVG populations varied significantly by transmission mode and host plant. Indeed, potato tubers harbored more unique DVGs than leaves. The second dataset revealed host-specific DVG profiles, with benthamiana showing high DVG diversity, while tomato and potato plants demonstrated more restrictive environments. PCA highlighted distinct clustering patterns of DVG types, but very consistent with all populations. SH

analysis revealed that forward deletion and insertion DVGs exhibit high conservation across PVY strains, while reverse deletion DVGs show greater variability. Also, deletion-type DVGs are the most diverse when different hosts were examined with benthamiana populations demonstrating the highest overall DVG diversity. These findings underscore the complex interplay between viral strain, transmission mode, passage and host plant in shaping DVG diversity and distribution. The results suggest that host species play a critical role in DVG formation and evolution, with implications for understanding PVY variability and optimizing management strategies.

Key-words: DVG diversity, DVG dynamics, Defective RNAs, *in silico* analysis, *Potyvirus*

Introduction

Viruses are among the smallest replicative forms found in almost all environments. However, even smaller entities, known as sub-viral agents, have been identified. These include satellite viruses (200-1800 nucleotides long), which require a helper virus for replication, viroids (200-400 nucleotides), which do not encode proteins nor need a helper virus, virophages (15-30 kbp), which parasitize giant viruses like mimiviruses (La Scola et al. 2008), which are composed solely of proteins and lack nucleic acids.

The concept of "replicators" was introduced by Richard Dawkins in 1976 in "The Selfish Gene". Replicators are entities that pass on their structure largely intact through generations and can be copied or replicated, propagating their form or information (Dawkins 2016). This concept encompasses a wide range of replicators, including inteins, introns, mini-inteins, plasmids, quasi-replicators, retrotransposons, and transposons (Koonin and Starokadomskyy 2016). From unicellular to multicellular organisms, a complex and interconnected network of replicators highlights the evolutionary dynamics of parasite-host coevolution (Koonin and Starokadomskyy 2016).

In 1947, Preben von Magnus described non-infectious, incomplete forms of the influenza A virus, leading to the discovery of defective RNAs (DRNAs) or defective viral genomes (DVGs) (Gard and von Magnus 1947; von Magnus 1954). Although DVGs are versions of the wild-type (WT) viral genome that cannot replicate autonomously, they can form heterogeneous or homogeneous subpopulations (Budzyńska et al. 2020). DVGs represent a complex and nuanced form of replicator, fitting into the broader conceptual framework that includes genes, memes, and other entities capable of replication and evolution. Despite having lost the ability to replicate autonomously, DVGs can propagate and be subject to evolutionary forces. Although they are not fully autonomous, they can still be replicated and passed on to successive generations of viral particles during co-infections with the WT virus.

Certain DVGs can interfere with the production of the WT virus, critically influencing infection outcomes, and are referred to as defective interfering particles (DIPs) (von Magnus 1954). Only DVGs that interfere with WT virus accumulation are

termed DIPs or defective interfering RNAs (DI-RNAs), a subclass of D-RNAs (Huang and Baltimore 1970).

The formation of DVGs is often attributed to the template switching of viral RNA-dependent RNA polymerase (RdRp), known as the copy-choice mechanism (Lazzarini et al. 1981). This process involves the premature dissociation of the viral RNA polymerase and the nascent strand from the RNA template, followed by the reinitiation of replication at a different site, resulting in incomplete RNA strands (Lazzarini et al. 1981). Thus, DVGs originate from the WT genome and often require co-infection with the WT virus to express all necessary viral proteins and package the DVG progeny (Lazzarini et al. 1981).

DVGs can be categorized into three types: (i) those with single or multiple internal deletions, (ii) those with mosaic genomes, which include insertions and deletions, and (iii) copy-back (or snap-back) genomes (Beauclair et al. 2018). Although initially detected in animal viruses, DVGs are also present in plant viruses. For instance, DVGs have been identified in *Bromovirus* (Damayanti et al. 1999; Llamas et al. 2004), *Closterovirus* (Che et al. 2002), *Crinivirus* (Rubio et al. 2000, 2002), *Cucumovirus* (Graves and Roossinck 1995), *Nepovirus* (Hasiów-Jaroszewska et al. 2012), *Orthotospovirus* (de Oliveira Resende et al. 1991, 1992), *Pomovirus* (Torrance et al. 1999), *Potexvirus* (White et al. 1992; Calvert et al. 1996), *Tobravirus* (Visser et al. 1999), and *Tombusvirus* (Burgyan et al. 1989; Knorr et al. 1991; Chang et al. 1995), mostly through serial passage experiments. However, their detection has been elusive for members of the *Potyviridae* family.

The mechanisms shaping the DVG population are host-specific, as evidenced by the formation, maintenance, and accumulation of DVGs in tomato bushy stunt virus (TBSV) populations in *Nicotiana benthamiana* but not in pepper plants (Omarov et al. 2004). These mechanisms can be expanded to other viral families. Although some plant viruses exhibit DVG formation, this field remains underexplored with significant potential for discovery. Given that DVGs have not been found in the *Potyviridae*, we investigated an agriculturally important virus, potato virus Y (PVY), the type member of the *Potyvirus* genus, responsible for substantial crop production losses. PVY has a single-strand positive sense RNA of approximately 9-10kb in length that encode a polyprotein that suffer autoproteolysis (Inoue-Nagata et al. 2022).

PVY is an ideal subject for studying DVGs due to its high mutation and recombination rates and fast replication cycle (Tromas et al 2014; Sanjuán et al. 2009). Furthermore, PVY can be classified in strains (Singh et al. 2008), based on symptoms expression in specific hosts and phylogeny.

Though often considered insignificant non-infectious byproducts in typical or infectious virus cultures, virus-like particles can engage in various biological processes. These processes include disrupting standard infections, initiating apoptosis or destroying host cells, and activating innate immune responses (Vignuzzi and López 2019). Here, we use an *in silico* approach to explore the formation of DVGs through two passage experiments of PVY: one using potato plants with different strains and transmission modes and another using mechanical inoculation with different host combinations.

Materials and methods

Datasets

In our study, we utilized two datasets to investigate the presence of DVGs in PVY populations. The first one, obtained from an online source (da Silva et al. 2020), was divided into subsets according to different criteria outlined in the original paper. These subsets included three PVY strains (N, N-Wi and O), inoculation methods (mechanical (MI), aphid-mediated (AT), or infected tuber (IT) transmission), plant organ (leaf or tuber), and passage number (1 to 5). It is important to note that passage 4 was not available to download and was excluded from the analysis. The raw sequencing reads were downloaded from the NCBI BioProject database (PRJNA601749). To generate the dataset, the potato plants were cultivated in a greenhouse and harvested 14 weeks after planting. The three strains were collected from Wisconsin (PVY^{O}), Minnesota ($PVY^{N\text{-}Wi}$), and Montana (PVY^{N}), all in the USA. The isolates were maintained in lyophilized tobacco tissue and used to inoculate a single founding plant. Then, three source plants were mechanically inoculated using the founding plant, and each source plant was used for each transmission mode (MI, AT, or IT).

The second dataset (author's dataset) was generated from a passage experiment involving two PVY isolates collected from different hosts: *Solanum tuberosum* (potato) (PVYSt) and *Nicotiana benthamiana* (PVYNb). These isolates were passed through three different hosts (benthamiana, potato and tomato) using various combinations over ten passages. The experiment consisted of five treatments: (i) Sl: viruses passed through tomato plants exclusively; (ii) St: viruses passed exclusively through potato plants; (iii) Nb: viruses passed exclusively through benthamiana plants; (iv) CTF (correlated temporal fluctuations): hosts alternated starting with tomato, followed by potato, benthamiana, and back to tomato; (v) MIX: a mix of all three hosts was used during inoculation and collection. Each treatment (T) had two independent lineages (L1 and L2). Each passage lasted ten days. Initial viruses (PVYNb and PVYSt), and different points of passages (ranging from 1st to 10th) to each treatment and respectively linages were also sequenced using high-throughput sequencing (HTS) Illumina, as for the first dataset (a list of all samples used can be found at Sup. Table 1). Unlike the first dataset, the second dataset involved collecting a pool of 16 mechanically inoculated plants per lineage rather than individual plant. The plants were cultivated in growing chambers with a controlled environment. PVYNb was maintained repeatedly in benthamiana plants, and PVYSt was collected from a potato production field.

DVGs identification

To detect the presence of DVGs, we employed DVGFinder (Olmo-Uceda et al. 2022), a metasearch tool designed for Illumina data. DVGFinder integrates ViReMa-a (Routh and Johnson 2014) and DI-tector (Beauclair et al. 2018), two algorithms specifically developed for DVG detection. We applied DVGFinder to the entire first dataset and each subset individually, including all three strains data and its subdivisions by transmission type, plant organ, and number of passages. The same conditions were applied to the second dataset, using the original dataset divisions by treatment, line, and passage. The standard command line of DVGFinder was used. The DVGFinder algorithm requires a reference genome to map the reads. The consensus genome of each sample was mapped against the NCBI database using BLASTn (Johnson et al. 2008), and the genome with the greatest identity was used as the reference genome for the analysis.

The output table from DVGFinder was further analyzed in R (R Core Team 2022). To refine the results and reduce false positives, we added new columns to the data frame: a 'start' column with the minimum values between the breakpoint (BP) and the reinitiating point (RI), and an 'end' column with the maximum values of BP and RI. The difference between 'end' and 'start' was calculated and stored in a 'delta' column. Additionally, a 'total reads' column was created by summing the read counts from ViReMa-a and DI-tector. Data preprocessing involved filtering out rows that did not meet the criteria of a delta greater than one and ViReMa-a read counts greater than 10, ensuring the remaining data was relevant and meaningful. This filter imposed a strong threshold on both datasets, greatly reducing the number of DVGs found but ensuring reliability.

DVG analysis

The analysis of high-throughput sequencing data was conducted to evaluate DVG diversity and distribution across various PVY populations.

Principal Component Analysis (PCA) was performed to explore patterns in DVG distribution and simplify the data by reducing the number of variables, making it easier to interpret complex relationships. PCA helps to identify clusters of samples or DVGs with similar characteristics, revealing underlying structures in the dataset that might not be immediately apparent. To perform PCA, the R stats package version 4.5.0 was used on original read counts.

Normalization of read counts was essential for accurate comparison across samples. Reads per million (RPM) were calculated for each sample using the dplyr (https://dplyr.tidyverse.org/) and tidyr (https://tidyr.tidyverse.org/) R packages. Normalization corrects for differences in sequencing depth, allowing for a fair comparison of DVG abundance across samples. To visualize DVG abundance, a heatmap of normalized read counts for the top found DVGs was generated. The top DVGs criterion refers to selecting the most abundant DVGs based on their read counts in the dataset. This approach is used to focus on DVGs that are most prevalent across samples, which are often of greater interest for analysis due to their higher abundance and potential biological significance.

Shannon Diversity Index (SH) was calculated using vegan R package (10.32614/CRAN.package.vegan) to measure the diversity of DVGs within each sample. SH accounts for both the number of DVG types and their relative abundance, providing a comprehensive measure of diversity.

All used packages were employed using RStudio version 2024.4.1.748 (Posit Team 2024), and all plots were generated using ggplot2 (Wickham 2016).

Results

DVGs profile

The analysis of the first dataset considered various populations. Initially, we focused on two main populations: the founding population and the source plants. The founding population comprised lyophilized PVY-infected tobacco tissue mechanically inoculated into potato plants, which were then divided by strain to initiate the experiment. The source plants, derived from these plants, were further categorized based on inoculation via leaf or tuber to begin the passage experiments. Using the source plants, five passages were conducted, with leaves divided by transmission modes: aphid-mediated transmission (AT), mechanical inoculation (MI), and two passages using infected tuber transmission (IT). Additionally, files containing strain-specific information were grouped into populations corresponding to the N, N-Wi and O strains, which were further subdivided by the plant organ collected in each sample: leaves (L) or tuber (T).

Despite the variations in read numbers across populations and subpopulations, a significant finding was the consistent detection of DVGs across all PVY samples, even when stringent threshold filters were applied. The initial analysis revealed a high number of DVGs without filtering, but applying the filters improved the visualization and removed poorly represented DVGs. In general, the number of DVGs varied significantly among the viral strains. Strain O exhibited the highest number of reads mapped to the reference genome (X12456) using ViReMa-a, as well as the largest number of unique DVGs (Fig. 1). In contrast, strain N-Wi had a smaller number of mapped reads but a DVG count close to that of strain O. Strain N, on the other hand, showed both a low number of mapped reads and a low number of DVGs.

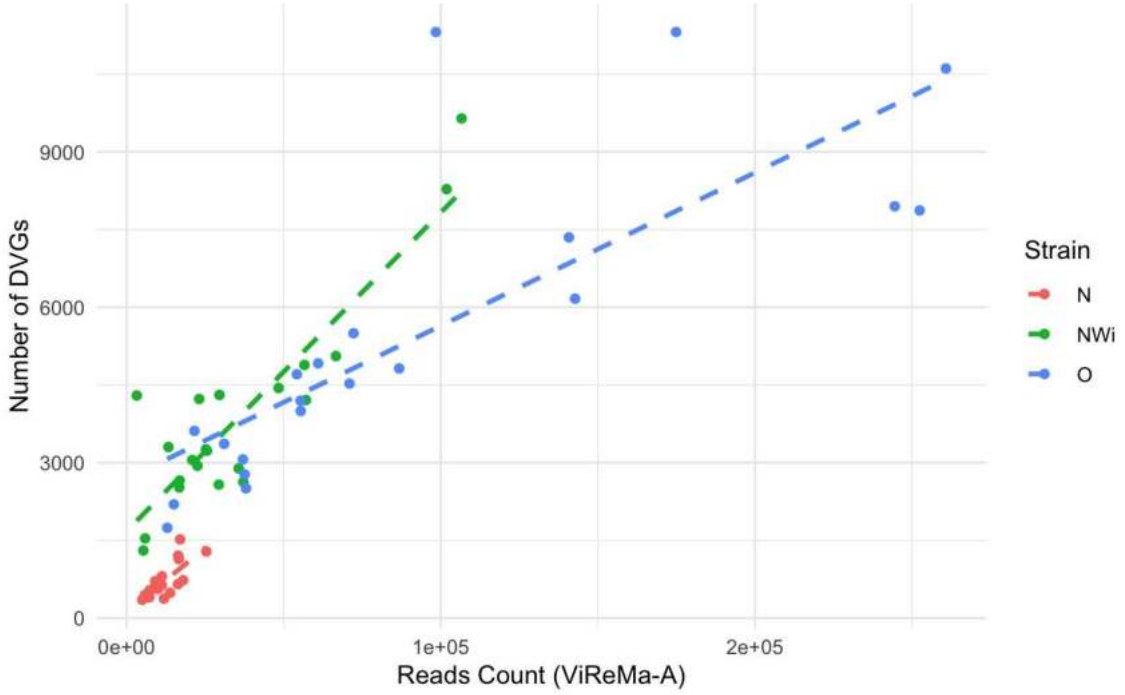


Fig 1. Number of reads mapped against the reference genome using ViReMa-A (*x*-axis) against the number of distinct DVGs detected in each PVY strain to the first dataset. Each point represents a population.

Representative figures for each population are provided in Sup. Fig. 1 and Sup. Table 1. The DVG population varied depending on the dataset, indicating that DVG formation is influenced by multiple factors. When considering all strains together, the foundation populations exhibited fewer DVGs compared to later passages, although some patterns of DVG formation emerged between populations.

The different founding populations of the strains exhibited distinct groupings and formations of DVGs, yet all classes of DVGs were present in these populations. In the first passage to the source plants, a noticeable change in behavior was observed. The number of DVGs in source plants was higher than in the foundation plants for all strains, maintaining diversity. However, when analyzed by plant organ, all strains had more unique DVGs in tubers than in leaves. For example, even though strain O had 4.6 times more mapped reads in leaves than in tubers, the number of unique DVGs was higher in the latter. Given the differences in mapped reads across samples, the number of unique DVGs per sample was considered in the analysis.

During aphid transmission, all strains exhibited a similar pattern, with the number of unique DVGs in leaves decreasing from the first to the fourth passage. In tubers, two patterns emerged: strain N showed a decrease in DVGs, while strains N-Wi and O showed an increase in DVGs with successive viral passages. For tuber transmission, the number of unique DVGs remained relatively constant for strain N, regardless of the organ analyzed. For strain N-Wi, the DVG population in leaves remained stable, while it increased from the first to the second passage in tubers. A similar increase was observed in strain O tubers, though the number of DVGs decreased sharply between the first and second passage when analyzing infected leaves.MI yielded results similar to aphid transmission. For all strains, the number of unique DVGs in leaves tended to decrease over the five viral passages. An exception was noted for strain O, where a significant increase occurred from the first to the second passage, followed by a decrease in subsequent passages. As with AT, the opposite trend was observed in tubers, where unique DVGs tended to increase with viral passages.

The results demonstrate that DVG populations in PVY are highly dependent on the virus strain, transmission mode, and the number of passages. Notably, bottlenecks were observed in some conditions, particularly in leaves during AT and MI, where the number of unique DVGs decreased over time. However, tubers, which exhibited more unique DVGs despite fewer mapped reads, appeared to reduce or mitigate these bottlenecks, suggesting a more permissive environment for viral replication and DVG formation. Strain-specific effects were also evident: strain O, with the highest number of reads and unique DVGs, showed intermittent bottleneck effects, especially in leaves, while tubers maintained or increased DVG diversity. In contrast, strain N exhibited more pronounced bottlenecks in both leaves and tubers, while strain N-Wi showed an intermediate behavior.

The analysis of the second dataset revealed that some populations were devoid of DVGs. Specifically, for PVYNb, DVGs were absent in Sl.L1 and Sl.L2 (both at the first passage), St.L1 and St.L2 (both at the fifth passage), Nb.L1 (at the fourth and tenth passages), CTF.L2 (at the eighth passage), and MIX.L2 (at the ninth passage). For PVYSt, DVGs were absent in Sl.L1 (fifth passage), Sl.L2 (first passage), St.L1 (sixth passage), St.L2 (tenth passage), Nb.L1 (tenth passage), CTF.L2 (first passage), and

MIX.L1 (second passage). Although 97 unique DVGs were detected in PVYNb/CTF.L1, this sample did not pass the threshold filtering and was discarded.

Initially, both PVYNb and PVYSt samples exhibited DVG formation. However, the number of unique DVGs in the PVYNb population (*n* = 22,557) was substantially higher compared to the initial PVYSt (*n* = 4,154) (Fig. 2, Sup. Table 2). Despite some samples having a low number of mapped reads, unique DVGs were still found.

This dataset highlights the critical role of the host in DVG formation (Sup. Table 2). Despite the initial population of unique DVGs of PVYNb in benthamiana, this population remained stable in intermediate and advanced. Similarly, benthamianas infected with PVYSt exhibited a significant increase in the number of unique DVGs, even though the initial PVYSt population was smaller. This indicates the permissiveness of benthamiana in the DVG formation process. Although fewer reads were mapped to St (potatoes), the ratio of reads to unique DVGs was considerably high (1.06 for PVYNb/St.L1, 0.72 for PVYSt/St.L1, and 0.62 for PVYSt/St.L2). Therefore, the number of unique DVGs in potatoes was even greater than in benthamiana for both isolates.

In PVYSt/Sl.L1, only 19 unique DVGs were found among 23 mapped reads, indicating poor sampling but suggesting a highly restrictive environment for DVG generation in tomato plants. For PVYNb, treatments that involved host alternation (CTF) and mixing (MIX) showed a decrease in the number of DVGs relative to the initial population, likely due to the presence of tomato plants during viral passages.

Representative figures for each population and subpopulation are provided in Sup. Fig. 2. Notably, none of the PVYNb subpopulations presented 5' copy back/snap back-type DVGs, which were exclusive to some PVYSt populations. Interestingly, different types of DVGs were present in early PVYNb, but only deletion-type DVGs were found in PVYSt. However, as PVYSt replicated in benthamiana, it recovered all DVG types, increasing diversity. In PVYNb, the same DVG formation pattern was observed in intermediate and final passages in benthamiana, demonstrating DVG stability in this host.

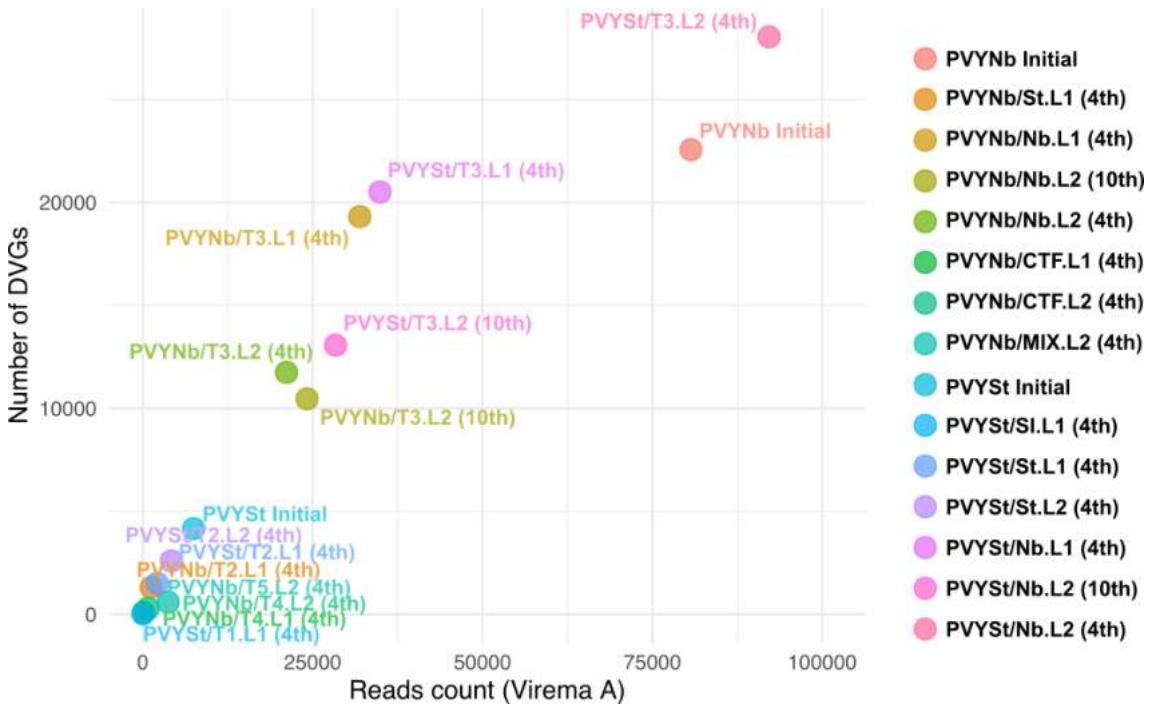


Fig 2. Number of reads mapped against the reference genome using ViReMa-A (*x*-axis) against the number of distinct DVGs detected in each PVY strain. Each point represents a subpopulation.

Curiously, although the number of DVGs decreased drastically when PVYNb transitioned to different hosts from benthamiana, different patterns were observed in different treatments. For example, in PVYNb/St.L1, only reverse deletion-type DVGs were maintained in potatoes. Conversely, in the alternating host treatment (CTF.L2), only insertion-type DVGs persisted. In the mixing hosts treatment, forward deletion and reverse insertion-type DVGs were maintained. Even more intriguingly, despite the significant decrease in DVG numbers, treatments using only potato plants presented exactly the same DVG types as the initial PVYSt, even after four passages.

This dataset underscores the influence of the host on DVG formation, with benthamiana showing high permissiveness and stability in DVG populations. The restrictive environment in tomato plants and the variable patterns observed in different treatments highlight the complexity of host-virus interactions in DVG dynamics.

Clustering DVGs

In this analysis, PCA was applied to all mapped reads with the goal of simplifying our dataset by focusing on DVG types. The results, illustrated in Fig. 3, reveal intriguing patterns across different datasets and subpopulations.

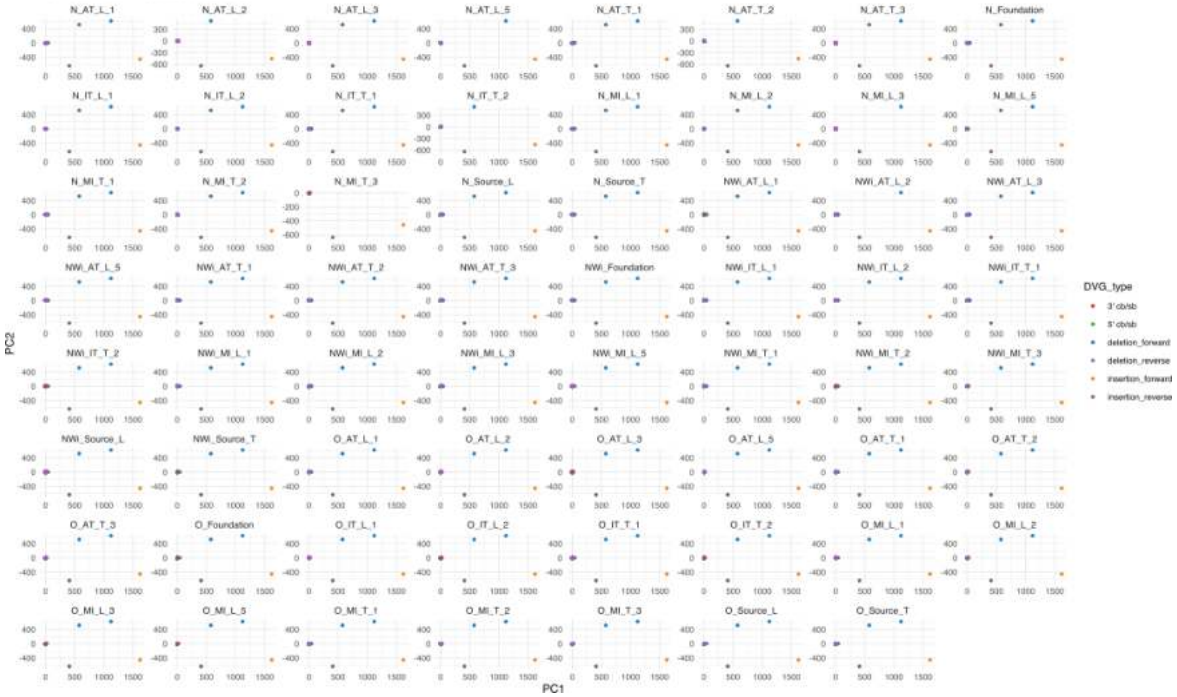


Fig 3. Clustering of DVGs type using PCA to first dataset. The "N, NWi or O" value is referenced to the PVY strain, followed by the type of transmission (AT to aphid-transmission, IT to infected tuber and MI to mechanical infection), the plant organ (L to leaf or T to tuber) and passage number.

For the first dataset, the PCA results show a consistent clustering pattern among the various DVG types across all samples. While small differences are evident between samples, these become more pronounced in later generations. An example is the subpopulation of the PVY^N transmitted by aphids in leaves, where a clear distinction emerges between the 1st and 2nd passages. However, this difference seems to be more related to the reduction in DVGs in the 2nd passage rather than any inherent change in the DVG types themselves.

A key observation is the distinct clustering of specific DVG types. For instance, DVGs formed by 3' and 5' cb/sb are closely grouped together, often alongside DVGs

created by reverse deletion. This pattern is consistent across most populations. In some cases, such as in the PVY$^{N\text{-}Wi}$-infected tubercles during the 2nd passage, two groups of reverse insertion DVGs are formed, with one group clustering closely with the 3', 5' cb/sb, and reverse deletion types. In contrast, forward insertion DVGs are clearly differentiated, consistently clustering far from the other DVG types. Additionally, forward deletion DVGs tend to form two distinct groups within nearly all populations.

The PCA analysis across both datasets underscores the complexity and variability of DVG formation and maintenance in PVY populations. The consistent clustering of certain DVG types, such as the 3' and 5' cb/sb (grouped on axis 0 along with reverse deletion), across different subpopulations suggests that some DVG types are inherently more stable or prevalent. However, the distinct clustering of forward insertion DVGs and the variability in forward deletion DVGs indicate that other types of DVGs may be more sensitive to factors such as viral passage, transmission mode, and host plant.

The PCA results for the second dataset reveal similar clustering patterns, with some nuances and slight variations in grouping, as shown in Fig. 4. Like the first dataset, the 3' and 5' cb/sb DVGs consistently cluster closely together across all populations. However, forward insertion DVGs again stand out by forming distinct clusters, separate from other DVG types.

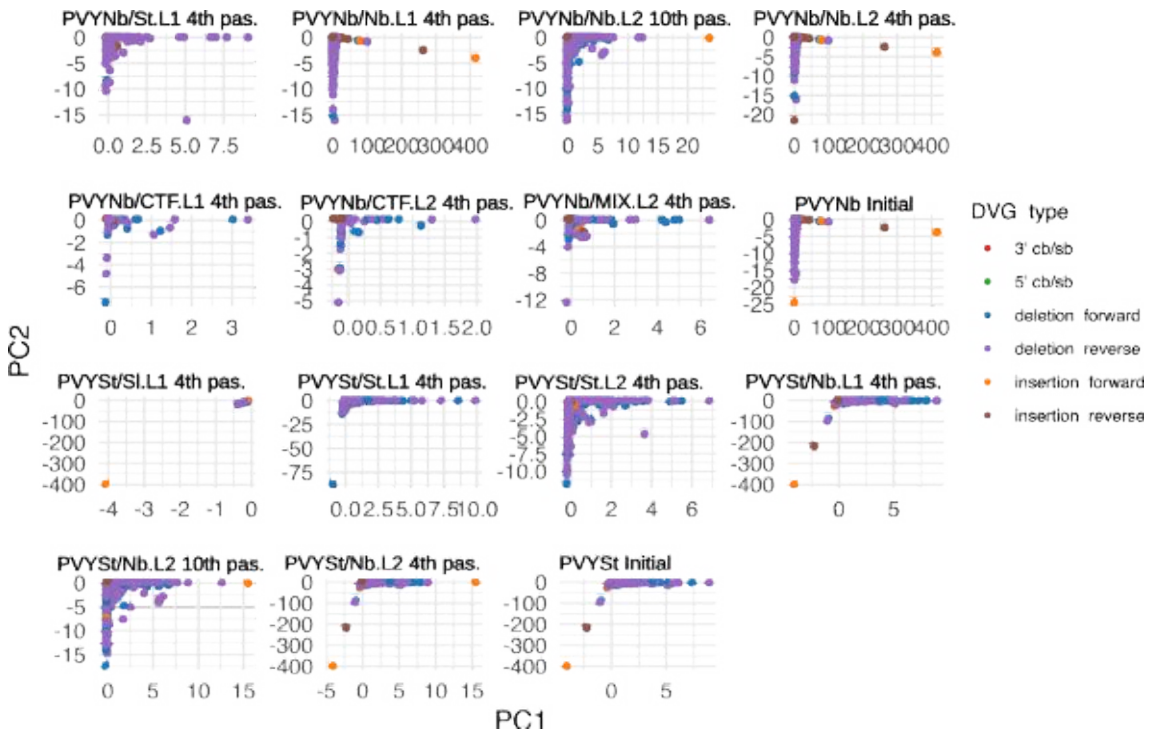


Fig 4. PCA analysis of DVGs using PCA to the second dataset, in which each dot represents a unique DVG separated by different colors by type.

Within the second dataset, there are several noteworthy observations. For example, the initial PVYNb and intermediate generations of CTF show high PC values, indicating a strong influence or differentiation along the main component. As these populations advance, such as in PVYSt Nb.L2 by the 10th generation, the PC values decrease significantly, suggesting reduced differentiation between the various DVG types formed. In benthamiana populations (Nb) infected with PVYSt, the clustering pattern remains similar to the initial isolate, but like PVYNb, this pattern only persists through intermediate generations. By the 10th generation in Nb.L2, the DVGs begin to form tighter clusters, indicating less variation in PVYSt population.

The PVYSt Sl.L1 tomato lineage presents a unique cluster, with forward insertion DVGs clearly distinct from the other DVG types within this subpopulation. Interestingly, when host alternation (CTF) or mixing (MIX) is applied, the established DVG patterns are disrupted, leading to distinct clusters that differ both from the initial PVYNb and from each other.

Furthermore, the influence of the host plant on DVG formation is particularly evident in the second dataset, where host-specific clusters emerge, and the introduction of host alternation or mixing disrupts established DVG patterns. This highlights the importance of the host environment in shaping DVG diversity and evolution.

Overall, these findings suggest that DVG populations in PVY are dynamic and could be influenced by multiple factors, including virus strain, transmission mode, passage number, and host plant.

DVGs formation and distribution

By filtering the most frequently observed DVGs by type, we were able to analyze their formation and distribution across different subpopulations. The results are detailed in Fig. 5.

The first dataset reveals a strikingly similar pattern of DVG formation between the N-Wi and O strains. Many DVGs appear consistently across all populations, with the forward insertion DVG (ID: ++2973-2973) being the most prevalent. However, not

all DVG types share this uniform presence-particularly, 3' cb/sb DVGs are absent in all populations.

Interestingly, the N strain demonstrates a distinct pattern compared to N-Wi and O. N populations have a high frequency of reverse insertion DVGs (ID: --2970-2970), which, although present in N-Wi and O, appear in much smaller numbers. Conversely, N-Wi and O show a higher prevalence of forward deletion DVGs (ID: ++7159-7166), which are less frequent in the N strain. Additionally, the O strain uniquely features a significant number of reads mapped to another forward deletion DVG (ID: ++7320-7322).

Fig 5. Heatmap showing the most found unique DVGs, divided by type and number of reads to first dataset. Each different population can be found in the abscissa and each unique DVG ID in the ordinate, divided by type.

While the strains generally form similar populations, some variations are observed, particularly within the N strain. A pattern of DVG formation is evident in the founder plant and is maintained in both source plants and 1st passage plants, regardless of transmission type or organ analyzed. However, after a 2nd inoculation, this pattern is

lost, leading to a reduction in DVG diversity and selection of fewer DVGs. This selective reduction does not occur in the N-Wi and O strains, where subpopulations maintain a consistent pattern despite some specific differences.

The second dataset does not exhibit a clear pattern of DVG formation (Fig. 6). Each population seems to possess unique DVGs with minimal similarity to other populations, and no DVGs are widely found across all samples. Notably, none of the 3' or 5' cb/sb DVGs met the threshold for inclusion in this analysis.

Despite the lack of a clear formation pattern, some trends are observed. In PVYNb, DVGs present in the initial population are filtered out in subsequent passages. Consecutive passages in the same host (*e.g.*, St and Nb) appear to refine the DVGs compared to the initial population. However, when hosts are alternated (CTF) or mixed (MIX), the DVG profile changes markedly, for example, eliminating all forward deletion DVGs and favoring the formation of reverse insertion DVGs.

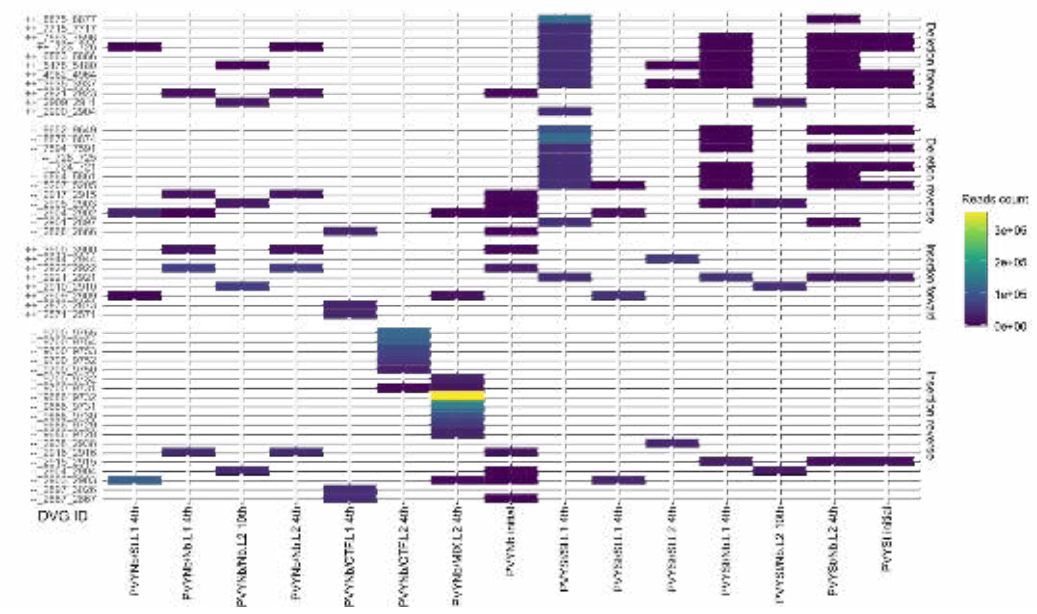


Fig 6. Heatmap showing the most found unique DVGs, divided by type and to number of reads count to the second dataset. Each different population can be found in the abscissa and each unique DVG ID in the ordinate, divided by type.

For PVYSt, there is a tendency to retain DVGs from the initial population in intermediate benthamiana passages (Nb.L1 and L2). Yet, by the 10th passage in the

advanced Nb.L2 population, the DVG profile undergoes significant filtering, drastically altering its composition. This filtering is also evident in potato populations (St). In contrast, the DVG diversity increases sharply in tomato (Sl) compared to the initial population, forming distinct patterns while still preserving some original DVGs.

The analysis of DVG formation across different strains and subpopulations reveals distinct patterns in DVG types and their persistence over successive passages. While N-Wi and O strains show a consistent DVG formation pattern, the N strain diverges significantly, particularly in its reverse insertion DVG prevalence. The second dataset highlights the impact of host consistency on DVG selection, with alternating or mixed hosts resulting in a more diverse DVG profile. These findings underscore the complex dynamics of DVG formation and selection, influenced by both viral strain and host interaction.

DVGs diversity

In our analysis, we calculated SH for all populations to directly compare the diversity of different types of DVGs. The results for the first dataset are presented in Fig. 7.

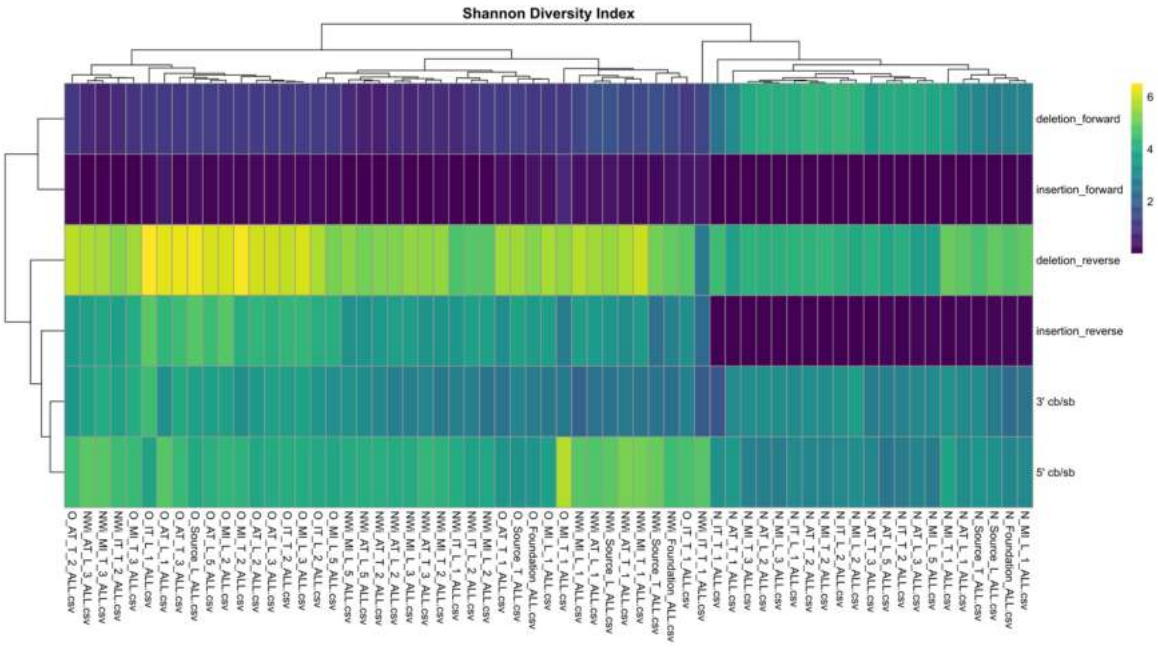


Fig 7. Shannon-Entropy graph to the first dataset, in which the bluer color represents less diversity and the yellow more diversity between each population, represented by each square.

To the first dataset, when examining diversity across samples, there is a noticeable similarity in diversity among the same types of DVGs. Forward deletion and insertion DVGs exhibit similar diversity levels, showing a high degree of conservation across all populations. A similar trend is observed for 3' and 5' cb/sb DVGs. Interestingly, the diversity of reverse insertion DVGs is more closely aligned with the 3' and 5' cb/sb group than with reverse deletion DVGs, which tend to show greater variability compared to other DVG types.

Diversity patterns also tend to cluster by strain, with some exceptions, such as the blending of diversity between PVY^{O} and $PVY^{N\text{-}Wi}$. Despite this, PVY^{N} generally displays a unique diversity pattern, distinct from the other strains, with the exception of $PVY^{N\text{-}Wi}$ IT from tubers in the 1st passage. Notably, the diversity of forward deletion DVGs is lower in PVY^{O} and $PVY^{N\text{-}Wi}$, while this trend is inversely proportional to the reverse insertion DVGs when compared to the N strain.

In $PVY^{N\text{-}Wi}$ and O, diversity is predominantly concentrated in reverse deletion DVGs, with lower diversity observed in forward deletion DVGs. Conversely, PVY^{N} shows lower overall diversity, but reverse deletion DVGs still tend to exhibit greater diversity among samples. An interesting contrast is evident between the diversity patterns of $PVY^{N\text{-}Wi}$ and O compared to PVY^{N}. Forward deletion DVGs are less diverse in N-Wi and O but show increased diversity in N. On the other hand, reverse insertion DVGs are highly conserved in PVYN but exhibit greater diversity in the other two strains.

The diversity patterns in the second dataset differ from those observed in the first (Fig. 8). Deletion-type DVGs generally show higher diversity similarity, making them the most diverse DVG type across all samples. Insertion and 3' cb/sb DVGs have similar diversity levels, while 5' cb/sb DVGs are the most conserved type among all samples.

Significant differences are also noted between samples. While there is a general tendency for clusters to originate from the same viral isolate, exceptions exist. For instance, DVG diversity is relatively conserved in tomato (PVYSt/Sl.L1), potatoes inoculated with PVYNb (St.L1), and in the alternate (CTF) and mix (MIX) treatments. However, diversity increases in the remaining populations. Populations found in benthamiana exhibit greater diversity compared to others, clustering together regardless

of the virus or passage. This suggests that benthamiana's more permissive cellular environment may promote higher DVG diversity. In support of this observation, correlation analysis shows a strong positive relationship between the number of DVGs and the reads count in benthamiana (correlation coefficient of 0.929), indicating that as viral replication increases, so does DVG diversity. The linear model further supports this, revealing a significant positive association between the number of DVGs and reads count across all samples, with an R-squared value of 0.6914, highlighting that a substantial proportion of the variability in DVG numbers can be explained by variations in viral replication levels. Thus, the increased diversity observed in benthamiana likely reflects the host ability to support more extensive viral replication and a higher likelihood of replication errors, resulting in greater accumulation of DVGs. This suggests that the more permissive nature of benthamiana cellular environment may facilitate the generation and persistence of a wider range of DVGs. Additionally, a more conserved diversity pattern is observed starting from the initial PVYSt, where both St lines tend to maintain the diversity conformation of the initial population. This finding suggests that DVG diversity formation is more dependent on the host species than on viral passage.

The analysis of DVG diversity across different strains and subpopulations reveals distinct patterns influenced by both DVG type and host-virus interactions. In the first dataset, diversity tends to cluster by strain, with forward deletion and insertion DVGs showing high conservation, while reverse deletion DVGs display more variability. The second dataset highlights the role of host species in shaping DVG diversity, with deletion-type DVGs being the most diverse and benthamiana populations exhibiting the greatest overall diversity. These findings underscore the complex interplay between viral strain, DVG type, and host species in determining the diversity of DVGs.

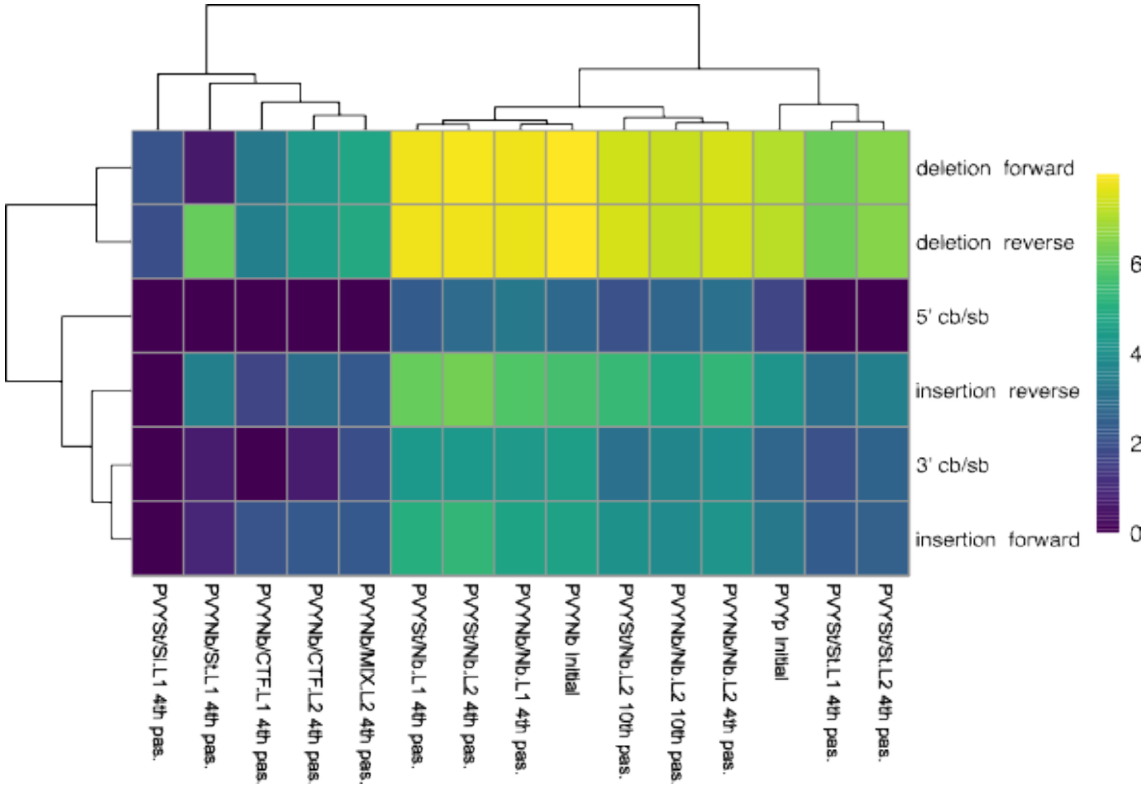


Fig 8. Shannon-entropy graph to the second dataset, in which the bluer color represents less diversity and the yellow more diversity between each population, represented by each square.

Discussion

The viral RdRp naturally introduces errors during replication, leading to high variability in viral genomes. This variability can result in the formation of DVGs, which coexist with WT genomes in infected cells (Vignuzzi and López 2019). Our detection approach successfully identified DVGs across nearly all PVY samples, even with stringent filters. The number of detected DVGs varied by viral strain: PVY^{O} exhibited the highest number of mapped reads and unique DVGs, while $PVY^{N\text{-}Wi}$ had fewer reads but a similar number of DVGs. PVY^{N} had the lowest counts for both metrics. Additionally, PVY accumulation in potato cultivars varies by strain and plant stage, which could affect DVG generation (Mondal et al. 2023). Higher viral accumulation often correlates with increased DVG numbers.

The DVG population also varied with virus strain, transmission mode (AT, IT, or MI), and plant organ (leaves or tubers). During AT, DVGs in leaves decreased with successive passages, while tubers showed variable trends depending on the strain. In MI, similar trends to AT were observed, with a general decrease in DVGs in leaves over time and an increase in tubers for some strains. This suggests an organ-specific influence on DVG dynamics. Although DVGs typically increase with multiple passages (Pogany et al. 1995; Hasiów-Jaroszewska et al. 2012), our results indicated a decrease in DVGs in potato leaves and an increase in tubers during sequential passages. This is important as DVG generation is not always random; virus-encoded sequences can actively promote specific DVGs (Vignuzzi and López 2019), indicating that host-virus interactions are specific.

These findings align well with the diversity analyses, showing high diversity in insect transmission, medium in mechanical inoculation, and low in tuber transmission due to bottlenecks (da Silva et al. 2020). Tubers generally exhibit higher diversity (π) than leaves for tuber and mechanical inoculation (da Silva et al. 2020), suggesting that increased diversity is related to higher DVG numbers. Metabolic activities, development, hormonal responses, and gene expression differences between potato tuber and leaf cells (Taiz et al. 2015) that likely contribute to these variations.

Host factors play a critical role in DVG formation. In benthamiana, the DVG population remained stable for PVYNb and increased significantly for PVYSt, indicating the plant's permissiveness (*discussed in Chapter III*). Conversely, potato samples showed a higher ratio of reads to unique DVGs, suggesting robust DVG formation even with fewer mapped reads. The formation of DVGs depends on both host factors and environmental conditions, such as temperature (Llamas et al. 2004). In this work, we used two datasets, the first one was entirely produced under greenhouse conditions, subject to environmental variations. On the other hand, the second dataset was conducted entirely under artificial conditions (*M&M from Chapter III*), which may have influenced the divergence between both datasets as well. Importantly, no work has yet addressed the issue of environmental factors that can shape the population of DVGs in plants. Host alternation or mixing disrupted established DVG patterns, leading to distinct clusters, further emphasizing the complex interaction between host and

environment. Additionally, DVG formation often results in the loss of specific DVG types and the emergence of new ones in mixed or alternate host passages.

DVGs can interfere with WT viruses, potentially reducing virulence, protecting the host, or generating an immune response (Rabinowitz and Huprikar 1979; Barrett and Dimmock 1984). Recent studies have demonstrated the efficacy of DVGs in treating viral infections in animals (also known as therapeutic interfering particles, TIPs) (Rezelj et al. 2021; Xiao et al. 2021). Studies using this strategy in planta are non-existent, although interfering DVGs has been previously reported to for others virus genera (Graves et al. 1996; Hasiów-Jaroszewska et al. 2018) and DVGs construction was done using plant hosts (Pathak and Nagy 2009; Lee and White 2014). But unlike the use of animal cells that can generate different DVGs *in vitro* and *in vivo* (Li et al. 2024), the use of plants for such studies may simplify testing and yield results closer to real-world conditions. This strategy holds potential for a non-transgenic and efficient viral control method. However, further research is necessary for effective implementation.

A major challenge in utilizing DVGs as therapeutic agents lies in isolating those with antiviral properties from the broader array of defective genomes produced during WTV replication. Three primary mechanisms through which DI RNAs interfere with viral processes have been identified: (i) competing with the virus and host for resources, thereby hindering viral replication and reducing symptom severity; (ii) inducing posttranscriptional gene silencing (PTGS), leading to gene silencing; and (iii) altering the functions of viral factors (Szittya et al. 2002; Pathak and Nagy 2009; Lukhovitskaya et al. 2013). Methodologies for interrogating DVGs and identifying potential therapeutic candidates have been explored (Rezelj et al. 2021). Our study revealed consistent clustering patterns of specific DVG types across various samples, with 5' DVGs showing a preference in therapeutic applications (Li et al 2024) due to their retention of essential replication regions and their ability to induce strong antiviral immune responses interfering the WT-virus replication. Additionally, 5' DVGs are less likely to revert to fully functional viruses, reducing the risk of generating pathogenic viruses during therapy. Despite the preliminary results found, our *in silico* methodology seeks to understand the formation of DVGs in different virus-host-environment interactions that still need to be validated in the future in bench work.

In conclusion, this study provides valuable insights into the diversity and evolution of DVGs within PVY populations, revealing patterns that reflect the complex interplay between viral genetics, host factors, and evolutionary pressures. Further research into the mechanisms driving these patterns and the functional consequences of different DVG types will be essential for a deeper understanding of DVG biology and its implications for viral fitness, pathogenicity, and host interactions. DVGs, particularly those retaining the 5' end of the viral genome, hold promise as therapeutic agents. Although preliminary, our findings suggest that DVGs could serve as a basis for developing novel antiviral strategies, particularly in plant systems. Further research is needed to validate the therapeutic potential of DVGs and to explore their application in viral control. Investigating the specific mechanisms driving DVG formation and their effects on viral fitness and host interactions will be crucial for advancing this area.

References

Barrett ADT, Dimmock NJ (1984) Modulation of semliki forest virus-induced infection of mice by defective-interfering virus. Journal of Infectious Diseases 150:98–104. https://doi.org/10.1093/infdis/150.1.98

Beauclair G, Mura M, Combredet C, et al (2018) DI-tector: defective interfering viral genomes detector for next-generation sequencing data. RNA 24:1285–1296. https://doi.org/10.1261/rna.066910.118

Budzyńska D, Minicka J, Hasiów-Jaroszewska B, Elena SF (2020) Molecular evolution of tomato black ring virus and *de novo* generation of a new type of defective RNAs during long-term passaging in different hosts. Plant Pathology 69:1767–1776. https://doi.org/10.1111/ppa.13258

Burgyan J, Grieco F, Russo M (1989) A defective interfering RNA molecule in cymbidium ringspot virus infections. Journal of General Virology 70:235–239. https://doi.org/10.1099/0022-1317-70-1-235

Calvert LA, Cuervo MI, Ospina MD, et al (1996) Characterization of cassava common mosaic virus and a defective RNA species. Journal of General Virology 77:525–530. https://doi.org/10.1099/0022-1317-77-3-525

Chang YC, Borja M, Scholthof HB, et al (1995) Host effects and sequences essential for accumulation of defective interfering RNAs of cucumber necrosis and tomato bushy stunt tombusviruses. Virology 210:41–53. https://doi.org/10.1006/viro.1995.1315

Che X, Mawassi M, Bar-Joseph M (2002) A novel class of large and infectious defective RNAs of citrus tristeza virus. Virology 298:133–145. https://doi.org/10.1006/viro.2002.1472

da Silva W, Kutnjak D, Xu Y, et al (2020) Transmission modes affect the population structure of potato virus Y in potato. PLoS Pathology 16:e1008608. https://doi.org/10.1371/journal.ppat.1008608

Damayanti TA, Nagano H, Mise K, et al (1999) Brome mosaic virus defective RNAs generated during infection of barley plants. Journal of General Virology 80:2511–2518. https://doi.org/10.1099/0022-1317-80-9-2511

Dawkins R (2016) The selfish gene, 40th anniversary. Oxford University Press

de Oliveira Resende R, de Haan P, de Avila AC, et al (1991) Generation of envelope and defective interfering RNA mutants of tomato spotted wilt virus by mechanical passage. Journal of General Virology 72:2375–2383. https://doi.org/10.1099/0022-1317-72-10-2375

de Oliveira Resende R, de Haan P, van de Vossen E, et al (1992) Defective interfering L RNA segments of tomato spotted wilt virus retain both virus genome termini and have extensive internal deletions. Journal of General Virology 73:2509–2516. https://doi.org/10.1099/0022-1317-73-10-2509

Gard S, von Magnus P (1947) Studies on interference in experimental influenza: purification and centrifugation experiments. Ark Kemi Mineral Geologi 24:1–4.

Graves MV, Pogany J, Romero J (1996) Defective interfering RNAs and defective viruses associated with multipartite RNA viruses of plants. Seminars in Virology 7:399–408. https://doi.org/10.1006/smvy.1996.0048

Graves MV, Roossinck MJ (1995) Characterization of defective RNAs derived from RNA 3 of the Fny strain of cucumber mosaic cucumovirus. Journal of Virology 69:4746–4751. https://doi.org/10.1128/jvi.69.8.4746-4751.1995

Hasiów-Jaroszewska B, Borodynko N, Figlerowicz M, Pospieszny H (2012) Two types of defective RNAs arising from the tomato black ring virus genome. Archives of Virology 157:569–572. https://doi.org/10.1007/s00705-011-1200-z

Hasiów-Jaroszewska B, Minicka J, Zarzyńska-Nowak A, et al (2018) Defective RNA particles derived from Tomato black ring virus genome interfere with the replication of parental virus. Virus Research 250:87–94. https://doi.org/10.1016/j.virusres.2018.04.010

Huang AS, Baltimore D (1970) Defective viral particles and viral disease processes. Nature 226:325–327. https://doi.org/10.1038/226325a0

Inoue-Nagata AK, Jordan R, Kreuze J, et al (2022) ICTV Virus Taxonomy Profile: *Potyviridae* 2022. Journal of General Virology 103:001738. https://doi.org/10.1099/jgv.0.001738

Johnson M, Zaretskaya I, Raytselis Y, et al (2008) NCBI BLAST: a better web interface. Nucleic Acids Research 36. https://doi.org/10.1093/nar/gkn201

Karasev AV, Gray SM (2013) Continuous and emerging challenges of potato virus Y in potato. Annual Review of Phytopathology 51:571–586. https://doi.org/10.1146/annurev-phyto-082712-102332

Karasev AV, Hu X, Brown CJ, et al (2011) Genetic diversity of the ordinary strain of Potato virus Y (PVY) and origin of recombinant PVY strains. Phytopathology 101:778–785. https://doi.org/10.1094/phyto-10-10-0284

Knorr DA, Mullin RH, Hearne PQ, Morris TJ (1991) *De novo* generation of defective interfering RNAs of tomato bushy stunt virus by high multiplicity passage. Virology 181:193–202. https://doi.org/10.1016/0042-6822(91)90484-S

Koonin EV, Starokadomskyy P (2016) Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 59:125–134. https://doi.org/10.1016/j.shpsc.2016.02.016

La Scola B, Desnues C, Pagnier I, et al (2008) The virophage as a unique parasite of the giant mimivirus. Nature 455:100–104. https://doi.org/10.1038/nature07218

Lazzarini RA, Keene JD, Schubert M (1981) The origins of defective interfering particles of the negative-strand RNA viruses. Cell 26:145–154. https://doi.org/10.1016/0092-8674(81)90298-1

Lee PKK, White KA (2014) Construction and characterization of an Aureusvirus defective RNA. Virology 452–453:67–74. https://doi.org/10.1016/j.virol.2013.12.033

Li X, Ye Z, Plant EP (2024) 5′ copyback defective viral genomes are major component in clinical and non-clinical influenza samples. Virus Research 339:199274. https://doi.org/10.1016/j.virusres.2023.199274

Llamas S, Sandoval C, Babin M, et al (2004) Effect of the host and temperature on the formation of defective RNAs associated with broad bean mottle virus infection. Phytopathology 94:69–75. https://doi.org/10.1094/phyto.2004.94.1.69

Lukhovitskaya NI, Thaduri S, Garushyants SK, et al (2013) Deciphering the mechanism of defective interfering RNA (DI RNA) biogenesis reveals that a viral protein and the DI RNA act antagonistically in virus infection. Journal of Virology 87:6091–6103. https://doi.org/10.1128/jvi.03322-12

Mondal S, Wintermantel WM, Gray SM (2023) Infection dynamics of potato virus Y isolate combinations in three potato cultivars. Plant Disease 107:157–166. https://doi.org/10.1094/pdis-09-21-1980-re

Olmo-Uceda MJ, Muñoz-Sánchez JC, Lasso-Giraldo W, et al (2022) DVGfinder: a metasearch tool for identifying defective viral genomes in RNA-seq data. Viruses 14. https://doi.org/10.3390/v14051114

Omarov RT, Rezende JAM, Scholthof HB (2004) Host-specific generation and maintenance of tomato bushy stunt virus defective interfering rnas. molecular plant-microbe interactions 17:195–201. https://doi.org/10.1094/mpmi.2004.17.2.195

Pathak KB, Nagy PD (2009) Defective Interfering RNAs: foes of viruses and friends of virologists. Viruses 1:895–919. https://doi.org/10.3390/v1030895

Pogany J, Romero J, Huang Q, et al (1995) *De novo* generation of defective interfering-like RNAs in broad bean mottle bromovirus. Virology 212:574–586. https://doi.org/10.1006/viro.1995.1515

Posit team (2024) RStudio: integrated development environment for R. Posit Software, PBC, Boston, MA.

Rabinowitz SG, Huprikar J (1979) The influence of defective-interfering particles of the PR-8 strain of influenza a virus on the pathogenesis of pulmonary infection in mice. Journal of Infectious Diseases 140:305–315. https://doi.org/10.1093/infdis/140.3.305

Rezelj VV, Carrau L, Merwaiss F, et al (2021) Defective viral genomes as therapeutic interfering particles against flavivirus infection in mammalian and mosquito hosts. Nature Communications 12:2290. https://doi.org/10.1038/s41467-021-22341-7

Roossinck MJ (1997) Mechanisms of plant virus evolution. Annual Review of Phytopathology 35:191–209. https://doi.org/ 10.1146/annurev.phyto.35.1.191

Routh A, Johnson JE (2014) Discovery of functional genomic motifs in viruses with ViReMa–a Virus Recombination Mapper–for analysis of next-generation sequencing data. Nucleic Acids Research 42:e11–e11. https://doi.org/10.1093/nar/gkt916

Rubio L, Tian T, Yeh H, et al (2002) *De novo* generation of lettuce infectious yellows virus defective RNAs in protoplasts. Molecular Plant Pathology 3:321–327. https://doi.org/10.1046/j.1364-3703.2002.00125.x

Rubio L, Yeh H-H, Tian T, Falk BW (2000) A heterogeneous population of defective RNAs is associated with lettuce infectious yellows virus. Virology 271:205–212. https://doi.org/10.1006/viro.2000.0318

Sanjuán R, Agudelo-Romero P, Elena SF (2009) Upper-limit mutation rate estimation for a plant RNA virus. Biology Letters 5:394–396. https://doi.org/10.1098/rsbl.2008.0762

Sanjuán R, Domingo-Calap P (2021) Genetic diversity and evolution of viral populations. In: Encyclopedia of Virology. Elsevier, pp 53–61

Singh RP, Valkonen JPT, Gray SM, et al (2008) Discussion paper: the naming of Potato virus Y strains infecting potato. Archives of Virology 153:1–13. https://doi.org/10.1007/s00705-007-1059-1

Szittya G, Molnár A, Silhavy D, et al (2002) Short defective interfering RNAs of tombusviruses are not targeted but trigger post-transcriptional gene silencing against their helper virus. Plant Cell 14:359–372. https://doi.org/10.1105/tpc.010366

Taiz L, Zeiger E, Møller IM, Murphy AS (2015) Plant physiology and development, 6th ed. Sinauer Associates, Incorporated, Publishers.

Torrance L, Cowan GH, Sokmen MA, Reavy B (1999) A naturally occurring deleted form of RNA 2 of potato mop-top virus. Journal of General Virology 80:2211–2215. https://doi.org/10.1099/0022-1317-80-8-2211

Tromas N, Zwart MP, Maïté P, Elena, SF (2014) Estimation of the *in vivo* recombination rate for a plant RNA virus. Journal of General Virology 95:724-732. https://doi.org/10.1099/vir.0.060822-0

Vignuzzi M, López CB (2019) Defective viral genomes are key drivers of the virus–host interaction. Nature Microbiology 4:1075–1087. https://doi.org/10.1038/s41564-019-0465-y

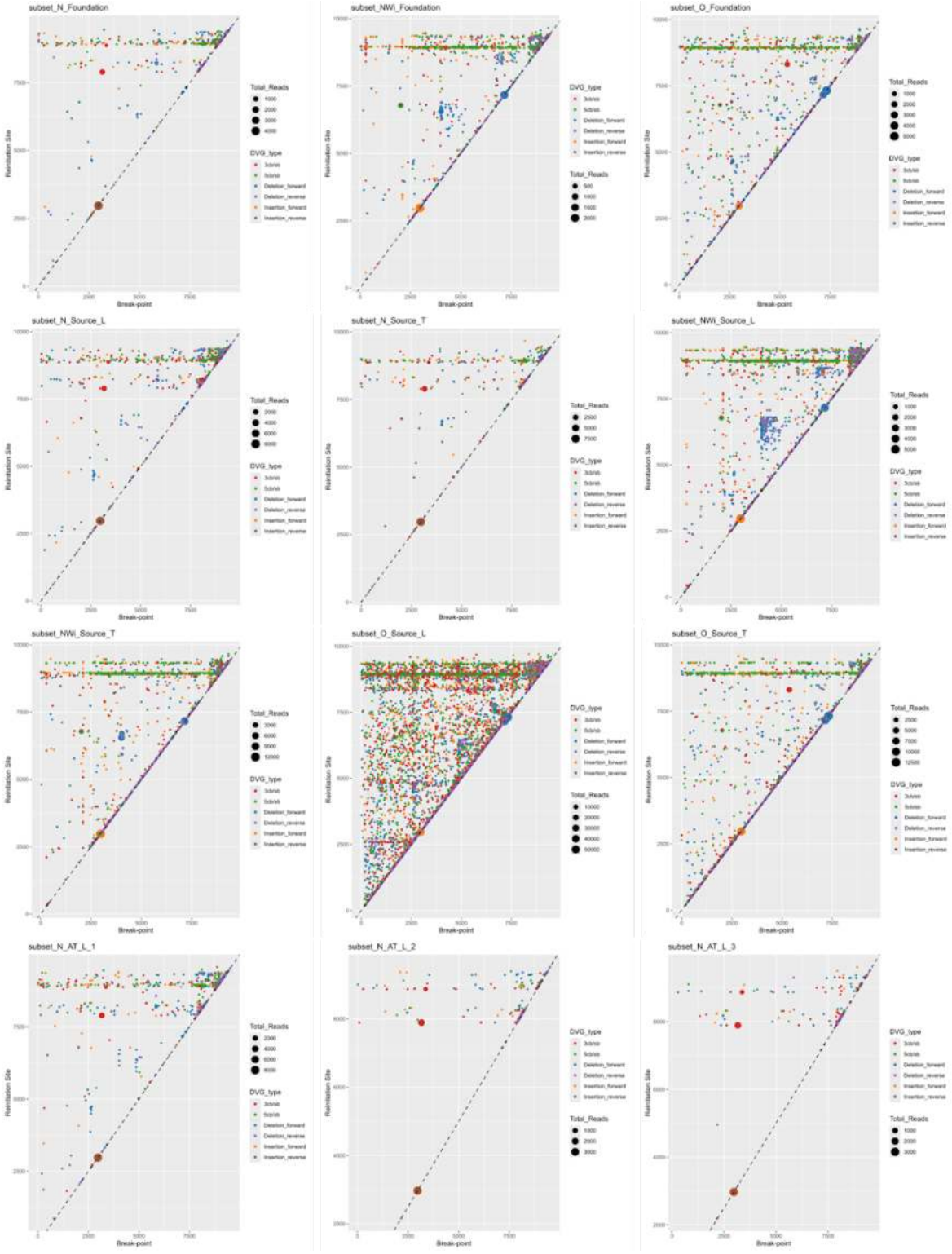
Visser PB, Brown DJF, Brederode FTh, Bol JF (1999) Nematode transmission of tobacco rattle virus serves as a bottleneck to clear the virus population from defective interfering RNAs. Virology 263:155–165. https://doi.org/10.1006/viro.1999.9901

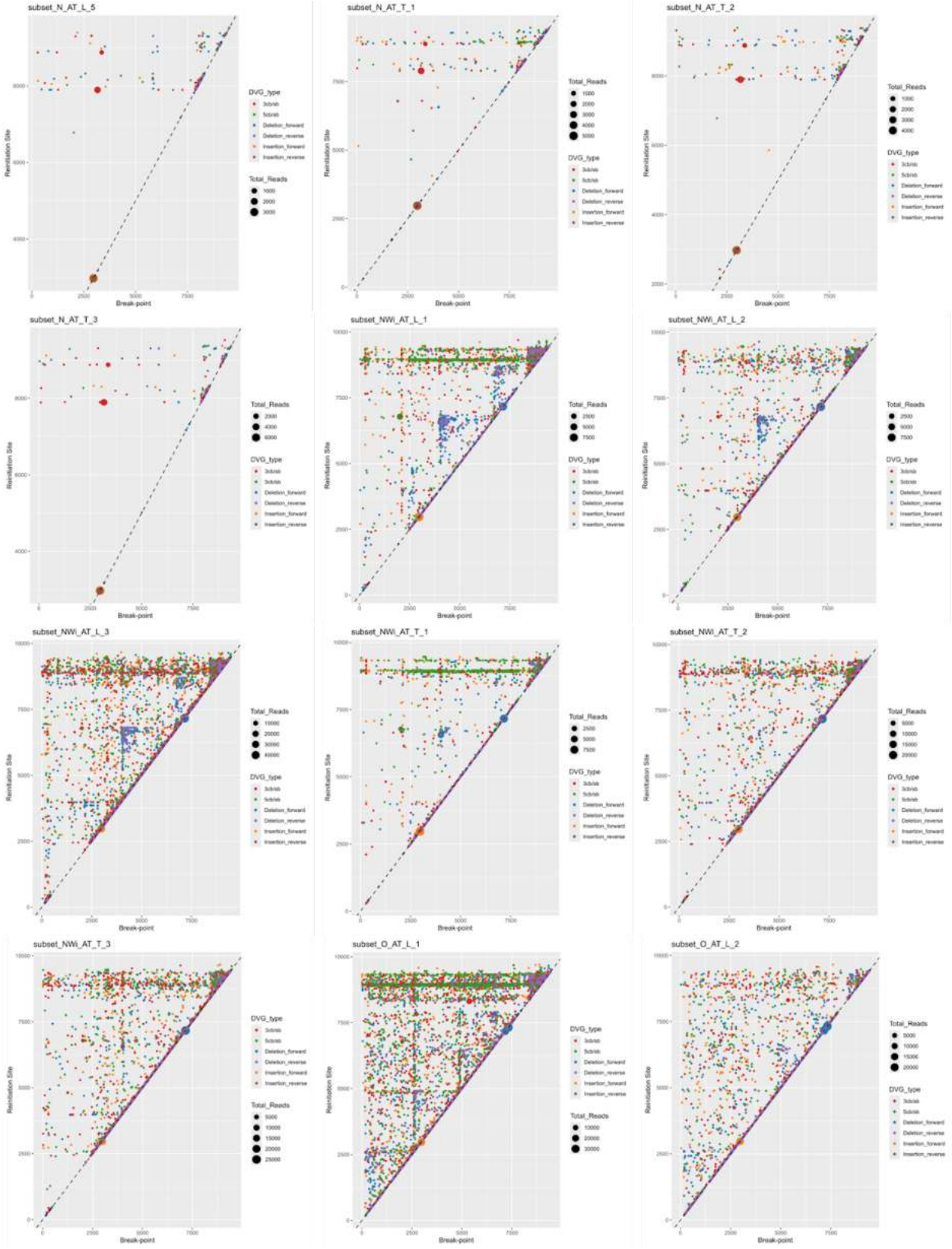
von Magnus P (1954) Incomplete Forms of Influenza Virus. Advances in Virus Research 2:59-79. https://doi.org/ 10.1016/s0065-3527(08)60529-1

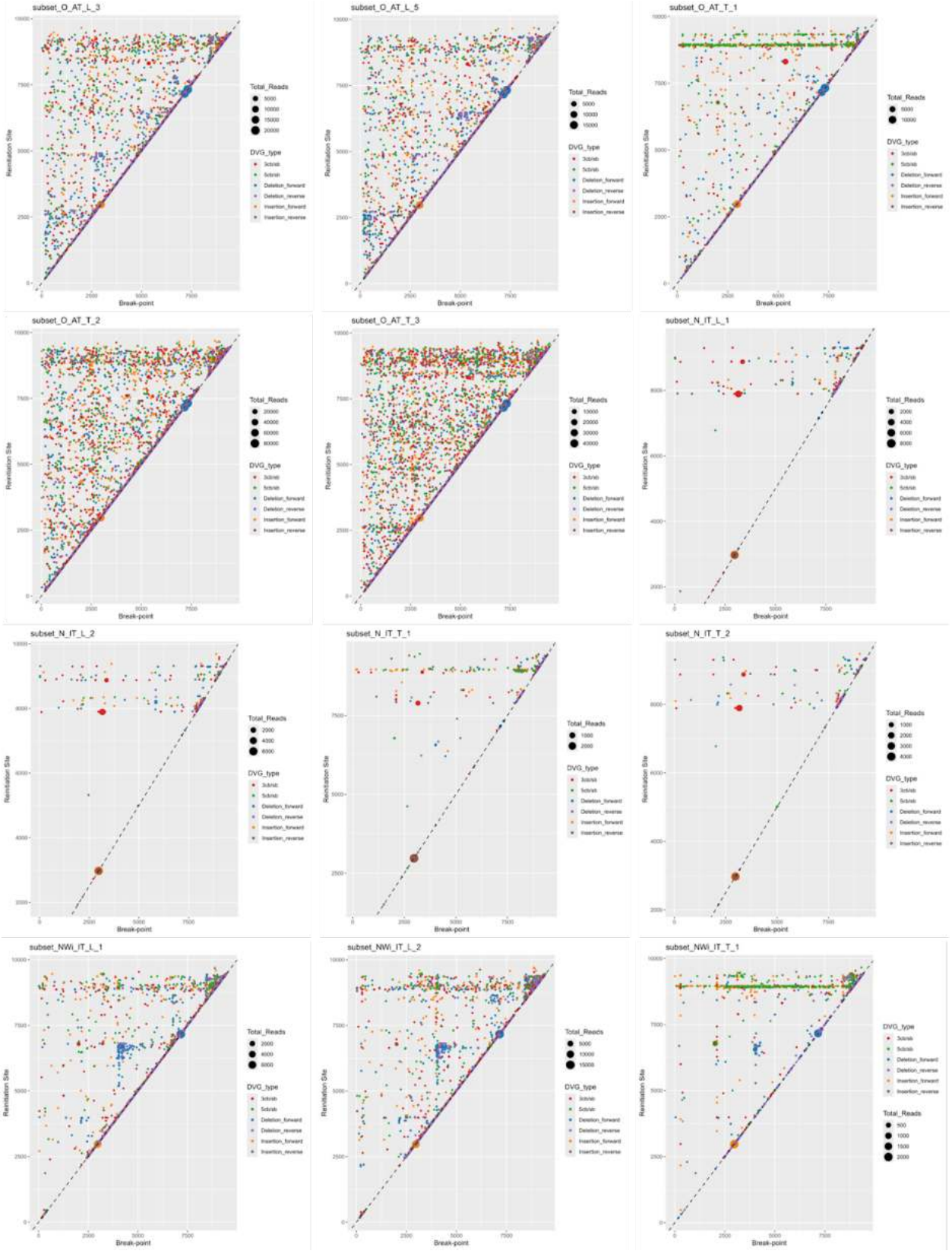
White KA, Bancroft JB, Mackie GA (1992) Coding capacity determines *in vivo* accumulation of a defective RNA of clover yellow mosaic virus. Journal of Virology 66:3069–3076. https://doi.org/10.1128/jvi.66.5.3069-3076.1992

Wickham H (2016) Ggplot2: elegant graphics for data analysis, 2nd ed. Springer International Publishing, Cham, Switzerland.

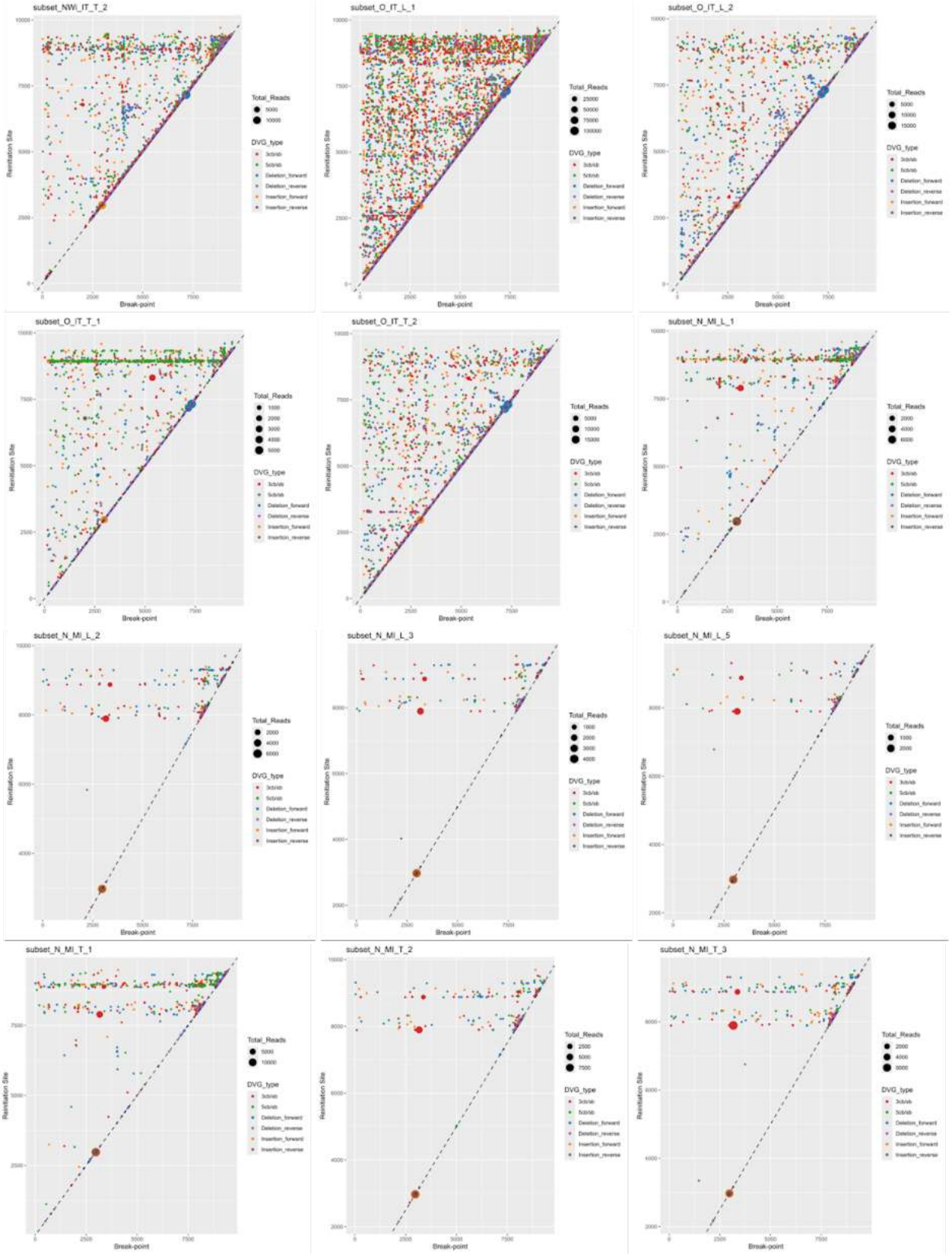
Xiao Y, Lidsky PV., Shirogane Y, et al (2021) A defective viral genome strategy elicits broad protective immunity against respiratory viruses. Cell 184:6037-6051.e14. https://doi.org/10.1016/j.cell.2021.11.023


subset_N_Foundation
subset_NWI_Foundation
subset_O_Foundation
subset_N_Source_L
subset_N_Source_T
subset_NWI_Source_L
subset_NWI_Source_T
subset_O_Source_L
subset_O_Source_T
subset_N_AT_L_1
subset_N_AT_L_2
subset_N_AT_L_3
Reinitiation Site
Break-point
Total_Reads
DVG_type

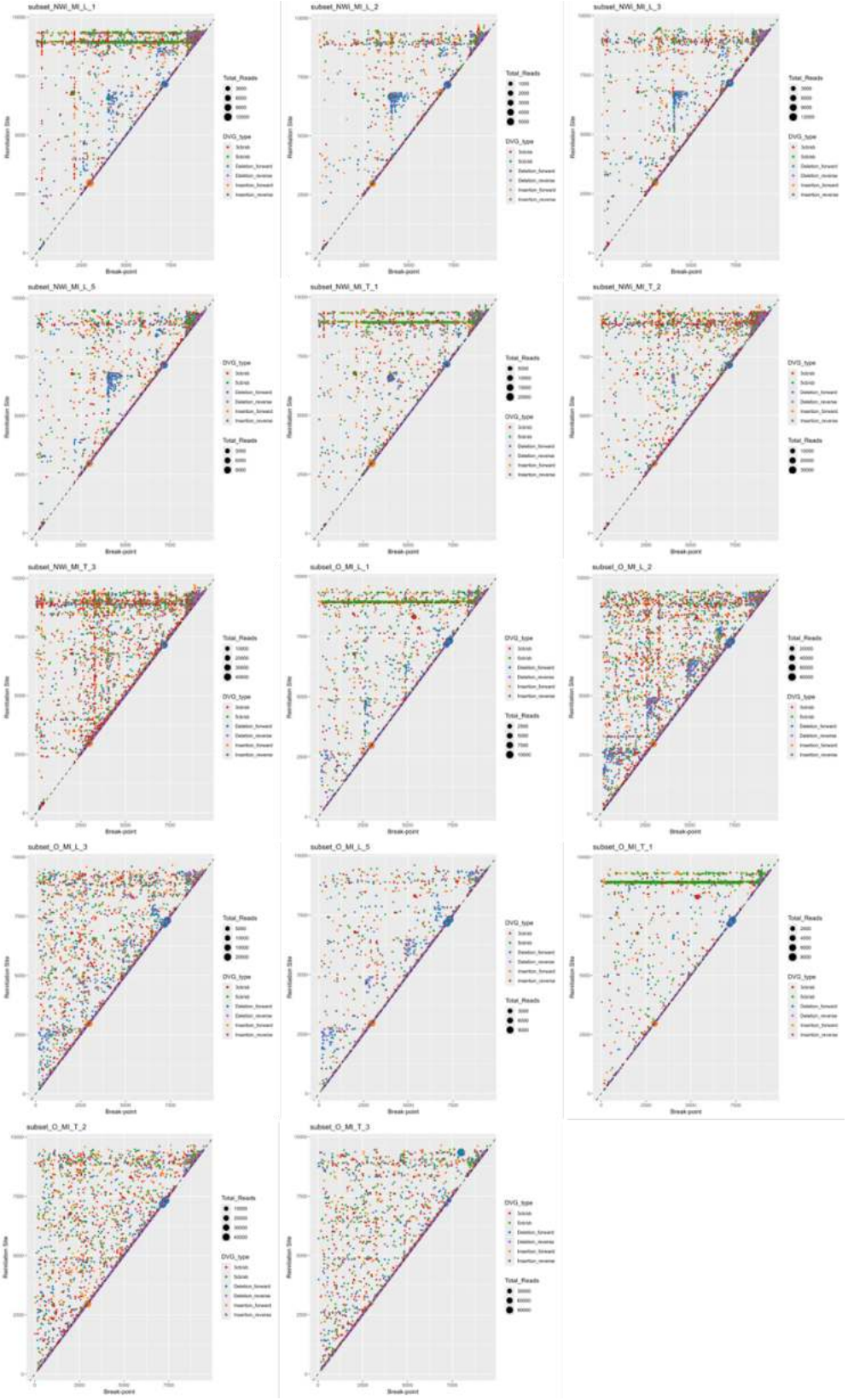

subset_N_AT_L_5
subset_N_AT_T_1
subset_N_AT_T_2
subset_N_AT_T_3
subset_NWi_AT_L_1
subset_NWi_AT_L_2
subset_NWi_AT_L_3
subset_NWi_AT_T_1
subset_NWi_AT_T_2
subset_NWi_AT_T_3
subset_O_AT_L_1
subset_O_AT_L_2
Reinitiation Site
Break-point
DVG_type
Total_Reads


subset_O_AT_L_3
subset_O_AT_L_5
subset_O_AT_T_1
subset_O_AT_T_2
subset_O_AT_T_3
subset_N_IT_L_1
subset_N_IT_L_2
subset_N_IT_T_1
subset_N_IT_T_2
subset_NWi_IT_L_1
subset_NWi_IT_L_2
subset_NWi_IT_T_1
Reinitiation Site
Break-point
Total_Reads
DVG_type
3dbrkb
5dbrkb
Deletion_forward
Deletion_reverse
Insertion_forward
Insertion_reverse

831

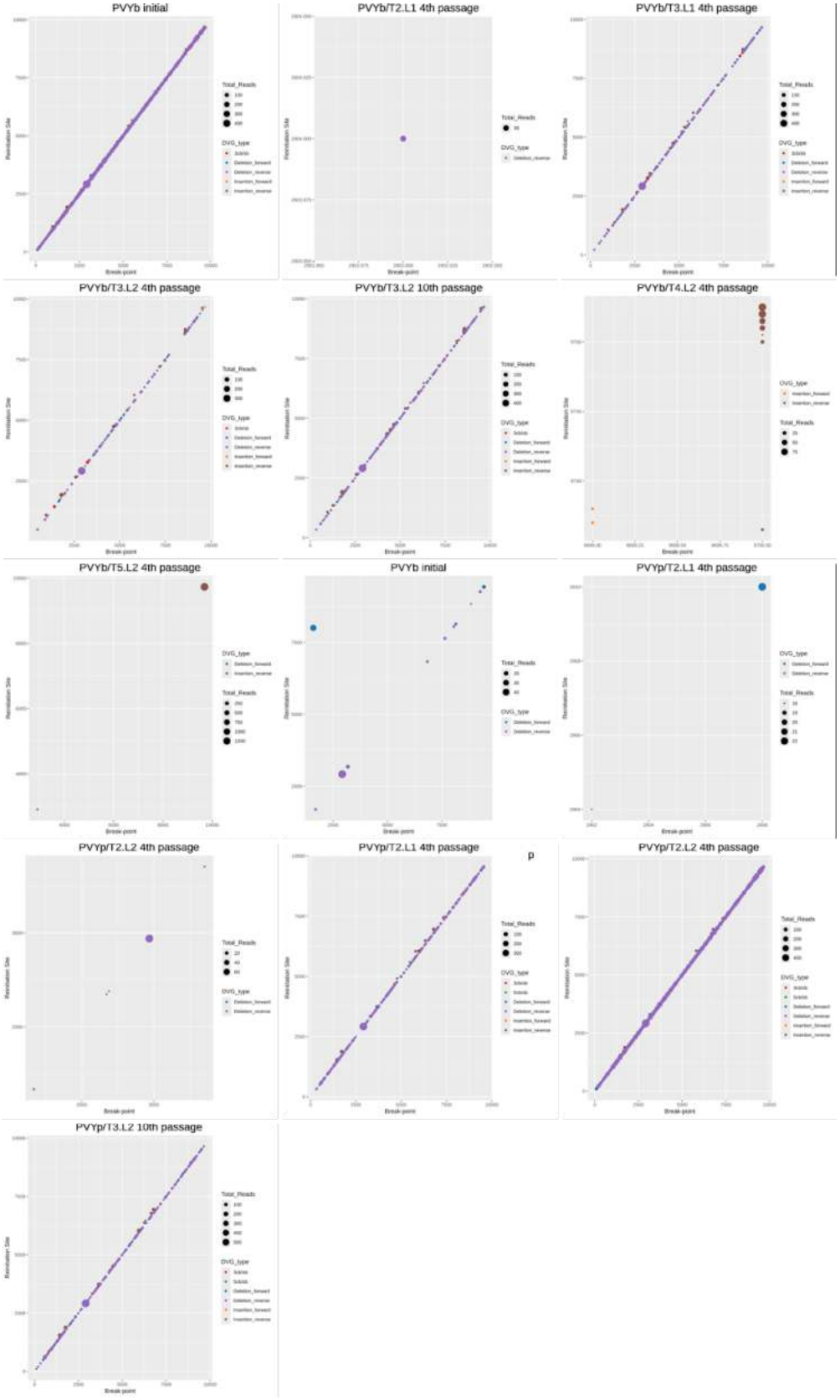

subset_NWi_IT_T_2
subset_O_IT_L_1
subset_O_IT_L_2
subset_O_IT_T_1
subset_O_IT_T_2
subset_N_MI_L_1
subset_N_MI_L_2
subset_N_MI_L_3
subset_N_MI_L_5
subset_N_MI_T_1
subset_N_MI_T_2
subset_N_MI_T_3
Reinitiation Site
Break-point
Total_Reads
DVG_type
3cb/sb
5cb/sb
Deletion_forward
Deletion_reverse
Insertion_forward
Insertion_reverse

832



Sup Fig 1. DVGs conformation to each population, divided by strain (O, N or NWi), type of transmission (AT, MI or IT), plant organ (L or T) and passage (1-5) of first

dataset, obtained in (da Silva, 2020). Each point represents a mapped read against the reference genome of PVY (X12456).


PVYb initial
PVYb/T2.L1 4th passage
PVYb/T3.L1 4th passage
PVYb/T3.L2 4th passage
PVYb/T3.L2 10th passage
PVYb/T4.L2 4th passage
PVYb/T5.L2 4th passage
PVYb initial
PVYp/T2.L1 4th passage
PVYp/T2.L2 4th passage
PVYp/T2.L1 4th passage
PVYp/T2.L2 4th passage
PVYp/T3.L2 10th passage

838

Sup Fig 2. DVGs conformation to each population, divided by two PVY isolates (PVYNb and PVYSt), treatment (Sl, St, Nb, CTF and MIX) and number of passages (1st to 10th) of second dataset, same as used in the Chapter III. Each point represents a mapped read.

Sup. Table 1. Total number of reads that mapped against the PVY reference genome using ViReMa-A and the number of DVGs found in each population to the first dataset.

Virus	Strain	Transmission mode	Organ	Passage	Reads count ViReMa-A	Number of DVGs
PVY	N			Foundation	9078	714
PVY	N		L	Source	16409	1206
PVY	N		T	Source	17974	731
PVY	N	AT	L	1	16608	1140
PVY	N	AT	L	2	6959	453
PVY	N	AT	L	3	5717	401
PVY	N	AT	L	5	6098	414
PVY	N	AT	T	1	10421	645
PVY	N	AT	T	2	8421	567
PVY	N	AT	T	3	11858	376
PVY	N	IT	L	1	13814	487
PVY	N	IT	L	2	11225	635
PVY	N	IT	T	1	5787	446
PVY	N	IT	T	2	7209	400
PVY	N	MI	L	1	17014	1519
PVY	N	MI	L	2	9947	563
PVY	N	MI	L	3	7320	537
PVY	N	MI	L	5	4958	354
PVY	N	MI	T	1	25376	1287
PVY	N	MI	T	2	16323	655
PVY	N	MI	T	3	11216	809
PVY	N-Wi			Foundation	5883	1538
PVY	N-Wi		L	Source	13290	3302
PVY	N-Wi		T	Source	29333	2577
PVY	N-Wi	AT	L	1	23094	4230
PVY	N-Wi	AT	L	2	20875	3052
PVY	N-Wi	AT	L	3	106606	9642
PVY	N-Wi	AT	L	5	37130	2630

PVY	N-Wi	AT	T	1	22524	2939
PVY	N-Wi	AT	T	2	48376	4441
PVY	N-Wi	AT	T	3	56586	4886
PVY	N-Wi	IT	L	1	16762	2521
PVY	N-Wi	IT	L	2	35662	2889
PVY	N-Wi	IT	T	1	5325	1305
PVY	N-Wi	IT	T	2	3220	4296
PVY	N-Wi	MI	L	1	29545	4307
PVY	N-Wi	MI	L	2	16878	2656
PVY	N-Wi	MI	L	3	25552	3231
PVY	N-Wi	MI	L	5	25227	3256
PVY	N-Wi	MI	T	1	57056	4210
PVY	N-Wi	MI	T	2	66610	5056
PVY	N-Wi	MI	T	3	101846	8278
PVY	O			Foundation	12939	1743
PVY	O		L	Source	174904	11311
PVY	O		T	Source	37556	2773
PVY	O	AT	L	1	98431	11310
PVY	O	AT	L	2	70946	4526
PVY	O	AT	L	3	60976	4917
PVY	O	AT	L	5	54190	4706
PVY	O	AT	T	1	37980	2505
PVY	O	AT	T	2	244520	7948
PVY	O	AT	T	3	140772	7348
PVY	O	IT	L	1	260815	10606
PVY	O	IT	L	2	55415	3998
PVY	O	IT	T	1	15016	2196
PVY	O	IT	T	2	55350	4196
PVY	O	MI	L	1	31045	3363
PVY	O	MI	L	2	252454	7868
PVY	O	MI	L	3	72229	5498
PVY	O	MI	L	5	37005	3063

PVY	O	MI	T	1	21571	3613
PVY	O	MI	T	2	142749	6165
PVY	O	MI	T	3	86750	4818

845

Sup Table 2. Total number of reads that mapped against the PVY reference genome using ViReMa-A and the number of DVGs found in each population to the second dataset.

Virus	Treatment	Line	Passage	Host	Reads count ViReMa-A	Number of DVGs
PVYNb	Initial			Benthamiana	80679	22557
PVYNb	St	L1	4	Potato	1231	1312
PVYNb	Nb	L1	4	Benthamiana	31936	19304
PVYNb	Nb	L2	4	Benthamiana	21145	11744
PVYNb	Nb	L2	10	Benthamiana	24171	10464
PVYNb	CTF	L1	4	Tomato	92	97
PVYNb	CTF	L2	4	Tomato	758	280
PVYNb	MIX	L2	4	Mix	3728	594
PVYSt	Initial			Potato	7454	4154
PVYSt	Sl	L1	4	Tomato	23	19
PVYSt	St	L1	4	Potato	2094	1501
PVYSt	St	L2	4	Potato	4192	2592
PVYSt	Nb	L1	4	Benthamiana	34921	20503
PVYSt	Nb	L2	4	Benthamiana	92179	28032
PVYSt	Nb	L2	10	Benthamiana	28378	13067

Concluding remarks

As the global population grows, so does the need for increased food production. However, various factors can affecting the productivity of cultivated plants, with viruses posing a significant challenge. PVY has long been known as an obstacle to sustainable agriculture, and addressing the development of crops with high resistance to PVY infection was always one of the top priorities. The most relevant challenges though are the lack of resistance sources for a specific crop (*e.g.*, potatoes) and the emergence of the so-called resistance-breaking isolates.

Our research employed diverse approaches to uncover the genetic variations and phenotypic impacts of different PVY isolates on various hosts. We focused on understanding the importance of identifying isolates from different crops, as even isolates of the same species can yield vastly different results in experimental settings. While genetic differences in PVY are influenced by multiple factors, we were specifically interested on the role of the host. From some advances in these aspects, future research should aim to unravel the molecular mechanisms that determine an isolate's ability to infect a particular host. This knowledge is vital for crafting effective resistance strategies.

In addition to exploring PVY genetic diversity, we wanted to facilitate generation of genome data by producing an easy and fast protocol. We found out that Nanopore sequencing technology offers a promising alternative, providing rapid, cost-effective, and accurate results comparable to Illumina sequencing.

There are still unresolved questions that need further investigation, such as identifying the most affected genomic regions during host switching and understanding the molecular interactions between viral and host factors. This includes studying the structural roles of proteins and intrinsically disordered proteins, as well as assessing the current level of protection in potato cultivars against PVY. These insights may be gained through a deeper analysis of the genome. Notably, we have detected the formation of DVGs in PVY populations for the first time. While this discovery requires further validation, it may represent the first step for developing new non-transgenic control strategies, which is one of our goals.

We have not yet answered all the questions posed at the outset of this research, but we believe our findings provide a crucial foundation for understanding and mitigating the impact of PVY on agriculture. Our study lays the groundwork for future research and control measures, including the development of resistant cultivars, targeted antiviral treatments, and integrated pest management strategies.