

RESEARCH

Open Access



Global phylodynamics of two relevant aphid-transmitted viruses in cucurbit crops: cucurbit aphid-borne yellows virus and watermelon mosaic virus

M. P. Rabadán^{1*}  and P. Gómez¹

Abstract

Cucurbit aphid-borne yellows virus (CABYV) and watermelon mosaic virus (WMV) are major plant pathogens that cause severe epidemics in cucurbit crops. While there has been an increasing interest in molecular epidemiological studies on both viruses at regional scales, their phylodynamic analysis by using the temporal data at global scale remains unexplored. In this study, we implemented the Nextstrain phylodynamic approach to comprehensively examine the coat protein gene and full-length genome sequences of the CABYV and WMV worldwide populations. Our analyses reconstructed a robust phylogeny of CABYV and confirmed the occurrence of isolates grouped into three clusters (Asian, Mediterranean, and Recombinant). Nextstrain analysis suggested that CABYV epidemics likely originated in Southeast Asia in fourteenth century, while the Mediterranean population emerged in Spain in seventeenth century. We also found a high divergence between Asian and Mediterranean isolates, with low genetic diversity and scarce evidence of selection, as reflected by the lack of gene flow. Moreover, the hierarchical analysis of molecular variance revealed a significant differentiation between CABYV populations grouped by geographical location and plant host. Additionally, the global phylogenetic reconstruction of the WMV population confirmed a clear differentiation among isolates, which grouped into two clusters (classical and emergent), and Nextstrain analysis suggested that WMV epidemics were most probably originated in USA during the sixteenth century. The initial WMV population diverged in the eighteenth century, with the origin of the emergent population in the nineteenth century. Our analysis confirmed that WMV population has a polyphyletic origin, defining an extensive genetic diversity. Overall, this work provides insights into the CABYV and WMV origin and evolutionary dynamics, gaining an understanding of the global spread of both viral diseases. Additionally, the integration of all spatio-temporal CABYV and WMV data, along with their genome sequence data by open access (<https://github.com/PiR92>) and the Nextstrain approach, provides a valuable tool for compiling and sharing current knowledge of these viral diseases in cucurbit crops.

Keywords CABYV, Cucurbits, Molecular epidemiology, Phylodynamics, WMV

*Correspondence:

M. P. Rabadán
mprabadan@cebas.csic.es

¹ Centro de Edafología y Biología Aplicada del Segura (CEBAS)-CSIC, Dpto de Biología del Estrés y Patología Vegetal, PO Box 164, 30100 Murcia, Spain

Background

Plant viral epidemics poses a major threat to global food security (Jones and Naidu 2019; Ristaino et al. 2021). The emergence and spread of viral diseases in crops are influenced by intrinsic viral and host factors, as well as ecological, agronomical, and socio-economic factors (Vurro



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

et al. 2010; Elena et al. 2014; González et al. 2019). For instance, RNA viruses exhibit rapid rates of molecular evolution that coupled with high population sizes promoting genetic diversity on the viral populations (Duffy and Holmes 2008). Likewise, diversified agricultural practices, with intensive and extensive production systems, make challenges to control viral diseases (Jones 2009). The use of monoculture practices can reduce ecosystem diversity and lead to the emergence of viral diseases (Alexander et al. 2014; Roossinck and García-Arenal 2015). Furthermore, environmental heterogeneity has been shown to shape plant-virus infection networks at temporal and spatial scales (Valverde et al. 2020). Concurrently, climate change has also been associated to changes in the viral disease progression, virulence, and viral load, as well as disease dispersal by influencing the spread of insect-vector (Trebicki 2020; Montes and Pagán 2022). These aspects indicate that viral diseases in plant crops are highly dynamic, and with the increasing global trade of food that may introduce severe disease outbreaks, it becomes important to conduct molecular epidemiological studies that allow understanding the temporal and spatial changes of the viral populations during epidemics (Jeger 2020; Jones 2021). In this sense, recent advances in sequencing techniques, bioinformatics, and computational analysis, have allowed accurate inference of the evolutionary histories of viral diseases (Grenfell et al. 2004; Drummond and Rambaut 2007; Pybus et al. 2012). As such, phylodynamic inference, which integrates molecular epidemiology and evolutionary biology, is therefore essential for understanding evolutionary histories and the structure and genetic diversity of viral populations at spatiotemporal scales. Despite the importance of this phylodynamic inference in examining the origins, virus genotype, and how it becomes established in new areas, this approach remains elusive in plant viral diseases. Increasing our understanding on plant viral phylodynamics can facilitate the development of accurate diagnoses, breeding programs, and efficient management strategies to protect the crops from viral diseases.

Within the framework of plant viruses, phylodynamic analyses have been used to examine the origin and dispersion patterns at different landscape scales. For instance, the spread of citrus tristeza virus (CTV) over three decades worldwide was found to have originated in North America and spread initially to South America, and then spread to Europe and Asia (Benítez-Galeano et al. 2017). Similarly, a global study of plum pox virus (PPV) analysed two groups of PPV strains and found that the virus emerged in Eurasia, but the population structure show that strains were from two different introductions in Europe (Hajizadeh et al. 2019). Further

studies have dated the origin of epidemic viruses and analyzed the timescale of the spread around the world. It has been reported that turnip mosaic virus (TuMV) was probably originated from a virus of wild orchids in Germany and observed a pattern of dispersal from Europe to Asia, along with a timescale that correlated with the adaptation to new hosts and agricultural development (Yasaka et al. 2017). Also, a study of migration patterns of begomoviruses in wild peppers in Mexico suggested that short-distance movement is caused by the insect-vectors, while long-distance movement is due to the trade of plant materials (Rodelo-Urrego et al. 2013). However, a crucial aspect of phylodynamic studies is the interpretation, visualization, and dissemination of results, which is increasingly becoming complex (Theys et al. 2019). The Nextstrain project has recently addressed this drawback by providing the visualization and dissemination of pathogens at global scale (Hadfield et al. 2018). This is a publicly accessible repository that uses genomic data (full or partial) of pathogens to provide a real-time interactive view of pathogen spread and evolution, in addition to maintain a repository on GitHub that allows collaborations between different entities (Hadfield et al. 2018). Thus, Nextstrain has been used to perform several transmission and spatio-temporal dynamics studies on viruses causing human diseases such as Zika, Ebola, and SARS-CoV-2, among others (Gardy and Loman 2018; Proserpi et al. 2022), and definitely provides a valuable basis for further investigation with the application in plant viruses, such as it has been performed for tomato brown rugose virus (ToBRFV)(van de Vossen et al. 2020) and cassava-infecting geminiviruses (<https://nextstrain.org/community/pestdisplace/CMDAFRICA>).

Potyvirus and polerovirus are plant RNA viruses that are globally distributed, with some species causing damaging diseases to a broad range of crops (Gibbs and Ohshima 2010; Garcia-Ruiz et al. 2020). Among them, cucurbit aphid-borne yellows virus (CABYV, genus *Polerovirus*) and watermelon mosaic virus (WMV, genus *Potyvirus*) are known to be widely distributed, affecting the quality and yields of cucurbit crops (Desbiez et al. 2020; De Moya-Ruiz et al. 2021; Radouane et al. 2021; Moya-Ruiz et al. 2023; Rabadán et al. 2023). In fact, recent CABYV outbreaks have been reported from Germany, Poland, Slovenia, Bulgaria, and Pakistan (Zarzyńska-Nowak et al. 2019; Ahsan et al. 2020; Mehle et al. 2020; Menzel et al. 2020; Radeva-Ivanova et al. 2022), including a novel CABYV variant causing more severe symptoms in watermelon in Spain (Rabadán et al. 2021), as well as other isolates affecting passion fruit in Brazil (Vidal et al. 2023), peanuts in Kenya (Mabele et al. 2022), or spiny gourd in Sri Lanka (Abeykoon et al. 2018). In this sense, molecular epidemiological studies have shown that

CABYV isolates can be classified into three phylogenetic groups based on full-length genomic sequence: Asian, Mediterranean, and Recombinant (Kwak et al. 2018; Vidal et al. 2023; Rabadán et al. 2023). Phylogenetic analysis among CABYV isolates have been mostly focused on partial sequences at regional level, revealing distinct subgroups within the Asian and Mediterranean CABYV populations (Bananej et al. 2009; Shang et al. 2009; Kassem et al. 2013). However, no large-scale phylodynamic study has been carried out on the CABYV population. Otherwise, WMV isolates can be classified into three molecular groups based on the amino acid sequences of the coat protein N-terminal region: Group 1 or classical (CL) isolates, Group 2 isolates, and Group 3 emergent isolates (EM), with its isolates divided into four genetic subgroups (EM-1, 2, 3, and 4) (Desbiez et al. 2009; Juarez et al. 2013; Bertin et al. 2020; Rabadán et al. 2023). The accumulation and transmission efficiency of EM isolates is higher than those of CL and Group 2 isolates, which could explain their wide distribution and displacement of both groups by the EM group in the Mediterranean basin (Desbiez et al. 2009; Juarez et al. 2013; Bertin et al. 2020). Furthermore, a close phylogenetic relationship has been observed between classical isolates from Italy, France, Iran, and Turkey, and similarly with emerging isolates from Italy with Chinese, South Korean, and Argentinean isolates (Hajizadeh et al. 2017). This similarity raises the hypothesis of multiple introductions of WMV across several countries (Bertin et al. 2020). However, this statement has yet to be supported by phylodynamic studies at large scale. Since molecular epidemiological investigations of CABYV and WMV have been conducted at the local or regional level, the aim of this study was to comprehensively examine full-length genomes and coat protein (CP) gene sequences of both viruses at global scale. In this study, we implemented the Nextstrain phylodynamic approach to understand their evolutionary dynamics and track the global spread of both viral diseases at real-time in a freely shared system. Furthermore, CABYV population structure and genetic diversity were also examined at complete viral entity.

Results

CABYV Nextstrain build

To explore the epidemic history and evolutionary dynamics of CABYV, we conducted a comprehensive analysis of the CP gene sequence of 393 isolates from 32 countries or regions, as well as, the full-length genome sequences of 74 CABYV isolates from 10 countries or regions, by using the Nextstrain approach. We first performed a principal component analysis (PCA) based on the geographical origin of CABYV isolates, and found three CABYV well-supported clades (Asian, Mediterranean,

and Recombinant) (Additional file 1: Figure S1a). These clades were confirmed by the Nextstrain phylogenies of both CP gene and full-length genome sequences (Fig. 1a, b, and for more detailed information on the phylogeny and the geographic distribution, visit the following domains: https://nextstrain.org/community/PiR92/CABYV_CP and <https://nextstrain.org/community/PiR92/CABYV>). Based on CP gene data, CABYV epidemic could have originated in Southeast Asia in the fourteenth century (1399–1588). Mediterranean isolates were probably originated in Spain in the seventeenth century (1759–1966), with a genetic divergence of 0.11 mutations compared to the Asian isolates. On one hand, the isolates of the Asian clade began to spread across the Asian continent, and five subgroups (AS 1–5) were formed based on their country of origin: AS-1 comprises isolates from Indonesia, AS-2 includes isolates from India and Thailand, AS-3 is composed of isolates from India, AS-4 is formed by isolates from China, South Korea, USA, Serbia, Pakistan, India, Thailand, and Turkey, and AS-5 is made up of the Recombinant group, which includes isolates from Taiwan region, India, Thailand, Papua New Guinea, and Spain. By the other hand, the Mediterranean isolates have mainly spread across the Europe and Africa continents, arising in several countries and regions, grouping into four subgroups (MS 1–4). The MS-1 subgroup consists in Philippines, Taiwan region of China, and Indonesia, MS-2 includes isolates from Taiwan region of China, MS-3 is formed by isolating from Spain and MS-4 is made up from isolates from Czech Republic, Egypt, Germany, Iran, Italy, Montenegro, Morocco, Pakistan, Poland, Saudi Arabia, Serbia, Slovakia, Slovenia, Sudan, Syria, and Tunisia, with a genetic divergence between 0.13 and 0.20 mutations. Two isolates from Pakistan appeared to have a recombination event between isolates from Asian and Mediterranean clades (Fig. 1a). Although the number of isolates in the CABYV populations for which full-length genome sequencing data was compiled was relatively lower than for CP sequencing data, we sought to perform Nextstrain analysis from the full-length genome data and reconstruct CABYV phylogenies. Similarly, the Asian isolates clustered into four subgroups according to the country/region (AS 1–4). The AS-1 subgroup is formed by being isolated from China and USA, AS-2 is composed from isolated from South Korea and China, AS-3 included isolated from South Korea and Japan, additionally, AS-4 contains recombinant isolates from India, Chinese mainland, and Taiwan region of China (Fig. 1b), with a genetic divergence that ranged between 0.16 and 0.18 mutations (Fig. 1b). Consistent with the CP analysis, Mediterranean isolates could have originated in the seventeenth century in Spain, comprising the Spanish and French subgroups.

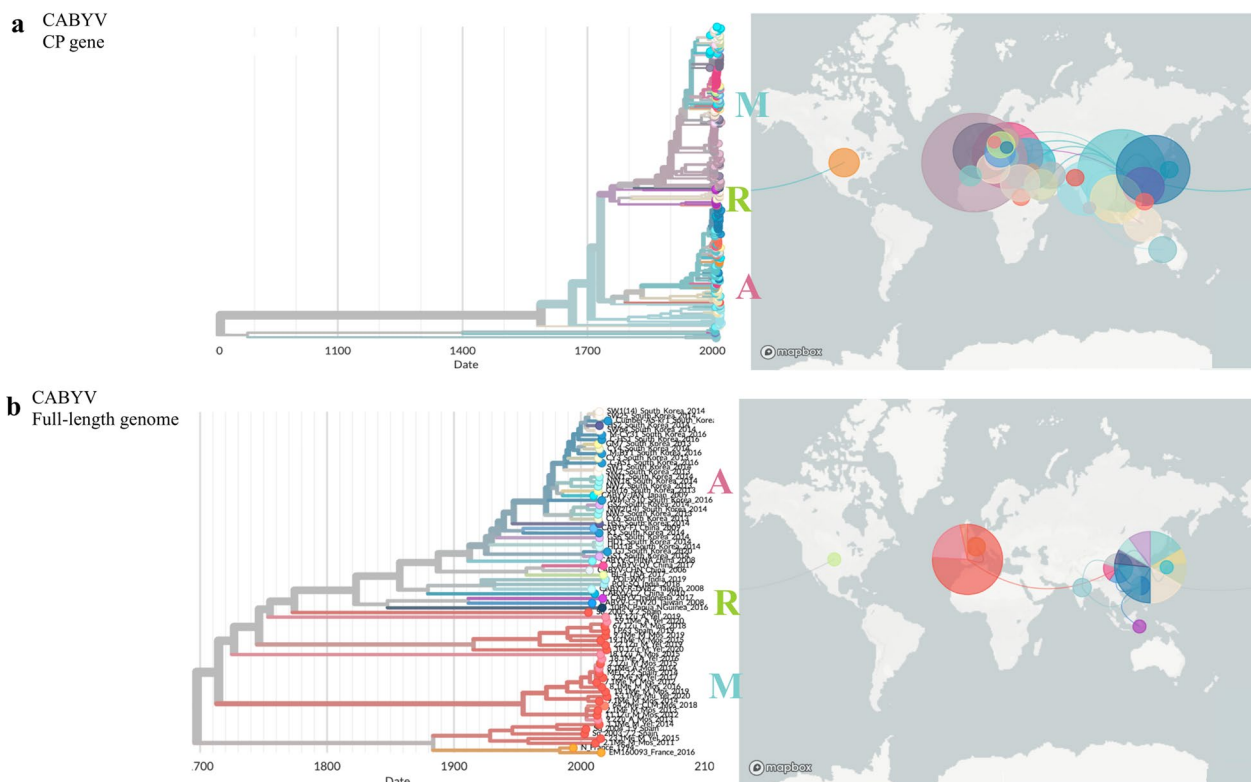


Fig. 1 Nextstrain builds of CABYV from coat protein and complete genome. Both tools are open and hosted on the following domains https://nextstrain.org/community/PiR92/CABYV_CP@main and <https://nextstrain.org/community/PiR92/CABYV@main>. **a** Phylodynamic of CABYV from 393 CP gene sequences collected from 1994 to 2021. **b** Phylodynamic of CABYV from 74 complete genomes collected from 1994 to 2021. In both panels, a time tree is shown on the left, and the geographic distribution is displayed on the right. The compilation information is presented in three main panels: clustering of genomic diversity, geographic origin of samples, and relative genetic diversity. With the associated metadata, users can color the tree nodes based on the region, country, and host plant. The internal node colors indicate the expected ancestral state of a given trait, and confidence in that state is conveyed by the colour saturation of the trait. Branch lengths are shown as a function of divergence or time. The viral genotypes represented in the tree are plotted on a map, which allows to observe the level of geographical spread and the most likely transmission events

The French population probably emerged in 1883 (1774–1951). Likewise, host-based genetic differences analysis revealed similar clades among the isolates. Furthermore, this CABYV phylodynamics population had an averaged substitution nucleotide rate of 1.23×10^{-4} per year.

Inferring genetic diversity and population structure of CABYV

To further understand the role of geographic differentiation in the CABYV population, we estimated genetic diversity (π), genetic differentiation between CABYV populations (K_{ST}^* , S_{nn} , and F_{ST}) and tests of population demography or neutrality (Tajima's D, Fu, and Li F^* , and Fu's FS). At this point, these analyses were performed using the full-length genome of CABYV isolates, grouped either by phylogenetic relationship (Asian, Mediterranean, or recombinant) or geographical origin (Chinese mainland, France, India, Indonesia, Japan, Papua New Guinea, South Korea, Spain, Taiwan of China, and the

United States). The recombinant isolates exhibited the highest diversity ($\pi=0.169$), while the Mediterranean isolates had the lowest nucleotide diversity ($\pi=0.029$) (Table 1). The Taiwanese and Brazilian populations showed the highest nucleotide diversity $\pi=0.189$ and $\pi=0.140$, respectively, while the South Korean populations showed the lowest nucleotide diversity ($\pi=0.020$). The recombinant group showed higher number of polymorphic sites and number of mutations ($S=1898$ and $\eta=2296$), with haplotype diversity (Hd) value of 1, indicating recent geographic expansion. The Asian populations had also a high number of polymorphic sites and mutations ($S=2182$ and $\eta=2660$), possibly associated with the high number of segregating sites observed in the Chinese populations, which was also expanding geographically (Hd=1). The Mediterranean population showed the lowest number of polymorphic sites, and thus, the lowest number of mutations ($S=698$ and $\eta=739$), with a Hd value of 0.995. The Spanish and South

Table 1 Estimation of genetic parameters and neutrality tests for CABYV populations grouped either by phylogenetic relationship or geographical origin

	Hd	S	η	K	π	dS	dN	ω	Tajima's D	Fu and Li's F*	Fu's Fs
Phylogroup											
Asian (n=41)	1	2182	2660	357.33	0.061	0.090±0.0041	0.063±0.002	0.7	-1.59	-1.80	-2.10
Mediterranean (n=29)	0.995	698	739	166.818	0.029	0.040±0.0033	0.025±0.0017	0.62	-0.45	-0.81	0.50
Recombinants (n=7)	1	1898	2296	906.071	0.169	0.217±0.009	0.174±0.008	0.8	0.13	0.29	3.75
Geographical origin											
South Korea (n=30)	1	647	676	116.39	0.020	0.025±0.002	0.013±0.0009	0.52	-1.24	-1.60	-3.424
Spain (n=29)	0.995	917	977	187.59	0.033	0.042±0.003	0.027±0.001	0.65	-0.97	-1.78	0.783
China (n=6)	1	1725	1908	773.73	0.140	0.195±0.009	0.14±0.005	0.71			
Brazil (n=3)	1	196	196	130.66	0.022	0.026±0.003	0.017±0.001	0.64			
France (n=2)	1	204	204	204	0.035	0.059±0.007	0.027±0.003	0.45			
India (n=2)	1	283	283	283	0.050	0.069±0.008	0.044±0.004	0.63			
Taiwan, China (n=2)	1	1068	1068	1068	0.189	0.269±0.017	0.018±0.009	0.066			

Hd = diversity of haplotypes/isolates, S = number of polymorphic (segregating) sites, η (Eta) = total number of mutations, k = average number of nucleotide differences between sequences, π = nucleotide diversity, SS = total number of synonymous sites analysed, NS = total number of non-synonymous sites analysed, dS = synonymous nucleotide diversity, dN = non-synonymous nucleotide diversity, D = Tajima's D, Fu and Li's F* and Fu's Fs

Korean populations had the higher number of segregating sites ($S=917$ and $S=647$) and mutations ($\eta=977$ and $\eta=676$), respectively (Table 1). The estimation of the Tajima's D and Fu and Li F* revealed non-significant negative values for the Asian and Mediterranean populations. However, the Fu's FS statistic value was non-significant negative for the Asian population and positive for the Mediterranean and recombinant populations. The diversity in synonymous (dS) and non-synonymous positions (dN) were calculated to infer the strength of selection pressure (Table 1). The ratio dN/dS was similar among CABYV phylogroups ($\omega=0.7$), and there were no populations under positive selection. Whereas, ω ranged between 0.06 and 0.71 among populations grouped by the geographical origin (Table 1). Moreover, selective constrains in each region of CABYV genome of the populations were evaluated, and several codons under positive selection were identified, in particular in ORF2 (Open reading frame 2) (Additional file 2: Table S1).

We then aimed to test the subdivision within and between CABYV populations, considering each of the geographical regions (Additional file 1: Figure S2). Our DAPC analysis (Discriminant principal component analysis) showed that CABYV populations were divided into three subpopulations associated with phylogenetic groups (Additional file 1: Figure S2a), with the isolates from India, Papua New Guinea, Indonesia, and Taiwan of China in the first component, the Spanish and French isolates in the second component, and rest Chinese, South Korean, US and Japanese isolates together. Minimum network analysis (MSN) analysis, as well as the SNP (Single nucleotide Polymorphism)-based PCA for the genome-wide nucleotide variation in CABYV

populations showed a significant genetic differentiation among the CABYV phylogroups (Fig. 2a, b), with the well-defined differentiation of Asian and Mediterranean isolates. Likewise, the geographical separation between Spanish and French isolates, and Chinese, South Korean, USA, and Japanese isolates (Additional file 1: Figure S2b) suggested that these populations likely originated from different populations. In turn, significant genetic differentiation was observed between the South Korean populations and those of Spain, Brazil, France, India, and Taiwan region of China, except for population from the Chinese mainland. We further estimated the genetic differentiation parameters K_{ST}^* , F_{ST} , and Snn, revealing also significant genetic differentiation among the CABYV phylogroups (Table 2). The Snn values were significantly high for the Asian, Mediterranean, and Recombinants populations. Moreover, a high gene flow differentiation was observed between Asian and Mediterranean populations ($F_{ST} = 0.588$), indicating that these populations likely originated from different populations. However, the F_{ST} value was similar in the Asian vs. recombinant and Mediterranean vs. recombinant populations ($F_{ST} = 0.28$), supporting the observation that these recombinants were parents of the Mediterranean and Asian populations. In addition, significant genetic differentiation was observed between the South Korean populations and those of Spain, Brazil, France, India, and Taiwan region of China ($F_{ST} = 0.293-0.896$), except for the China mainland population, where the $F_{ST} = 0.20$ value suggested that these populations are more closely related. Similarly, genetic differentiation was observed among the Spanish, Brazilian, Indian, and Chinese (including Taiwan) populations ($F_{ST} = 0.317-0.859$). In contrast, gene flow

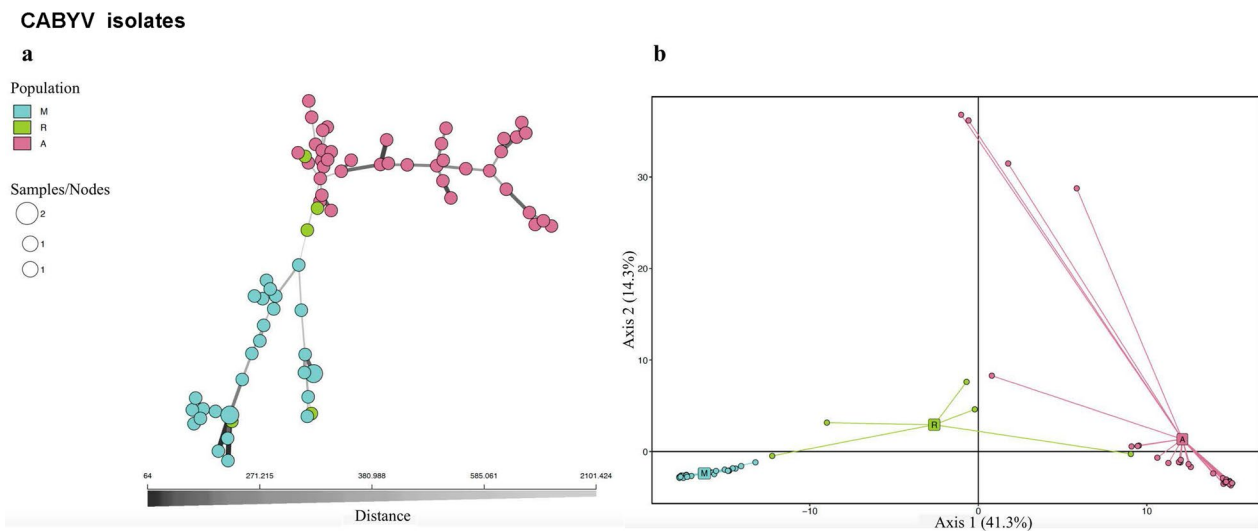


Fig. 2 **a** Minimum spanning network (MNS) of the CABYV genotype variants based on phylogenetic group. Each node represents a distinct genetic variant, with the size node indicating the number variant. The color of the node indicates the phylogenetic population in which the variant is present, where CABYV population is represented by phylogenetic group, Mediterranean (blue), Asian (pink), and recombinant (green). The thickness of the edges connecting the nodes indicates the genetic distance between the variants, with thicker edges representing greater relatedness. The scale for genetic distance is shown in the insets. **b** Principal component analysis (PCA) of CABYV phylogenetic group. CABYV population is represented by phylogenetic group, Mediterranean (blue), Asian (pink), and recombinant (green). PCA was performed in Rstudio with ade4 based on SNP difference among population Mediterranean (M), Asian (A), and Recombinant (R)

between the Spanish and French populations was low ($F_{ST} = 0.241$), indicating that these populations are more closely related. Notably, the Brazilian population showed higher gene flow with the other CABYV populations ($F_{ST} = 0.599\text{--}0.896$), indicating that the Brazilian isolates are quite different from the other CABYV isolates. Finally, we conducted a hierarchical AMOVA (analysis of molecular variation) to examine whether there was any correlation between location, host identity, and timing of the isolate collection. Our analysis consistently showed significant differentiation within different geographical locations ($\phi = 0.48$, $p = 0.0009$), representing 51.6% of the variation. However, there was no significant difference among sub-populations by timing ($\phi = 0$, $p = 0.98$). When the host identity was included as another hierarchical level in the model, this AMOVA analysis revealed significant differentiation between geographical locations and within host ($\phi = 0.55$, $p = 0.0009$), representing 65.2% of the variation. Together these results showed that despite the low genetic variability within the Mediterranean and Asian populations, phylodynamic differences can still be identified, likely as a result of different dispersal pathways of the CABYV populations and host adaptation.

WMV Nextstrain build

To explore the epidemic history and evolutionary dynamics of WMV, we analyzed the CP protein gene sequence of 472 isolates from 28 countries or regions, as well as the

full-length genome sequences of 129 WMV isolates from 13 countries or regions, using the Nextstrain approach. Our PCA and Nextstrain analyses found no geographical separation of WMV isolates based on their origins (Fig. 3a and Additional file 1: Figure S1b) and for more detailed information on the phylogeny and the geographic distribution, visit the following domains: https://nextstrain.org/community/PiR92/WMV_CP and <https://nextstrain.org/community/PiR92/WMV>). In particular, our Nextstrain analysis based on the CP gene revealed that the WMV epidemic may have originated in the USA during the sixteenth century (1284–1815), with a confidence level of 6%. But there are another three geographical areas, South Korea, Italy, and China, with close confidence levels where could also have been the origin (Fig. 3a). The WMV population is subdivided into 3 groups: classical (CL), emergent (EM), and one particular Chinese group formed for the isolates from the host *Ailanthus altissima*. The CL group may have originated in the eighteenth century (1861–1965) in France with a confidence level of 36%, with a divergence of 0.11 mutations from the initial population. These isolates were further clustered into 2 different CL (A and B) subgroups. The CL-A subgroup is composed of isolates from Argentina, Algeria, Chile, French Polynesia, India, Iran, Pakistan, South Korea, Spain, Turkey, USA, and Venezuela. CL-B consists of isolates from Argentina, Azerbaijan, Belgium, Bosnia, Herzegovina, Chile, Croatia, France, Egypt, Iran,

Table 2 Estimates of genetic differentiation and gene flow between CABYV populations grouped either by phylogenetic relationship or geographical origin

Comparasion	K _{ST} *	p value	Snn	p vlau	F _{ST}
Phylogroup					
Asian versus Mediterranean	0.113	0.000 ***	1.000	0.000 ***	0.588
Asian versus Recombinants	0.044	0.000 ***	0.980	0.000 ***	0.284
Mediterranean versus Recombinants	0.062	0.000 ***	0.932	0.000 ***	0.286
Geographical origin					
South Korea versus Spain	0.131	0.000 ***	0.981	0.000 ***	0.745
South Korea versus China	0.046	0.000 ***	0.917	0.002 **	0.203
South Korea versus Brazil	0.084	0.000 ***	1.000	0.000 ***	0.896
South Korea versus France	0.045	0.000 ***	1.000	0.013 *	0.735
South Korea versus India	0.061	0.002 **	1.000	0.010 *	0.823
South Korea versus Taiwan, China	0.053	0.002 **	0.937	0.162 ns	0.293
Spain versus China	0.062	0.000 ***	1.000	0.003 **	0.403
Spain versus Brazil	0.071	0.000 ***	1.000	0.300 ns	0.859
Spain versus France	0.014	0.208 ns	0.968	0.009 **	0.241
Spain versus India	0.050	0.001 **	1.000	0.009 **	0.794
Spain versus Taiwan, China	0.045	0.006 **	0.967	0.211 ns	0.317
China versus Brazil	0.094	0.013 *	1.000	0.014 *	0.666
China versus France	0.010	0.336 ns	1.000	0.065 ns	0.402
China versus India	0.030	0.101 ns	0.875	0.119 ns	0.480
China mainland versus Taiwan region	0.026	0.167 ns	0.500	0.728 ns	-0.033
Brazil versus France	0.253	0.093 ns	1.000	0.093 ns	0.848
Brazil versus India	0.301	0.089 ns	1.000	0.089 ns	0.887
Brazil versus Taiwan, China	0.296	0.092 ns	1.000	0.092 ns	0.599
France versus India	1.000	1.000 ns	1.000	0.349 ns	0.787
France versus Taiwan, China	1.000	1.000 ns	0.750	0.332 ns	0.317
India versus Taiwan, China	1.000	1.000 ns	0.750	0.321 ns	0.270

ns nonsignificant *0.01 < p < 0.05; **0.001 < p < 0.01. significantly rejects the null hypothesis that there is no genetic differentiation between two populations. K_{ST}* and Snn are test statistics of genetic differentiation. F_{ST}, coefficient of gene differentiation, which measures inter-population diversity

***p value < 0.001

Italy, Saudi Arabia, Syria, Slovakia, Spain, Turkey, and Ukraine (Fig. 3a). In contrast, the EM group could have originated in the nineteenth century (1858–1967) in China with a confidence level of 83%, with the population also subdivided into 4 (EM 1–4) subgroups. The EM-1 subgroup comprises by isolates from Argentina, France, Italy, South Korea, and Spain. EM-2 consists Argentina, China, France, Italy, Iran, South Korea, and USA. EM-3 includes isolates from Argentina, China, France, Italy, Iran, South Korea, and USA. EM-4 is formed by isolates from Belgium, China, France, Italy, and the Netherlands.

In turn, our Nextstrain analysis based on the full-length genome revealed that the WMV population may have originated in the seventeenth century (1722–1767) in China, with a confidence level of 57% (Fig. 3b, <https://nextstrain.org/community/PiR92/WMV>). In this context, the WMV population appeared to be divided into two groups: the Chinese group likely originated in the eighteenth century (1873–1897), and the group including

the CL and EM groups also originated in the eighteenth century (1834–1861). Thus, the CL isolates clustered into two subgroups (CL-A and B), with CL-A subgroup consisting of isolates from Chile, China, France, Iran, Pakistan, South Korea, Spain, Turkey, and Venezuela, while CL-B includes isolates from South Korea. In contrast, the EM isolates consisted of 4 subgroups (EM 1–4). EM-1 consists of isolates from France, South Korea, Spain, and USA. EM-2 comprise isolates from Argentina, China, France, and South Korea. EM-3, includes France, India, and South Korea. EM-4 is composed of isolates from France, Italy, Spain, and USA (Fig. 3b). Likewise, the same clades were identified in the host-based analysis. Furthermore, WMV phylodynamic population had an averaged substitution nucleotide rate of 4.13×10^{-4} per year.

Inferring population structure of WMV

To further examine the genetic differentiation within and between WMV populations, we analyzed the 129

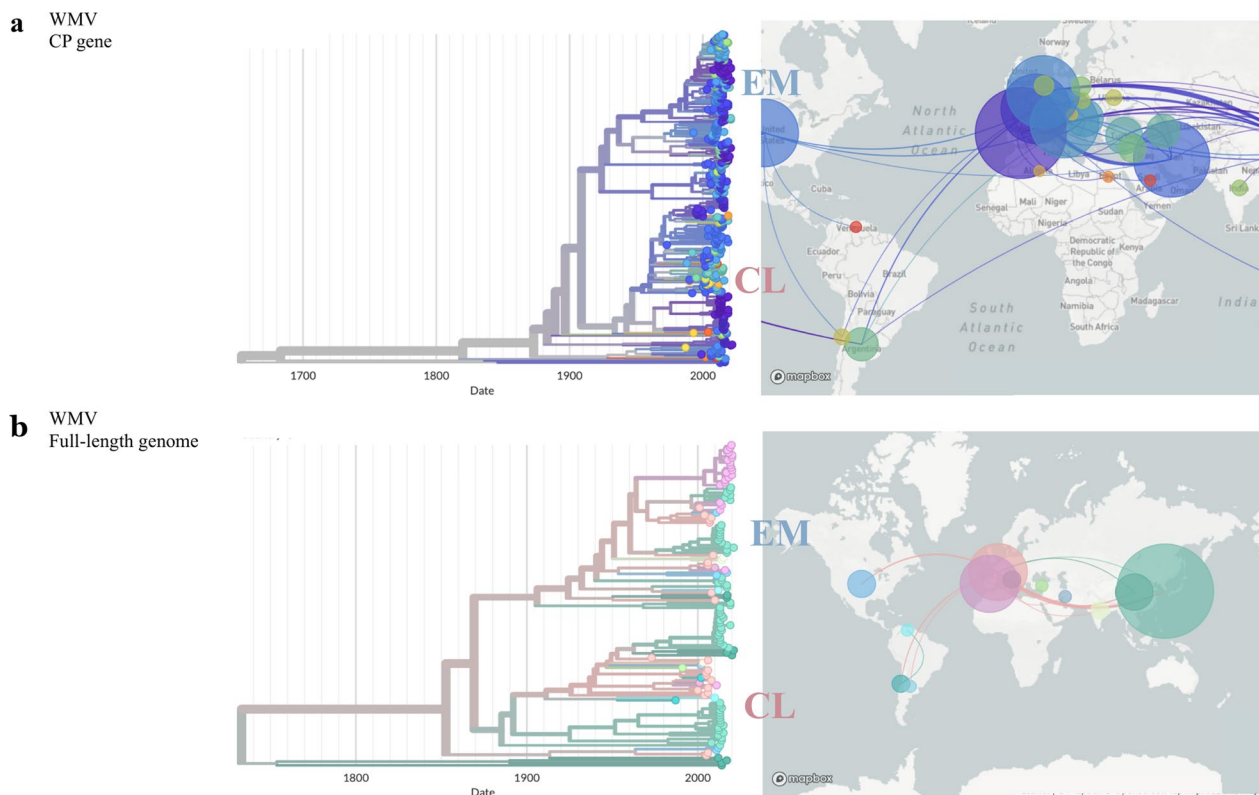


Fig. 3 Nextstrain builds of WMV from coat protein and complete genome. Both are open and hosted on the following domains <https://nextstrain.org/community/PiR92/WMV@main> and https://nextstrain.org/community/PiR92/WMV_CP@main. **a** Phylodynamic of WMV from 472 CP gene sequences collected from 1972 to 2020. **b** Phylodynamic of WMV from 129 complete genomes collected from 1972 to 2020. In both panels, a time tree is shown on the left, and the geographic distribution is displayed on the right. The compilation information is presented in three main panels: clustering of genomic diversity, geographic origin of samples and relative genetic diversity. With the associated metadata, users can color the tree nodes based on the region, country, and host plant. The internal node colors indicate the expected ancestral state of a given trait, and confidence in that state is conveyed by the colour saturation of the trait. Branch lengths are shown as a function of divergence or time. The viral genotypes represented in the tree are plotted on a map, which allows to observe the level of geographical spread and the most likely transmission events

full-length genome sequences considering their geographical regions. At first, our DAPC analysis revealed that there was no geographic or temporal discrimination (data not shown) among WMV populations. However, SNP-based PCA analysis showed a well-defined separation between CL and EM groups, which was confirmed by the MSN analysis, with no differentiation by geographic location or time (Fig. 4a, b). Our subsequent AMOVA analysis revealed a significant differentiation within phylogenetic groups ($\phi = 0.34$, $p = 0.0009$), representing 65.7% of the variation. In contrast, host and timing had no significant differentiation ($\phi = 0.26$, $p = 0.7$ and $\phi = -0.07$, $p = 0.9$, respectively). These results indicate that WMV populations exhibited different phylodynamics, possibly as a result of distinct pathways of virus dispersal.

Discussion

Phylodynamic studies play a crucial role in understanding the spatio-temporal information and spread patterns of pathogens (Grenfell et al. 2004; Pybus et al. 2012). The development of interactive platforms has made it easier to gather all available information in order to gain insights into the evolution and epidemiology of pathogens, which allows to share the epidemiological data with health organizations, scientists, and the public at real-time (Croucher and Didot 2015). In this sense, the Nextstrain platform offers a unique and actionable perspective of disease epidemics, utilizing a bioinformatic architecture associated to sequence metadata to represent the diversity and epidemiology of the pathogen globally, as well as to facilitate the scientific collaboration through data dissemination (Hadfield et al. 2018).

WMV isolates**a**

Population



Samples/Nodes

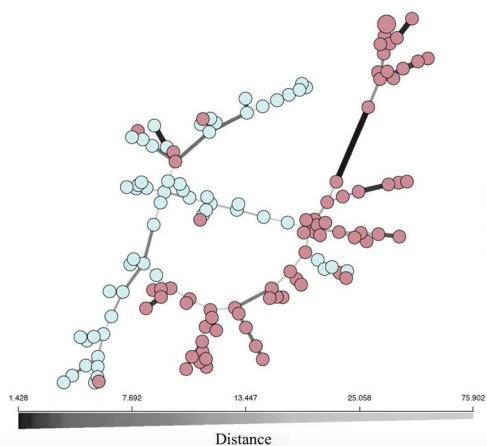
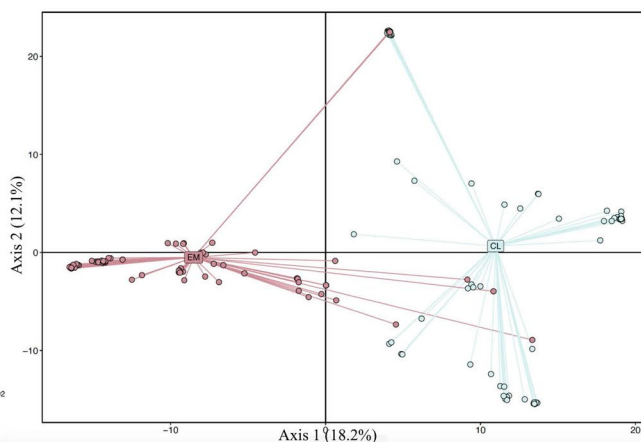
**b**

Fig. 4 **a** Minimum spanning network (MNS) of the WMV genotype variants based on phylogenetic origin. Each node represents a distinct genetic variant, with the size node indicating the number variant. The color of the node indicates the phylogenetic population in which the variant is present, where is represented classic group (light blue) and emergent group (light pink). The thickness of the edges connecting the nodes indicates the genetic distance between the variants, with thicker edges representing greater relatedness. The scale for genetic distance is shown in the insets. **b** PCA of WMV phylogenetic group. WMV population is represented by phylogenetic group, classic (light blue), and emergent (light pink). PCA was performed in Rstudio with ade4 based on SNP difference among classic (CL) and emergent (EM) populations

In this study, we inferred the phylodynamic of two RNA plant viruses, CABYV and WMV, that have important global epidemics on cucurbit crops. On one hand, the Nextstrain analysis showed that the CABYV epidemic may have originated in Southeast Asia in the fourteenth century, while the Mediterranean isolates probably originated in Spain during the seventeenth century. Additionally, the global phylogenetic reconstruction of the CABYV population confirmed a clear differentiation among isolates based on geographic location, revealing a close genetic relationship among the isolates within the Asian and Mediterranean group, with no gene flow between them. Although genetic diversity within Asian and Mediterranean populations was low, it appears that host adaptation may have contributed to the CABYV genetic differentiation. On the other hand, the WMV epidemic appears to have originated in the USA during the sixteenth century, with the classical group possibly originating in the eighteenth century, and the emergent in the nineteenth century. The WMV population had a polyphyletic origin, with extensive genetic diversity within sub-populations.

The phylodynamic analysis of CABYV shows that its geographical distribution is complex and may have been influenced by factors such as agricultural practices, environmental conditions, short- and long-term dispersal, including the aphid-vector ecology, and genetic diversity. CABYV is widely distributed in cucurbit crops, and recent outbreaks have been reported in Europe, Asia,

and Africa (Lecoq and Desbiez 2012; Radouane et al. 2021; Moya-Ruiz et al. 2023; Rabadán et al. 2023). The population structure of CABYV appears to be separated according to the geographical origin of the isolates, with three well-defined clades (Asian, Mediterranean, and Recombinant), which is consistent with previous studies (Kwak et al. 2018; Rabadán et al. 2023). The Nextstrain analysis of the CP gene sequence data suggests that the CABYV epidemic originated in the Asian continent dating back to the fourteenth century, grouping the Asian isolates, while the Mediterranean population probably originated in Spain in the seventeenth century. However, the full-length genome analysis contradicts this and suggests that the origin of the CABYV epidemic was likely in the Mediterranean and dated to the seventeenth century. Considering that CABYV was first identified in 1988 in cucumber and zucchini crops in France (Lecoq et al. 1992). This discrepancy between the CABYV origin could be attributed to the limited number of CABYV isolates for complete genome scans. This evidence highlights the potential limitations of using partial genomic data in understanding eco-evolutionary dynamics and the need of using complete genome data for phylogenetic traits. We found that the rate of CABYV molecular evolution was 1.23×10^{-4} substitutions/site/year, a value lying within the range reported for other ssRNA viruses (Duffy and Holmes 2008; Fargette et al. 2008).

Genetic diversity analysis showed that diversity in the Mediterranean CABYV population was lower than the

Asian and recombinant populations. This low genetic diversity may be due to the modularity or dynamics of cucurbit host species in the Mediterranean ecosystem, as exemplified for plant-virus infections in an agricultural landscape (Valverde et al. 2020). Furthermore, our analysis shows that there are codons under positive selection in the ORF2, which has been also evidenced from similar studies at regional level (Rabadán et al. 2021, 2023). However, despite this, the CABYV population is under purifying selection, suggesting that most mutations may be deleterious. The statistical tests of neutrality (Tajima's D and Fu and Li's F^*) supported the pattern of genetic structure observed in the population. The non-significant negative values observed in the Mediterranean and Asian populations indicate a recent population expansion. However, the Fu's FS statistic showed non-significant negative values for the Asian population, indicating an excess of rare mutations, while the values for the Mediterranean and recombinant populations were positive, suggesting a recent bottleneck. These increases or decreases in population size may be associated with demographic events that may be affecting demographic variation through genetic drift, mutation, or recombination. A similar aspect has been described for the PVY population in China and Japan, where despite not being too far apart geographically, the virus population has been observed to have a spatial structure (Gao et al. 2017). This difference in the genetic structure of the two populations seems to be attributed to higher potato production in China compared to Japan in recent years leading to an expansion of the population in China while the population in Japan has remained stable. Additionally, the DAPC analysis confirmed the subdivision of the CABYV population based on the geographical origin of the isolates. The genetic variability within Asian and Mediterranean CABYV isolates was low, while both groups were genetically differentiated. The F_{ST} values (0.6) also confirmed the genetic differentiation between the Mediterranean and Asian populations, with no gene flow between them.

We further show that population structure of CABYV is largely consistent with the geographical location. However, we also found that a significant proportion of variation was detected between geographical populations and within hosts. These results suggest that despite a substantial proportion of the CABYV isolates are coming from melon hosts, there could be a differentiation in CABYV attributed to the host adaptation. This contrasts with our recent study where there was no evidence of CABYV genetic differentiation by the host (melon and zucchini crops), and however, there was a moderate differentiation among the type of infection (i.e. whether isolates coming from single or mixed

infections) (Rabadán et al. 2023). It is likely that the effect of limiting the analysis to melon isolates, in addition to the uncertainty whether the CABYV isolate was present or absent with another viral infection in the same host plant, could lead to an underestimation of the significant factors contributing to this particular genetic differentiation. In this sense, a recent study has reported that CABYV population was structured by the host (*Ecballium elaterium* and melon), with no evidence of migration fluxes between both host populations (Maachi et al. 2022). Another example for host adaptation could be the influence of ecological barriers or low virus transmission due to host species diversity, which often lead to patterns of isolation and spatial population structure (García-Arenal et al. 2001). Overall, our study represents the first comprehensive CABYV phylodynamic analysis on a large scale, showing how the CABYV population is globally structured, with limited gene flow migration rates. This may suggest that agro-ecological factors have been contributing to the CABYV population differentiation.

Phylodynamic analysis of WMV shows that the population structure of WMV appears to be separated according to the genetic diversity of the isolates, with two well-defined clades (classical and emergent), as previously described (Desbiez et al. 2009; Juarez et al. 2013; Verma et al. 2020). The WMV epidemic was first described in Israel in 1963 (Cohen and Nitzany 1963). Nextstrain analysis showed that the WMV population is polyphyletic, spreading into more than one phylogenetic group, suggesting that WMV isolates have been exposed in other geographical regions, evolving by genetic drift (García-Andrés et al. 2007). Nextstrain analysis based on CP sequences suggests that the WMV epidemic may have originated in the sixteenth century in USA. The classical group is likely to have originated in the eighteenth century in France, while the emergent group may have originated in the nineteenth century in China. However, analysis based on full-length genome analysis suggests that the WMV epidemic may have originated in the seventeenth century in China. A CP-based study has described the association between WMV isolates (Hajizadeh et al. 2017), where Chinese, French, and US populations exhibited higher gene flow, while the Spanish and Iranian populations were more differentiated from the Chinese, French, and South Korean populations. Otherwise, Italian isolates were found to be closely related to isolates from France, Iran, and Turkey, while emergent isolates showed polyphyletic relatedness between Italian, South Korean, Chinese, and Argentinian isolates (Bertin et al. 2020). A plausible explanation for these polyphyletic relationships is that the WMV populations have spread rapidly from different origins, which could have

contributed to the genetic variability of the WMV population (Desbiez and Lecoq 2008; Peláez et al. 2021).

Additionally, the MSN analysis confirmed the subdivision between CL and EM isolates. It should be noted that some of those isolates showed genetic relatedness, which could be explained by increased gene flow between both groups (Desbiez and Lecoq 2008). Moreover, a study based on the Nib/CP sequence data from Italian isolates found a shift from CL to EM populations likely due to multiple introductions of EM isolates in different regions of Italy (Bertin et al. 2020). Otherwise, a recent molecular epidemiological study of WMV population in Spain showed that although most of collected isolates belongs to the EM group, there are still coexisting CL isolates without its displacement (Rabadán et al. 2023). In this regard, it should be noted that a recent work found that the genetic diversity of WMV population was structured by host species and habitat type, suggesting that crops had higher number of haplotypes than anthropic habitats, although both effects were not mutually exclusive, with other ecological factors contributing to the variability of WMV (Peláez et al. 2021). Nevertheless, we further found that the WMV population structure is largely consistent with phylogenetic groups, with no significant proportion of molecular variation either by host or time. As aforementioned, it is likely that other agro-ecological factors may be contributing to WMV diversification. Thus, further studies on alternative hosts, including wild populations, are needed to understand what factors are influencing WMV population dynamics. Our Nextstrain analysis indicates that the WMV population spread rapidly across different geographical locations, likely through infected plant materials. Additionally, the EM phylogroup may have originated in Asia and subsequently migrated to the European continent. This migration may have led to recombination with CL isolates, resulting in a population differentiation (Desbiez and Lecoq 2008; Desbiez et al. 2009). Thus, the migration of WMV isolates from different countries or regions may have influenced the structure and epidemiology of the WMV population.

It is noteworthy that population genomic surveillance for viral diseases can help to examine the evolutionary dynamics of viral populations across spatial and temporal scales, which stands as a crucial factor in managing viral diseases (Acosta-Leal et al. 2011). Within breeding programs for the development of resistant cultivars, the characterization of the population genetic structure is fundamental to use the appropriate virus genotypes that allow to ensure the durability of the resistance (García-Arenal and McDonald 2003; Gómez et al. 2009). As an example, the appearance of virulent variants in the viral populations may allow the breaking of cultivar resistance

usually by the occurrence of mutations in the so-called avirulence gene encoded by the viral genome (Moury et al. 2004; Chiba et al. 2008; Fraile et al. 2011). Additionally, the evolutionary study of potyviruses is a well-described example that particular migration pathways may have resulted from viral life-history traits (Fuentes et al. 2019; Gibbs et al. 2020), and provide valuable predictions to assess in which extent factors, such as human activity, could impact on the effectiveness of implemented control measures. Therefore, this comprehensive phylodynamic characterization for CABYV and WMV within an open phylogenetic reporting platform offers potential for the identification and track of novel genotypes and also valuable insights for the development of effective control measures.

Conclusions

Nextstrain platform offers a unique and actionable perspective on disease epidemics by facilitating the scientific collaboration through data dissemination. Our Nextstrain CABYV analysis suggests that the epidemic originated in Southeast Asia during the fourteenth century, emerging subsequently in the Mediterranean basin. The CABYV geographical distribution indicates that its phylodynamics are complex and may be influenced by several factors, including the location and host identity. Regarding WMV, the analysis suggests an epidemic origin in USA during sixteenth century. The WMV population appears to be polyphyletic, with recombination events among isolates from different geographical locations, and the gene flow playing an important role within populations. It is worthwhile to note that the possibility of new viral outbreaks remains as a notable consideration, and this study emphasizing the importance of extensive and continued viral population genomic surveillance programs to gain insights into the population structure of viruses, which can elucidate their evolutionary histories in order to determine the outbreak origin and address effective preventive and control measures.

Methods

CABYV and WMV nucleotide sequences

A total of 74 full-length genome and 393 CP gene sequences of CABYV were retrieved from the GenBank database, originating from 10 to 32 different countries or regions, respectively, and spanning a period of 27 years (1994–2021). For WMV, 129 full-length genome and 472 CP sequences were obtained from 13 to 28 countries or regions, respectively, over a period of 48 years (1972–2020). All sequence data were organized in a metadata table, which includes information on strain, virus, date, accession, host, region, country, database (db), segment, and URL.

Nextstrain implementation

The phylodynamic reconstruction of CABYV and WMV was carried out by using Nextstrain command, including the metadata input for each virus. Multiple sequence alignment of CABYV and WMV sequences was performed by MAFFT (Kato and Standley 2013) and then clustered by RAxML (Stamatakis 2014). The initial tree was further refined with metadata, and a temporal tree was constructed using TreeTime (Sagulenko et al. 2018), with coalescent scalar time optimisation. Internal nodes were assigned to their marginally most likely dates, and confidence intervals for node dates were estimated. Amino acid changes in the CABYV and WMV populations were determined by using the NCBI (National Center for Biotechnology Information) accession numbers MW051362.1 for CABYV and MW147356.1 for WMV, which both were characterized in our laboratory (De Moya-Ruiz et al. 2021; Rabadán et al. 2021). The resulting Augur output was exported and visualised in auspice. The compiled CABYV and WMV Nextstrain data are deposited on GitHub (<https://github.com>).

Resource availability and interoperability

The Nextstrain builds for full-length genomes and CP gene sequences of CABYV and WMV are available in the following domains:

<https://nextstrain.org/community/PiR92/CABYV@main>

https://nextstrain.org/community/PiR92/CABYV_CP@main

<https://nextstrain.org/community/PiR92/WMV@main>

https://nextstrain.org/community/PiR92/WMV_CP@main

These web domains provide users with the ability to colour tree nodes based on the host plant, year, or region. Internal node colors indicate the expected ancestral state of a given trait, and the saturation of the internal node represents the confidence in that state. The cladogram can be displayed in various styles, such as rectangular, radial, and unrooted. The tree branch lengths can be displayed as a function of divergence or time. Using the information from the construction, Nextstrain determines the most likely transmission events, which can be viewed as an animation on the website. Genotypes represented on the tree are plotted on a map, and users can set different levels of geographical resolution, allowing for simultaneous querying of phylogenetic and geographic relationships, along with additional relevant metadata. Nextstrain estimates the genetic divergence between

isolates by measuring the number of changes (mutations) in the nucleotide sequences, relative to the root of the tree, as well as indicating the nucleotide substitution rate per site per year. Furthermore, the tree phylogenetic analysis can be conducted based on the host, region, or country by using the “Breakdown tree by” tab. The Augur input, with all metadata and scripts, is available on GitHub via <https://github.com/PiR92>. Users can download metadata and tree files from the website, as well as create screenshots of their visualisations.

Analysis of CABYV and WMV population structure

Population structure of CABYV was examined using model-free DAPC, which is implanted in the R package adegenet (Jombart et al. 2010). DAPC maximizes the variation between groups while minimizing variations within groups. DAPC first performs a PCA on the data and then assigns individuals to clusters using discriminant analysis (DA). This approach overcomes the drawbacks of direct application of DA by transforming the data so that the input variables are uncorrelated and their number is less than the number of individuals analysed (Wang et al. 2017). To visualize genetic distances among variants, a minimum-spanning network (MSN) was generated using the function poppr.msn implemented in the R poppr package (Kamvar et al. 2014). Additionally, an AMOVA was carried out in R with Poppr package to estimate how genetic variation is distributed within- and between sub-populations that are partitioned by host, geographical location, and time. This multivariate analysis was performed using the Euclidean distance between populations with multiple pairwise tests and a significance level (p -value) set at <0.05 . The population subdivision based on the phylogenetic group and geographical location was represented using a principal component analysis (PCA) implemented in the R package ade4 (Dray and Dufour 2007).

Genetic variability of CABYV populations

CABYV populations were separated according to phylogroup (Asian, Mediterranean, and Recombinant) or geographic population (China mainland, South Korea, India, Spain, France, Brazil, and Taiwan region of China). The strength of the natural selection was estimated by the ratio of the number of non-synonymous substitutions per non-synonymous site (dN) and the number of synonymous substitutions per synonymous site (dS) using the Pamilo-Banchi-Li method in MEGA7 (Pamilo and Bianchi 1993). The ratio $dN/dS \approx 1$ indicates neutral evolution, < 1 negative or purifying selection, and > 1 positive or adaptive selection. Examination the neutral selection hypothesis operating on the complete genome between phylogroups by Tajima's D and Fu and Li's F statistical

test were done using DnaSP6 (Rozas et al. 2017). Tajima's D test identifies evolutionary events by comparing the estimated number of segregating sites with the mean pairwise difference among sequences (Nei and Tajima 1981). If Tajima's D is positive it may indicate balancing or bottleneck selection in the population or founder events, whereas if, Tajima's D is negative it indicates population increase or positive selection. Fu and Li's F^* statistics are also neutrality statistics distinguishing between old or recent mutations occurring in the population, comparison of the number of derived singleton mutations and the total number of derived nucleotide variants (the asterisk indicates 'without an outgroup') (Fu and Li 1993). The examination of demographic expansion of CABYV population were calculated with Fu's FS. Fu's FS is based on Ewens' sampling distribution, considering the number of different haplotypes in the sample and usually displays a negative value (Fu 1997) in expanding populations.

Genetic differentiation of CABYV population was inferred by DnaSP6 (Rozas et al. 2017) software, with the estimation of the number of haplotypes (H), number of polymorphic (segregating) sites (S), total number of mutations η (Eta), average number of nucleotide differences between sequences (k), genetic diversity (p), total number of synonymous sites analysed (SS), and total number of non-synonymous sites analysed (NS). H values can range from 0.5 to 5, where, <2 low diversity and >3 high diversity. Population differentiation was calculated as described in Rabadán et al. (2023) K_{ST}^* , S_{nn} , and F_{ST} between phylogroups and geographic population was also performed by using DnaSP6. In particular, under the null hypothesis (no genetic differentiation), K_{ST}^* is expected to be close to zero. The hypothesis of deviation from null population differentiation was tested by 1000 permutations of the original data. S_{nn} is defined as the nearest neighbours of sequences, and its value can range between 1, when population are distinctly differentiated, to 0.5 in the case of panmixia (Hudson 2000).

Abbreviations

CL	Classical isolates
EM	Emergent isolates
CP	Coat protein
PCA	Principal component analysis
DAPC	Discriminant principal component analysis
MSN	Minimum network analysis
AMOVA	Analysis of molecular variation
π	Nucleotide diversity
F_{ST}	Gene flow index diversity
K_{ST}	Nucleotide sequence based statistic
D	Tajima's D
Hd	Haplotype diversity
dS	Number of synonymous substitutions per synonymous site
dN	Number of non-synonymous substitutions per non-synonymous site
S_{nn}	Statistics of genetic differentiation

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42483-023-00207-8>.

Additional file 1: Figure S1. Principal component analysis (PCA) of CABYV and WMV populations. **a** The CABYV population is represented by geographical location, countries or regions: SP (Spain), FR (France), IND (Indonesia), IN (India), TW (Taiwan of China), PG (Papua New Guinea), CH (China), SK (South Korea), JP (Japan), and USA (United States of America). PCAs were performed in Rstudio with ade4 based on SNP difference among Asian, Mediterranean, and Recombinant groups. **b** The WMV population is also represented by the following countries or regions: SP, ARG (Argentina), IND, USA, SK, FR, IT (Italy), CH, VZ (Venezuela), CHIL (Chile), PK (Pakistan), IR (Iran), and TK (Turkey). PCAs were performed in Rstudio with ade4 based on SNP difference among classic and emergent WMV groups. **Figure S2. a** Discriminant principal components analysis (DAPC) cluster of CABYV isolates based on their geographical origin. Eigenvalues retained for principal component and discriminant functions in the analysis are displayed in inset. DAPC is represented the CABYV population by geographical location, countries or regions: SP (Spain), FR (France), IND (Indonesia), IN (India), TW (Taiwan of China), PG (Papua New Guinea), CH (China), SK (South Korea), JP (Japan), and USA (United States of America). **b** Minimum spanning network (MSN) of the CABYV genotype variants based on their geographical origin. Each node represents a distinct genetic variant, with the size of the node indicating the number variant. The color of the node indicates the geographic population in which the variant is prevalent. The thickness of the edges connecting the nodes indicates the genetic distance between the variants, with thicker edges representing greater relatedness. The scale for genetic distance is shown in the insets. The CABYV population is represented by geographical location, countries, or regions: SP (Spain), FR (France), IND (Indonesia), IN (India), TW (Taiwan of China), PG (Papua New Guinea), CH (China), SK (South Korea), JP (Japan), and USA (United States of America).

Additional file 2: Table S1. List of codons within the ORFs of the CABYV genome that have been inferred to be under positive selection (p -value < 0.05), along with the corresponding amino acid substitution.

Author contributions

MPR and PG designed the study. MPR performed the study. MPR analyzed the data. MPR and PG wrote the manuscript. All authors read and approved the final manuscript.

Funding

We thank that this work was supported by the Spanish research grants (AGL2017-89550-R) from MICINN and EU FEDER funds, and also that M.P.R. was supported by funding of the Ministry of Economy, Industry and Competitiveness (MINECO, Spain) within a PhD programme grant (PRE2018-083915).

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 10 April 2023 Accepted: 9 October 2023

Published online: 14 November 2023

References

- Abeykoon AMSK, Basnayake BMVS, Salim N. First report of Cucurbit aphid-borne yellow virus infecting bitter melon (*Momordica charantia*) and spiny gourd (*Momordica dioica*) in Sri Lanka. *New Dis Rep.* 2018;38:22. <https://doi.org/10.5197/j.2044-0588.2018.038.022>.
- Acosta-Leal R, Duffy S, Xiong Z, et al. Advances in plant virus evolution: translating evolutionary insights into better disease management. *Phytopathology.* 2011. <https://doi.org/10.1094/PHYTO-01-11-0017>.
- Ahsan M, Ashfaq M, Mukhtar T, et al. First report of cucurbit aphid borne yellows virus (CABYV) infecting melon in Pakistan. *J Plant Pathol.* 2020;102:563–4. <https://doi.org/10.1007/s42161-019-00450-z>.
- Alexander HM, Mauck KE, Whitfield AE, et al. Plant-virus interactions and the agro-ecological interface. *Eur J Plant Pathol.* 2014;138:529–47. <https://doi.org/10.1007/s10658-013-0317-1>.
- Bananej K, Vahdat A, Predajbà L, Glasa M. Molecular characterization of geographically different cucurbit aphid-borne yellows virus isolates. *Acta Virol.* 2009;53:61–4. https://doi.org/10.4149/av_2009_01_61.
- Benítez-Galeano M, Castells M, Colina R. The evolutionary history and spatiotemporal dynamics of the NC lineage of citrus tristeza virus. *Viruses.* 2017;9:272. <https://doi.org/10.3390/v9100272>.
- Bertin S, Manglli A, McLeish M, Tomassoli L. Genetic variability of watermelon mosaic virus isolates infecting cucurbit crops in Italy. *Arch Virol.* 2020;165:937–46. <https://doi.org/10.1007/s00705-020-04584-9>.
- Chiba S, Miyaniishi M, Andika IB, et al. Identification of amino acids of the beet necrotic yellow vein virus p25 protein required for induction of the resistance response in leaves of *Beta vulgaris* plants. *J Gen Virol.* 2008;89:1314–23. <https://doi.org/10.1099/vir.0.83624-0>.
- Cohen S, Nitzany FE. Identity of viruses affecting cucurbits in Israel. *Phytopathology.* 1963;53:193–6.
- Croucher NJ, Didelot X. The application of genomics to tracing bacterial pathogen transmission. *Curr Opin Microbiol.* 2015;23:62–7. <https://doi.org/10.1016/j.mib.2014.11.004>.
- De Moya-Ruiz C, Rabadán P, Juárez M, Gómez P. Assessment of the current status of potyviruses in watermelon and pumpkin crops in Spain: epidemiological impact of cultivated plants and mixed infections. *Plants.* 2021;10:1–14. <https://doi.org/10.3390/plants10010138>.
- Desbiez C, Lecoq. Evidence for multiple intraspecific recombinants in natural populations of watermelon mosaic virus (WMV, *Potyvirus*). *Arch Virol.* 2008;153:1749–54. <https://doi.org/10.1007/s00705-008-0170-2>.
- Desbiez C, Joannon B, Wipf-Scheibel C, et al. Emergence of new strains of watermelon mosaic virus in South-eastern France: evidence for limited spread but rapid local population shift. *Virus Res.* 2009;141:201–8. <https://doi.org/10.1016/j.virusres.2008.08.018>.
- Desbiez C, Wipf-Scheibel C, Millot P, et al. Distribution and evolution of the major viruses infecting cucurbitaceous and solanaceous crops in the French Mediterranean area. *Virus Res.* 2020;286:198042. <https://doi.org/10.1016/j.virusres.2020.198042>.
- Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw.* 2007;22:1–20. <https://doi.org/10.18637/jss.v022.i04>.
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:1–8. <https://doi.org/10.1186/1471-2148-7-214>.
- Duffy S, Holmes EC. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA Begomovirus *Tomato Yellow Leaf Curl Virus*. *J Virol.* 2008;82:957–65. <https://doi.org/10.1128/JVI.01929-07>.
- Elena SF, Fraile A, García-Arenal F. Chapter three—evolution and emergence of plant viruses. In: Maramorosch K, Murphy FA, editors *Advances in virus research*. New York: Academic Press. 2014. p 161–91. <https://doi.org/10.1016/B978-0-12-800098-4.00003-9>.
- Fargette D, Pinel-Galzi A, Sérémé D, et al. Diversification of *Rice Yellow Mottle Virus* and related viruses spans the history of agriculture from the neolithic to the present. *PLoS Pathog.* 2008;4:1–8. <https://doi.org/10.1371/journal.ppat.1000125>.
- Fraile A, Pagán I, Anastasio G, et al. Rapid genetic diversification and high fitness penalties associated with pathogenicity evolution in a plant virus. *Mol Biol Evol.* 2011;28:1425–37. <https://doi.org/10.1093/molbev/msq327>.
- Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics.* 1997;147:915–25. <https://doi.org/10.1093/genetics/147.2.915>.
- Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics.* 1993;133:693–709. <https://doi.org/10.1093/genetics/133.3.693>.
- Fuentes S, Jones RAC, Matsuoka H, et al. Potato virus Y; the Andean connection. *Virus Evol.* 2019. <https://doi.org/10.1093/ve/vez037>.
- Gao F, Zou W, Xie L, Zhan J. Adaptive evolution and demographic history contribute to the divergent population genetic structure of Potato virus Y between China and Japan. *Evol Appl.* 2017;10:379–90. <https://doi.org/10.1111/eva.12459>.
- García-Andrés S, Accotto GP, Navas-Castillo J, Moriones E. Founder effect, plant host, and recombination shape the emergent population of begomoviruses that cause the tomato yellow leaf curl Disease in the Mediterranean basin. *Virology.* 2007;359:302–12. <https://doi.org/10.1016/j.virol.2006.09.030>.
- García-Arenal F, McDonald BA. An analysis of the durability of resistance to plant viruses. *Phytopathology.* 2003;93:941–52. <https://doi.org/10.1094/PHYTO.2003.93.8.941>.
- García-Arenal F, Fraile A, Malpica JM. Variability and genetic structure of plant virus populations. *Annu Rev Phytopathol.* 2001;39:157–86. <https://doi.org/10.1146/annurev.phyto.39.1.157>.
- García-Ruiz H, Holste NM, LaTourrette K. Poleroviruses (Luteoviridae). In: Bamford DH, Zuckerman M, editors *Encyclopedia of Virology*, Volume 1–5, 1st ed. Academic Press: Oxford, UK; 2020. p. 594–602. <https://doi.org/10.1016/B978-0-12-809633-8.21343-5>.
- Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet.* 2018;19:9–20. <https://doi.org/10.1038/nrg.2017.88>.
- Gibbs A, Ohshima K. Potyviruses and the digital revolution. *Annu Rev Phytopathol.* 2010;48:205–23. <https://doi.org/10.1146/annurev-phyto-073009-114404>.
- Gibbs AJ, Hajizadeh M, Ohshima K, Jones RAC. The potyviruses: an evolutionary synthesis is emerging. *Viruses.* 2020;12:2. <https://doi.org/10.3390/v12020132>.
- Gómez P, Rodríguez-Hernández AM, Moury B, Aranda M. Genetic resistance for the sustainable control of plant virus diseases: breeding, mechanisms and durability. *Eur J Plant Pathol.* 2009;125:1–22. <https://doi.org/10.1007/s10658-009-9468-5>.
- González R, Butković A, Elena SF. Role of host genetic diversity for susceptibility-to-infection in the evolution of virulence of a plant virus. *Virus Evol.* 2019;5:1–12. <https://doi.org/10.1093/ve/vez024>.
- Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science.* 2004;303:327–32. <https://doi.org/10.1126/science.1090727>.
- Hadfield J, Megill C, Bell SM, NextStrain, et al. Real-time tracking of pathogen evolution. *Bioinformatics.* 2018;34:4121–23. <https://doi.org/10.1093/bioinformatics/bty407>.
- Hajizadeh M, Bahrapour H, Abdollahzadeh J. Genetic diversity and population structure of watermelon mosaic virus. *J Plant Dis Prot.* 2017;124:601–10. <https://doi.org/10.1007/s41348-017-0114-8>.
- Hajizadeh M, Gibbs AJ, Amirnia F, Glasa M. The global phylogeny of Plum pox virus is emerging. *J Gen Virol.* 2019;100:1457–68. <https://doi.org/10.1099/JGV.0.001308>.
- Hudson RR. A new statistic for detecting genetic differentiation. *Genetics.* 2000;155:2011–14. <https://doi.org/10.1093/genetics/155.4.2011>.
- Jeger MJ. The epidemiology of plant virus disease: towards a new synthesis. *Plants.* 2020;9:1–50. <https://doi.org/10.3390/plants9121768>.
- Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* 2010;11:94. <https://doi.org/10.1186/1471-2156-11-94>.
- Jones RAC. Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res.* 2009;141:113–30. <https://doi.org/10.1016/j.virusres.2008.07.028>.
- Jones RAC. Global plant virus disease pandemics and epidemics. *Plants.* 2021;10:1–41. <https://doi.org/10.3390/plants10020233>.
- Jones RAC, Naidu RA. Global dimensions of plant virus diseases: current status and future perspectives. *Annu Rev Virol.* 2019;6:387–409. <https://doi.org/10.1146/annurev-virology-092818-015606>.
- Juarez M, Legua P, Mengual CM, et al. Relative incidence, spatial distribution and genetic diversity of cucurbit viruses in eastern Spain. *Ann Appl Biol.* 2013;162:362–70. <https://doi.org/10.1111/aab.12029>.

- Kamvar ZN, Tabima JF, Grunwald NJ, Poppr. An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*. 2014;2014:1–14. <https://doi.org/10.7717/peerj.281>.
- Kassem Ma, Juárez M, Gómez P, et al. Genetic diversity and potential vectors and reservoirs of Cucurbit aphid-borne yellows virus in Southeastern Spain. *Phytopathology*. 2013;103:1188–97. <https://doi.org/10.1094/PHYTO-11-12-0280-R>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80. <https://doi.org/10.1093/molbev/mst010>.
- Kwak HR, Lee HJ, Kim EA et al. Complete genome sequences and evolutionary analysis of Cucurbit aphid-borne yellows virus isolates from Melon in Korea. *Plant Pathol J*. 2018;34:532–43. <https://doi.org/10.5423/PPJ.OA.03.2018.0049>.
- Lecoq H, Desbiez C. Viruses of Cucurbit Crops in the Mediterranean Region. An Ever-Changing Picture. In: *Advances in Virus Research* (2012). pp 67–126. <https://doi.org/10.1016/B978-0-12-394314-9.00003-8>.
- Lecoq H, Bourdin D, Wipf-Scheibel C, et al. A new yellowing Disease of cucurbits caused by a luteovirus, cucurbit aphid-borne yellows virus. *Plant Pathol*. 1992;41:749–61. <https://doi.org/10.1111/j.1365-3059.1992.tb02559.x>.
- Maachi A, Donaire L, Hernando Y, Aranda MA. Genetic differentiation and Migration fluxes of viruses from melon crops and crop edge weeds. *J Virol*. 2022;96:16. <https://doi.org/10.1128/jvi.00421-22>.
- Mabele AS, Muyekho FN, Were HK. First report of cucurbit aphid-borne yellows virus infecting groundnut in Kenya. *Plant Health Prog*. 2022. <https://doi.org/10.1094/PHP-08-22-0071-RS>.
- Mehle N, Kutnjak D, Jakoš N, et al. First report of cucurbit aphid-borne yellows virus in cucurbita pepo and cucurbita maxima in Slovenia. *Plant Dis*. 2020;104:599. <https://doi.org/10.1094/PDIS-07-19-1524-PDN>.
- Menzel W, Maeritz U, Seigner L. First report of cucurbit aphid-borne yellows virus infecting cucurbits in Germany. *New Disease Reports*. 2020;41:5197. <https://doi.org/10.5197/j.2044-0588.2020.041.001>.
- Montes N, Pagán I. Challenges and opportunities for plant viruses under a climate change scenario. Volume 114, 1st ed. New York: Elsevier Inc; 2022. p. 1–6. <https://doi.org/10.1016/bs.aivir.2022.08.001>.
- Moury B, Morel C, Johansen E, et al. Mutations in *Potato virus Y* genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Mol Plant Micro Int*. 2004;17:322–9. <https://doi.org/10.1094/MPMI.2004.17.3.322>.
- Moya-Ruiz C, De Gómez P, Juárez M. Occurrence, distribution, and management of aphid-transmitted viruses in cucurbits in Spain. *Pathogens*. 2023;12:422. <https://doi.org/10.3390/pathogens12030422>.
- Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. *Genetics*. 1981;97:145–63. <https://doi.org/10.1093/genetics/97.1.145>.
- Pamilo P, Bianchi NO. Evolution of the Zfx and Zfy genes: Rates and interdependence between the genes. *Mol Biol Evol*. 1993;10:271–81. <https://doi.org/10.1093/oxfordjournals.molbev.a040003>.
- Peáléaz A, McLeish MJ, Paswan RR, et al. Ecological fitting is the forerunner to diversification in a plant virus with broad host range. *J Evol Biol*. 2021;1917–31. <https://doi.org/10.1111/jeb.13672>.
- Prosperi M, Rife B, Marini S, Salemi M. Transmission cluster characteristics of global, regional, and lineage-specific SARS-CoV-2 phylogenies. In: *Proceedings—2022 IEEE international conference on bioinformatics and biomedicine, BIBM*. 2022. p. 2940–44. <https://doi.org/10.1109/BIBM55620.2022.9995364>.
- Pybus OG, Suchard MA, Lemey P, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci USA*. 2012;109:15066–71. <https://doi.org/10.1073/pnas.1206598109>.
- Rabadán MP, Juárez M, De Moya-Ruiz C, Gómez P. Aphid-borne viruses infecting cultivated watermelon and squash in Spain: characterization of a variant of cucurbit aphid-borne yellows virus (CABYV). *Plant Pathol*. 2021;70:1476–85. <https://doi.org/10.1111/ppa.13390>.
- Rabadán M, Juárez M, Gómez P. Long-term monitoring of aphid-transmitted viruses in melon and zucchini crops: genetic diversity and population structure of cucurbit aphid-borne yellows virus and watermelon mosaic virus. *Phytopathology*. 2023. <https://doi.org/10.1094/PHYTO-10-22-0394-V>.
- Radeva-Ivanova V, Pasev G, Lyall R, et al. First report of cucurbit aphid-borne yellows virus causing yellowing disease on pumpkin (*Cucurbita pepo*) and cucumber (*Cucumis sativus*) in Bulgaria. *Plant Dis*. 2022;106:2538. <https://doi.org/10.1094/PDIS-09-21-2030-PDN>.
- Radouane N, Ezrari S, Belabess Z, et al. Viruses of cucurbit crops: current status in the Mediterranean region. *Phytopathol Mediterr*. 2021;60:493–519. <https://doi.org/10.36253/phyto-12340>.
- Ristaino JB, Anderson PK, Bebber DP, et al. The persistent threat of emerging plant Disease pandemics to global food security. *Proc Natl Acad Sci USA*. 2021;118:1–9. <https://doi.org/10.1073/pnas.2022239118>.
- Rodelo-Urrego M, Pagán I, González-Jara P, et al. Landscape heterogeneity shapes host-parasite interactions and results in apparent plant–virus codivergence. *Mol Ecol*. 2013;22:2325–40. <https://doi.org/10.1111/mec.12232>.
- Roossinck MJ, García-Arenal F. Ecosystem simplification, biodiversity loss and plant virus emergence. *Curr Opin Virol*. 2015;10:56–62. <https://doi.org/10.1016/j.coviro.2015.01.005>.
- Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol*. 2017;34:3299–302. <https://doi.org/10.1093/molbev/msx248>.
- Sagulenko P, Puller V, Neher RA, TreeTime. Maximum-likelihood phylodynamic analysis. *Virus Evol*. 2018;4:1–9. <https://doi.org/10.1093/ve/vex042>.
- Shang QX, Xiang HY, Han CG, et al. Distribution and molecular diversity of three cucurbit-infecting poleroviruses in China. *Virus Res*. 2009;145:341–6. <https://doi.org/10.1016/j.virusres.2009.07.017>.
- Stamatakis A. RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–13. <https://doi.org/10.1093/bioinformatics/btu033>.
- Theys K, Lemey P, Vandamme AM, Baele G. Advances in visualization tools for phylogenomic and phylodynamic studies of viral diseases. *Front Public Health*. 2019;7:1–18. <https://doi.org/10.3389/fpubh.2019.00208>.
- Trebecki P. Climate change and plant virus epidemiology. *Virus Res*. 2020;286:198059. <https://doi.org/10.1016/j.virusres.2020.198059>.
- Valverde S, Vidiella B, Montañez R, et al. Coexistence of nestedness and modularity in host–pathogen infection networks. *Nat Ecol Evol*. 2020;4:568–77. <https://doi.org/10.1038/s41559-020-1130-9>.
- van de Vossenberg BTLH, Visser M, Bruinsma M, et al. Real-time tracking of Tomato brown rugose fruit virus (ToBRFV) outbreaks in the Netherlands using Nextstrain. *PLoS ONE*. 2020;15:1–14. <https://doi.org/10.1371/journal.pone.0234671>.
- Verma RK, Mishra M, Marwal A, Gaur RK. Identification, genetic diversity and recombination analysis of Watermelon Mosaic Virus isolates. *3 Biotech* (2020) 10:1–8. <https://doi.org/10.1007/s13205-020-02248-8>.
- Vidal AH, Lacorte C, Sanches MM, et al. Characterization of cucurbit aphid-borne yellows virus (CABYV) from passion fruit in Brazil: evidence of a complex of species within CABYV isolates. *Viruses*. 2023;15:2. <https://doi.org/10.3390/v15020410>.
- Vurro M, Bonciani B, Vannacci G. Emerging infectious diseases of crop plants in developing countries: impact on agriculture and socio-economic consequences. *Food Secur*. 2010;2:113–32. <https://doi.org/10.1007/s12571-010-0062-7>.
- Wang J, Fernández-Pavía SP, Larsen MM, et al. High levels of diversity and population structure in the potato late blight pathogen at the Mexico centre of origin. *Mol Ecol*. 2017;26:1091–107. <https://doi.org/10.1111/mec.14000>.
- Yasaka R, Fukagawa H, Ikematsu M, et al. The timescale of emergence and spread of turnip mosaic potyvirus. *Sci Rep*. 2017;7:1–14. <https://doi.org/10.1038/s41598-017-01934-7>.
- Zarzyńska-Nowak A, Hasiów-Jaroszevska B, Budzyńska D, Borodynko-Filas N. First report of cucurbit aphid-borne yellows virus infecting zucchini plants (*Cucurbita pepo* convar. Giromontina) in Poland. *Plant Dis*. 2019;103:1047–7. <https://doi.org/10.1094/PDIS-10-18-1831-PDN>.