Article

Not peer-reviewed version

Metagenomic Analysis Reveals Viral Diversity of *Vanilla planifolia* in South Florida

Jesse Potts , Vincent N. Michael , Xingbo Wu *

Posted Date: 17 April 2026

doi: 10.20944/preprints202604.1247.v1

Keywords: *Vanilla planifolia*; *Cymbidium mosaic virus* (CymMV); *Vanilla latent virus* (VLV); metagenomics; viral diversity; phylogenetics; plant virome

 doi:10.20944/preprints202604.1247.v1

Article

Metagenomic Analysis Reveals Viral Diversity of *Vanilla planifolia* in South Florida

Jesse Potts, Vincent N. Michael and Xingbo Wu *

Tropical Research and Education Center, Horticultural Sciences Department, University of Florida IFAS, Homestead, FL 33031, USA

* Correspondence: xingbo.wu@ufl.edu

Abstract

Vanilla planifolia, a high-value tropical orchid, is significantly impacted by viral pathogens that threaten its cultivation and productivity. This study employs metagenomic techniques to detect and characterize the viral communities associated with *V. planifolia* in south Florida. Using high-throughput RNA sequencing, the *Cymbidium mosaic virus* (CymMV) and *Vanilla latent virus* (VLV) were prevalent in the plant system, with CymMV being the dominant viral species. Phylogenetic analysis of the CymMV coat protein gene revealed notable genetic divergence in the Homestead isolate, forming a distinct clade from global reference strains, suggesting local adaptation or host-specific evolution. Viral distribution the across plant system revealed higher viral loads in stem tissue, consistent with their role in systemic transport, whereas leaves exhibited greater viral richness, likely due to increased environmental exposure. The low abundance of other viral species, including Garlic viruses and *Senna severe yellow mosaic virus*, highlights the complex viral ecology associated with *V. planifolia*. This study underscores the value of metagenomic approaches for uncovering both well-characterized and novel viruses in plant systems and highlights the need for continuous viral surveillance to guide disease management strategies in economically important crops such as vanilla.

Keywords: *Vanilla planifolia*; *Cymbidium mosaic virus* (CymMV); *Vanilla latent virus* (VLV); metagenomics; viral diversity; phylogenetics; plant virome

1. Introduction

Vanilla (*Vanilla planifolia*) is an orchid with substantial economic importance to the tropical areas of the world. The pollinated flower of matured vanilla plants will develop into a bean-like fruit commonly known as vanilla capsule, which reaches to maturity after 8-9 months. The extract of cured mature beans possesses a unique aroma known as "vanilla flavor". Due to the complex production and processing, vanilla was largely produced overseas, with Madagascar, Uganda and Indonesia being the leading production countries[1]. The United States and European Union are the largest consumers of vanilla in the world [2].

As a well-known spice crop, global vanilla production is constantly being challenged with various abiotic and biotic stress. Though frequent extreme weather events such as excessive drought, heat and natural disaster impact the stability of vanilla supply regionally, biotic stresses resulting from fungal, bacterial, and viral pathogens remain the biggest challenge in vanilla cultivation worldwide [3]. Due to the disease pressures, a vanilla plantation only lasts averagely ten years with less than six years in good production [4], resulting in significant economic losses [5,6]. Among the various diseases affecting vanilla plants, viral infections constitute one of the most significant long-term threats to sustainable vanilla production. Viral infected vanilla plants can continue to grow with no visual symptoms at vegetative stage. As plants transition to reproductive phase, competition between fruit development and viral propagation quickly leads to declined plants with yellowing senescent leaves and reducing yield.

 doi:10.20944/preprints202604.1247.v1

Mechanical transmission during hand-pollination is a major contributor to the spread of viruses in commercial vanilla cultivation [4,7]. At present, no virus-resistant vanilla plant material is available, making viral diseases the most challenging constraint in vanilla management [6,8]. As a result, early detection and preventative practices remain the most effective control strategies. Among the viruses reported to infect vanilla, *Cymbidium mosaic virus* (CymMV) is regarded as the most virulent and economically damaging [4,7,9]. Although several immune-based strip assays are available for its detection, the sensitivity and diagnostic reliability vary due to high viral mutation rates [10,11]. Moreover, many viruses infecting vanilla remain uncharacterized, underscoring the need for a comprehensive understanding of the vanilla viral community to support early detection and effective disease prevention [3,12,13].

Traditional pathogen detection methods, including serological and PCR-based techniques, often fall short in sensitivity, particularly when addressing complex or novel viruses [14]. Additionally, these methods are typically designed for specific known pathogens, potentially overlooking emergent or uncharacterized viral threats [12,15]. The metagenomic techniques, which rely on high-throughput sequencing can facilitate the detection of broad spectrum viruses, including novel or low abundance, without requiring prior knowledge of the pathogen [13]. Metagenomics represents a transformative field that has unlocked unprecedented insights into the intricate microbial communities associated with various organisms, including plants [16]. By enabling the direct analysis of microbial genomes from environmental samples, this approach circumvents the traditional need for culturing, a process often limited in scope and time-consuming [17–19]. The advent of metagenomics has significantly advanced our understanding of plant-microbe interactions, uncovering complex networks of symbiotic relationships, pathogenic invasions, and overall microbiome structures that critically influence plant health, growth, and productivity [20,21]. This approach has been increasingly applied to various crops, offering fresh insights into plant virology and contributing to more effective disease management strategies [6,23]. Recent advancements in metagenomics have established a robust framework for detecting and characterizing viral populations in plants, including economically critical crops like grapevines (*Vitis vinifera*), tomatoes (*Solanum lycopersicum*), and cassava (*Manihot esculenta*) [15,23,24]. In grapevines, metagenomic approaches have led to the discovery of novel viral species and variants, deepening our understanding of grapevine viromes and aiding in disease management [15]. Similarly, in cassava, metagenomic studies have uncovered multiple begomoviruses linked to cassava mosaic disease, a major constraint on cassava production in Africa [23].

While metagenomic approaches have been successfully applied to characterize viral communities across a wide range of plant species [13,25], their use in high-value specialty crops such as vanilla remains limited, with only a small number of studies characterizing the vanilla virome to date [26,27]. Viral diseases pose well-documented threats to vanilla cultivation, with surveys in major production regions reporting infection rates of 20–50% in commercial plots and significant pre-harvest vine damage [28,29]. Given the scale of these losses, a comprehensive characterization of the viral community associated with *V. planifolia* represents an important step toward more effective disease management. Virome data of this kind have enabled the development of targeted diagnostic tools and evidence-based management strategies in other cropping systems [12,14,24] and the present study aims to generate comparable foundational knowledge for vanilla cultivation in the United States.

In this study, metagenomic approaches were used to identify and characterize the viral populations in south Florida, an area with emerging potential for vanilla cultivation. This represents the first study in the United States to profile the viral community associated with vanilla by analyzing different tissues. The research aims to expand the understanding of vanilla viromes and support the development of effective disease management strategies for sustainable vanilla cultivation.

 doi:10.20944/preprints202604.1247.v1

2. Materials and Methods

Sample Collection and RNA Extraction

Three replicates of leaf, stem and pod tissue samples were collected from two seven-year-old *V. planifolia* accessions (namely #23 and #85) representing the most severe symptomatic viral infection in the vanilla germplasm collection located in Homestead, FL described by[30]. Symptomatic plants exhibited characteristic mosaic patterns, chlorotic mottling of leaves, yellowing of older leaves, and reduced vine vigor. The collected plant tissues were immediately frozen in liquid nitrogen before homogenization using a mortar and pestle. Total RNA, including microbial RNA, was extracted using the E.Z.N.A.® Plant RNA Kit (Omega Bio-Tek, Norcross, GA, USA), following the manufacturer's protocol. Homogenized tissues (100 mg) were placed into 1.5 mL microcentrifuge tubes, and 500 µL of RB buffer with 20 µL of β-mercaptoethanol (BME) per 1 mL of buffer was added to each tube. The samples were vortexed to thoroughly mix and centrifuged at 14,000 x g for 2 minutes at room temperature. The cleared lysate was transferred to new microcentrifuge tubes.

To bind the RNA, 1 volume of 100% isopropanol was added to the lysate, and the mixture was transferred to the HiBind® RNA Mini Column placed in a 2 mL collection tube. The column was centrifuged at 10,000 x g for 1 minute. The column was then washed with 700 µL of RNA Wash Buffer I and centrifuged at 10,000 x g for 30 seconds. This was followed by a second wash with 500 µL of RNA Wash Buffer II. The RNA was eluted with 50 µL of Nuclease-free water by centrifuging the column for 1 minute at maximum speed. The extracted RNA was then checked for purity and integrity using electrophoresis. Following extraction, RNA samples were treated with DNase I to remove residual host DNA prior to library preparation.

Quality Control of RNA

The RNA concentration was quantified using fluorometry on a Qubit™ 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA**)** with the Qubit™ RNA Broad Range (BR) Assay Kit. For the quantification process, the required number of Qubit™ assay tubes were set up using two tubes of reference standards and one tube for each RNA sample. Thin-wall and clear 0.5 mL PCR tubes (Cat. No. Q32856) were used for the Qubit™ 4 Fluorometer. The Qubit™ RNA BR Reagent was diluted 1:200 in Qubit™ RNA BR Buffer to prepare the working solution, with 190 µL of working solution added to the standard tubes and 199 µL added to the sample tubes, leaving space for the RNA samples. A total of 1 µL of each RNA sample was added to its corresponding assay tube using a 2.5 µL pipette for precise measurement. The average RNA concentration across samples was 142.4 ng/µL, with an average A260/A280 purity ratio of 1.9. All tubes were vortexed for 3–5 seconds to ensure proper mixing and incubated at room temperature for 2 minutes. The RNA concentrations were then measured using the Qubit™ 4 Fluorometer, and the results were recorded for further analysis.

RNA Sequencing

For RNA sequencing, a cDNA library was prepared from the RNA samples obtained from the two *V. planifolia* accessions. The RNA was reverse transcribed into complementary DNA (cDNA) using amfiRivert Sensi cDNA Synthesis Master Mix(4X) (GenDEPOT). The cDNA was fragmented, and sequencing adapters were attached using NEBNext Ultra II DNA Library Prep Kit (New England Biolabs, USA). The library was amplified to ensure an adequate quantity of cDNA for sequencing. The cDNA libraries were then sequenced using the Illumina platform, enabling a comprehensive analysis of RNA from viral, microbial and plant community.

Viral Genome Assembly and Abundance analyses

Following sequencing, the raw reads were processed by filtering low quality reads using BBtools v38.90 [31] and mapped to vanilla reference genome [32]using HISAT2 v2.2.1 [33]. The mapped reads

 doi:10.20944/preprints202604.1247.v1

were discarded for further analysis. The unmapped reads were extracted with SAMtools v1.17 [34] and mapped to the VirSeq database [24]. Viral hits are defined as the number of distinct viral genome sequences in the VirSeq database to which at least one read from a given sample was mapped, serving as a measure of viral taxon richness per sample. Contigs mapping to viruses were assembled into full length genomes with RagTag [35].

The sequencing data were then analyzed to determine the composition and abundance of viral and microbial communities. Relative viral abundance was determined from the proportion of reads aligning to each viral genome in the VirSeq database, divided by the total number of mapped reads per sample, and visualized as percentage contributions using Krona charts [24,36]. This analysis provided insights into the relative abundance of different viral species and their potential roles in plant health. The results were visualized using Krona [36], facilitating the interpretation of the diversity and dynamics of viral community associated with *V. planifolia*.

Statistical Analysis

Virus hit data were analyzed by one-way ANOVA with Tukey's HSD for within-accession comparisons and two-sample t-tests for between-accession comparisons within each tissue type, with $\alpha = 0.05$ as the significance threshold. All statistical analyses were performed using R (version 4.3.2; R Core Team, 2023)[37].

Bioinformatic Analyses of CymMV and VLV

The coat protein gene sequences were aligned using Snap Gene 7.1.2[38]. A SnapGene Alignment was performed with the following parameters: Global alignment with free end gaps, Identity (1.0/0.0) as the cost matrix, a gap open penalty of 12, a gap extension penalty of 3, and 2 refinement iterations. This alignment allowed for the detection of mutations and variations in the coat protein gene across viral isolates, enabling a comparative analysis of CymMV sequences against global reference strains.

A phylogenetic tree was constructed to visualize the evolutionary relationships among the viral isolates using maximum likelihood implemented in iqtree2 [39]. The Interactive Tree of Life (iTOL) tool, available at https://itol.embl.de/, was used to visualize the tree based on the aligned coat protein sequences. iTOL facilitated the detailed visualization and annotation of the tree, highlighting the evolutionary divergence of the CymMV isolates from the vanilla samples.

3. Results

RNA Sequencing

The Illumina platform generated a total of 1.6 billion raw sequencing reads, of which 10% were removed for quality control, resulting 1.44 billion reads for downstream analyses. The number of clean reads, mapped reads, mapping efficiency and viral hits for each sample are summarized in Table 1.

Various mapping percentages were observed across the collected samples and different plant tissues. In both accessions, there were more reads mapped to leaf tissues (54.42% in #23 and 58.98% in #85) as compared to stem tissues (39.23% in #23 and 47.18% in #85), which coincide with the number of viral hits identified between the two tissue types (119 and 47 identified in the leaf tissue of #23 and #85, and 133 and 17 identified in the stem tissue of #23 and #85) (Table 1). Pod tissues displayed lower mapping percentages (21.90 - 29.10% in #23; 23.60–26.10% in #85) but exhibited a wide range of viral hit counts (38 - 168). There was more virus detected in #23 than #85 in both tissue types, while in both accessions the pod tissues exhibited intermediate viral abundance compared to leaves and stems. An average of 119 and 133 viral hits were found in leaf and stem tissues of #23, while only 47 and 17 viral hits were identified in the corresponding tissue of #85. No clear association was observed between mapping percentage and viral hit count across tissues or accessions. For example, on average 54.42% of sequencing reads and 119 viral hits were found in the leaf tissue, yet

 doi:10.20944/preprints202604.1247.v1

39.23% of sequence reads and 133 viral hits were found in the stem tissue of #23, while pod tissues showed intermediate mapping percentages but highly variable viral loads (38 - 168). Leaves exhibited greater viral richness, defined here as the number of distinct viral taxa detected, while stems accumulated the highest viral loads, particularly in #23. The viral hit distribution across tissues and accessions was shown on Figure 1. Panel A shows within-accession comparisons (Leaf, Stem, Pod) for #23 and #85, with bars representing the mean ± SD (Figure 1A). Letters above bars indicate Tukey groupings from one-way ANOVA (same letter = not significantly different at $\alpha = 0.05$). Panel B compares accessions (#23 vs #85) within each tissue, showing that stems differed significantly ($p = 0.0168$), leaves displayed a marginal trend ($p = 0.0519$), and pods were not significantly different ($p = 0.175$) (Figure 1B). These quantitative trends correspond with the mapping efficiencies summarized in Table 1, where tissues with higher mapping percentages also exhibited greater viral hit counts, and are further explored in relation to viral diversity patterns described below.

Table 1. Virus mapping statistics for vanilla tissue samples.

Sample	Reads passing QC	Mapped reads	Percentage mapped reads	Hits on VirSeq database	Sample Type
L_23_1	79,294,778	47,598,614	60.03%	132	Leaf
L_23_2	55,725,532	33,953,793	60.93%	104	Leaf
L_23_3	66,324,754	28,062,409	42.31%	120	Leaf
L_85_1	56,042,493	26,116,915	46.60%	86	Leaf
L_85_2	61,799,702	48,077,108	77.80%	24	Leaf
L_85_3	53,396,392	28,047,542	52.53%	30	Leaf
S_23_1	71,582,827	24,084,370	33.65%	134	Stem
S_23_2	58,295,915	20,466,786	35.11%	106	Stem
S_23_3	59,519,801	29,128,490	48.94%	160	Stem
S_85_1	59,568,487	23,401,252	39.28%	18	Stem
S_85_2	52,918,163	18,772,986	35.48%	14	Stem
S_85_3	54,753,363	36,566,888	66.78%	18	Stem
P_23_1	96,197,204	26,189,535	26.80%	88	Pod
P_23_2	129,441,824	28,621,993	21.90%	92	Pod
P_23_3	119,862,116	35,273,358	29.10%	92	Pod
P_85_1	112,795,066	29,744,931	26.10%	168	Pod
P_85_2	136,827,970	32,596,105	23.60%	38	Pod
P_85_3	120,955,084	29,062,325	23.80%	92	Pod

 doi:10.20944/preprints202604.1247.v1

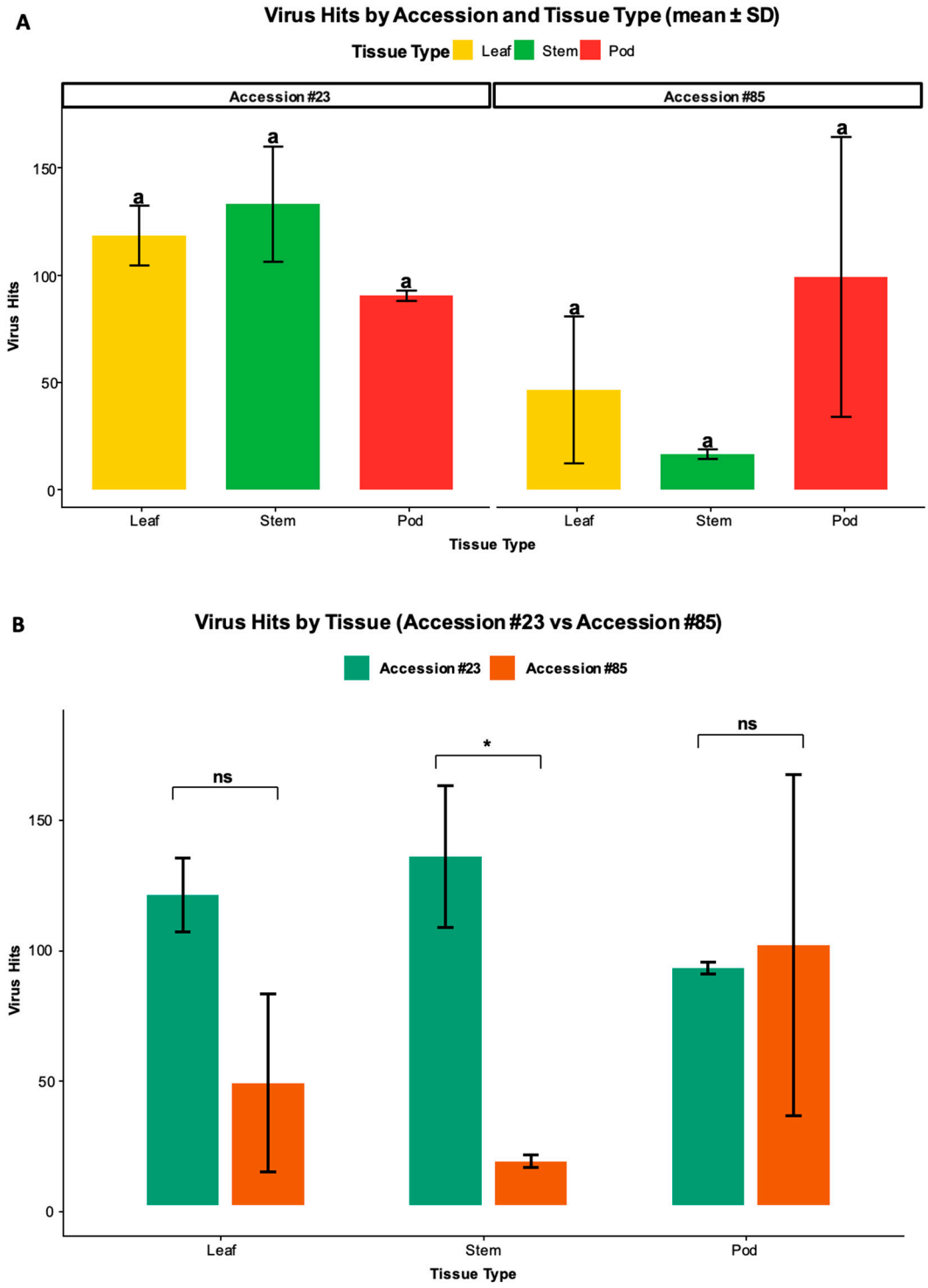


Figure 1. Virus hits by accession and tissue type (mean ± SD). (A) Within-accession comparisons of Leaf, Stem, and Pod tissues for #23 and #85, analyzed by one-way ANOVA followed by Tukey's HSD (letters above bars indicate significance groupings). (B) Between-accession comparisons (#23 vs #85) within each tissue using two-sample t-tests (stems significant, $p = 0.0168$; leaves not significant, $p = 0.0519$; pods, $p = 0.840$). Error bars represent SD ($n = 3$ for Leaf and Stem; $n = 3$ for Pod).

Viral Abundance in Vanilla

The virome of vanilla samples was found to be highly diverse, encompassing six viral families *Alphaflexiviridae, Phycodnaviridae, Tymoviridae, Geminiviridae, Betaflexiviridae* and *Mimiviridae.* Viral composition was consistent across leaf, stem, and pod tissues, with pods displaying comparable viral

 doi:10.20944/preprints202604.1247.v1

profiles to vegetative tissues. This confirms the systemic spread of dominant viruses such as *Cymbidium mosaic virus* (CymMV) and *Vanilla latent virus* (VLV) into reproductive organs.

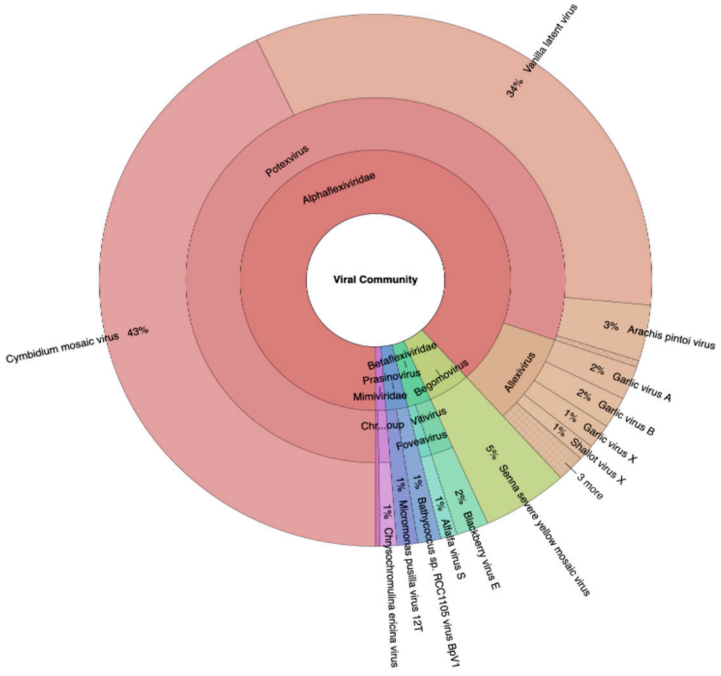
Besides these two prevailing viruses, numerous viral species with low prevalence were discovered, indicating a more intricate viral community than previously estimated. Several members of the Alphaflexiviridae family were identified, including Garlic viruses A, B, C, D, and E, and Shallot virus X, which are normally found in other plant hosts (Figure 2). Virus identification in vanilla indicates the possibility of cross-contamination through common tools in farming, insect vectors, or environmental contact. Furthermore, Senna severe yellow mosaic virus of Geminiviridae was found at quite a low incidence. This virus, which is mostly recognized to infect Senna species, leads to the questioning of its presence in vanilla and whether it signifies a new host relationship, a transient infection, or an artifact of environmental contamination.

Systemic Distribution and Co-Occurrence of CymMV and Vanilla Latent Virus

In addition to CymMV, VLV was consistently detected across all sampled tissues, including leaf, stem, and pod samples. The concurrent detection of both viruses across vegetative and reproductive organs indicates that infection is systemic rather than restricted to localized tissues or surface contamination. While CymMV represented the dominant viral species in terms of relative abundance, VLV constituted the second most prevalent component of the vanilla virome.

Although CymMV showed greater accumulation in stem tissues, VLV exhibited a more uniform proportional distribution across tissues, suggesting potential differences in replication dynamics or tissue tropism between the two viruses. Notably, the detection of VLV in pod tissues provides direct evidence that this virus can persist within reproductive organs of *V. planifolia*, extending previous reports that primarily relied on leaf-derived sequencing datasets.

The frequent co-occurrence of CymMV and VLV within the same tissues indicates that co-infection is common in vanilla plants. Such coexistence suggests that interactions between these viruses may influence infection dynamics, either through synergistic effects that enhance viral accumulation or through competitive interactions that modulate symptom development. Although the biological consequences of CymMV–VLV co-infection remain unresolved in vanilla, their consistent association across tissues highlights the importance of considering both viruses in disease monitoring and management strategies.



 doi:10.20944/preprints202604.1247.v1

Figure 2. *Proportional distribution of viruses detected in Vanilla planifolia.* Krona chart illustrating the hierarchical taxonomic composition of viruses detected across leaf, stem, and pod tissues. The innermost ring represents major virus families (e.g., Alphaflexiviridae), while outer rings show genus and species levels. *Cymbidium mosaic virus* (CymMV; ~43 %) and *Vanilla latent virus* (VLV) were predominant across all tissues. Additional families such as *Geminiviridae* and *Tymoviridae* appeared in smaller proportions. The inclusion of pod data highlights systemic viral movement into reproductive tissues.

Bioinformatic Analyses of CymMV and VLV

The phylogenetic analyses of Cymbidium mosaic virus showed a high degree of divergence of this virus across the globe, indicating its wide distribution and genetic stability (Figure 3). This included the CymMV Homestead isolates from this study, which showed a high degree of divergence from globally prevailing CymMV strains (Figure 3). This indicates that this strain carries important genetic variations with respect to the reference strain. It is possible that such genetic variations could be brought on by local environmental factors, host-specific adaptation, and/or an accumulation of mutations over time, which eventually caused divergence from the reference strain that may affect its pathogenicity, transmissibility, or interaction with the host plant *V. planifolia*.

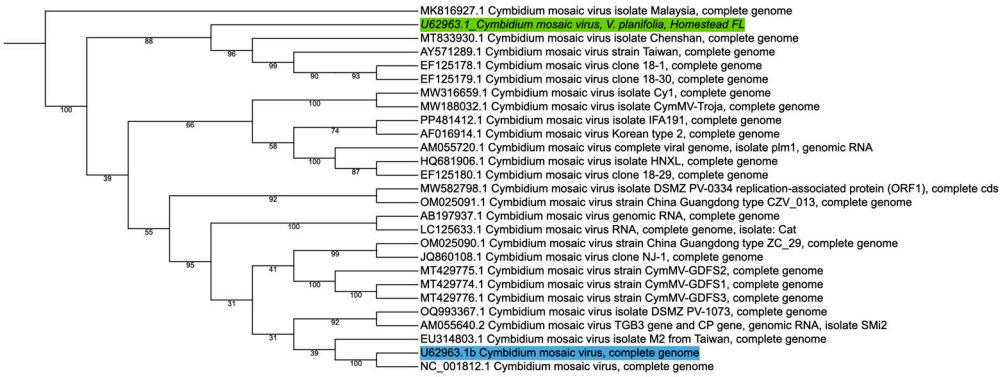


Figure 3. Phylogenetic analyses of *Cymbidium mosaic virus* (CymMV) isolates from *V. planifolia* in Homestead, FL. The labels highlight each virus isolate's accession number and host/source information.

Genetic variation analyses of CymMV Homestead isolate highlighted regions of genetic differences that may impact the structure and function of the viral coat protein (Figure 4). Nucleotide sequences analyses of identified CymMV isolates identified two 20 bp genomic hotspots that harbor the main genetic variations. This first hotspot ranges from 5,830 bp to 5,850 bp (Figure 4A), and the second genomic hotspot ranges from 5,840 bp to 5,860 bp (Figure 4B). Those hotspot genomic regions contain both transitions (purine-to-purine or pyrimidine-to-pyrimidine) and transversions (purine-to-pyrimidine) that are unique to the Homestead, FL isolate. Within these regions, most of the substitutions donate minor changes represented through transitions while the major changes are represented by transversions which might result in structural changes of the protein encoded by the virus.

 doi:10.20944/preprints202604.1247.v1

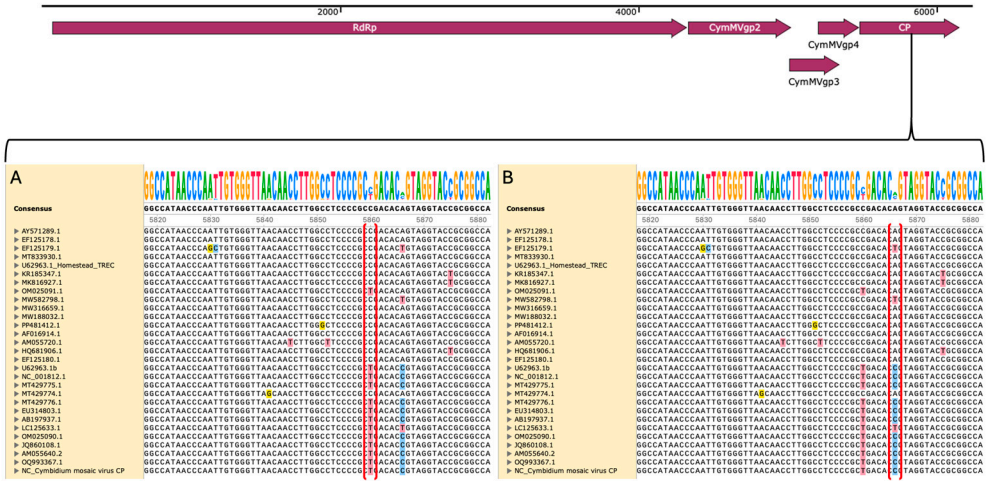


Figure 4. Comparative alignment of the ***Cymbidium mosaic virus*** **(CymMV)** coat protein gene from the Homestead, FL isolate and representative global reference isolates. The top panel shows genome-level alignment, with green regions indicating high sequence identity. Bottom panels (A–B) highlight localized nucleotide variation within the coat protein region (~5,830–5,870 bp), with red boxes indicating substitutions unique to the Homestead isolate.

Phylogenetic analysis of *Vanilla latent virus* (VLV) isolates revealed limited but discernible genetic variation among available reference sequences and the Homestead isolate (Figure 5A). All VLV sequences clustered within the genus *Allexivirus*, consistent with previous taxonomic assignments, and showed close relatedness to other reported vanilla-associated isolates. The Homestead VLV isolate formed a well-supported clade, confirming its identity and evolutionary placement. The comparatively reduced genetic variability observed for VLV likely reflects its persistent and largely asymptomatic infection strategy, which may impose lower selective pressure for rapid diversification. These results showed that VLV represents a stable and enduring component of the vanilla virome rather than a transient or incidental viral presence.

Comparative alignment of the Vanilla latent virus (VLV) coat protein (CP) gene revealed a high degree of sequence conservation across analyzed isolates, with only minor nucleotide substitutions detected (Figure 5B). No pronounced mutation hotspots or extensive transversion events were observed within the VLV CP region. This conservation suggests strong structural or functional constraints acting on the coat protein.

The limited CP divergence observed for VLV is consistent with its cryptic and largely asymptomatic infection behavior, which may facilitate long-term persistence within host tissues without eliciting strong defense responses.

 doi:10.20944/preprints202604.1247.v1

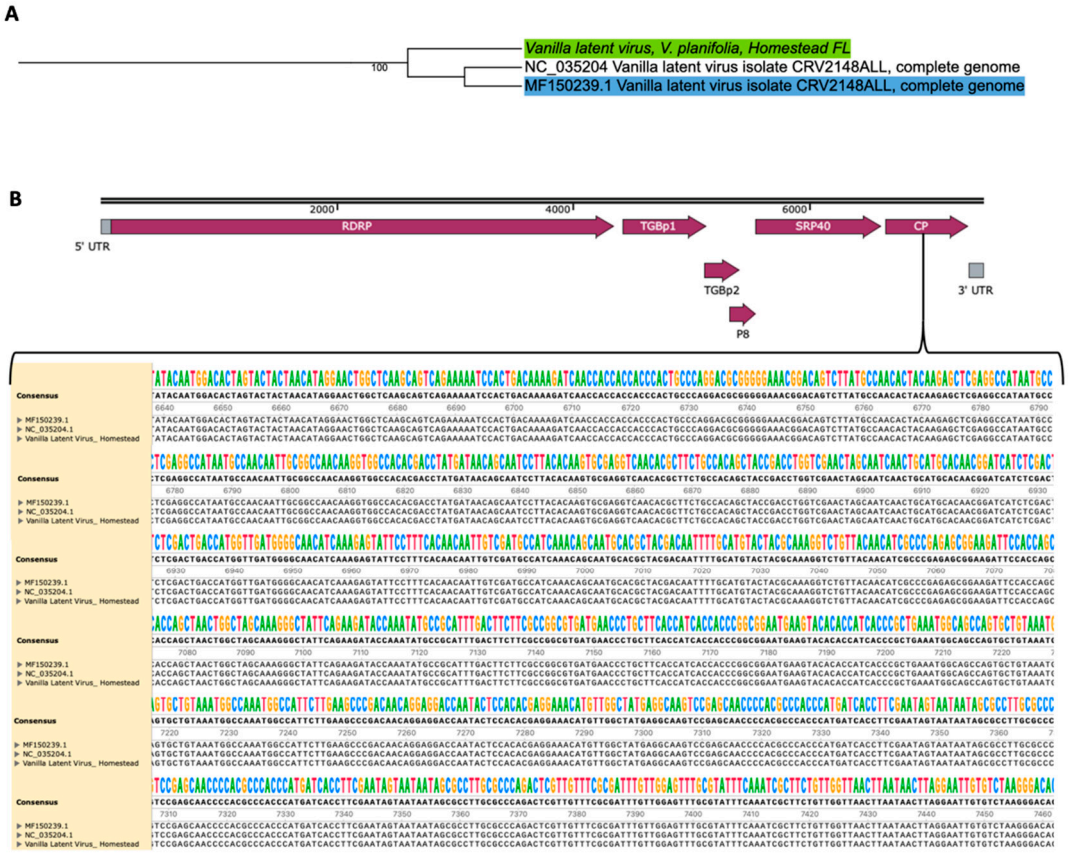


Figure 5. Phylogenetic and sequence analyses of Vanilla latent virus (VLV) isolates from V. planifolia in Homestead, FL. (A) Phylogenetic analysis of VLV isolates; labels indicate accession number and host/source. (B) Comparative alignment of the VLV coat protein (CP) gene from the Homestead isolate and representative global reference isolates.

4. Discussion

Despite being an important spice crop, vanilla is primarily cultivated in agroforestry systems worldwide and has undergone limited genetic improvement for modern commercial production [40]. Although these open and relatively unprotected cropping systems allow low production input, the average lifespan of vanilla plantations is typically only 6–8 years due to disease pressures such as viral infections [8]. With the expansion of the vanilla market and increasing demand for high-quality, natural products, vanilla cultivation has been steadily growing worldwide, including recent rapid expansion in Florida. The present study provides new insights into the viral ecology of *V. planifolia*, particularly highlighting *Cymbidium mosaic virus* (CymMV) and *Vanilla latent virus* (VLV) as the predominant viral agents detected in tropical production systems of the southern United States.

The widespread detection and high abundance of CymMV across multiple plant tissues corroborates previous reports identifying this virus as a major pathogen capable of systemic infection and efficient replication in vanilla plants [7]. In addition to these dominant viruses, the detection of viral sequences commonly associated with garlic, shallots, and other non-orchid hosts suggests potential horizontal transmission across plant species. Such cross-host transmission may be mediated by insect vectors including aphids, thrips, and mealybugs [3,41,42]. These observations indicate that orchid viruses can persist in mixed cropping systems and may spread through vegetative propagation practices commonly used in commercial vanilla cultivation.

Variation in viral abundance within stem tissues may reflect underlying genetic differences in virus tolerance among vanilla accessions. Detection of viruses in pod tissues further confirms that viral movement in *V. planifolia* is systemic and extends beyond vegetative organs. The presence of CymMV and VLV in fruits suggests that reproductive tissues may serve as potential viral reservoirs

 doi:10.20944/preprints202604.1247.v1

or transmission pathways. This observation aligns with studies in other crops demonstrating vertical or seed-associated viral transmission [4,43], and highlights the importance of phytosanitary screening of propagation materials to limit the spread of latent infections.

Stem tissues generally exhibited higher viral mapping efficiency, suggesting greater viral loads in these organs. This pattern likely reflects the role of stems in long-distance viral transport through the vascular system, facilitating systemic movement and replication of viruses such as CymMV. Similar patterns have been reported in other plant–virus systems, where viral RNA accumulates in vascular tissues involved in systemic transport [44]. In contrast, the greatest viral richness was observed in leaf samples. Leaves are highly exposed to environmental factors and potential vectors, which may explain the broader range of viral species detected in these tissues. This observation supports the hypothesis that leaves function as primary entry points for many plant viruses and therefore encounter greater viral richness[41,45]. The frequent detection of multiple viral species within individual plants also indicates the presence of co-infections, which may lead to synergistic interactions that increase disease severity or, conversely, competitive interactions that suppress symptom expression.

Phylogenetic analysis of CymMV isolates revealed notable genetic divergence in the Homestead isolate, which formed a distinct clade separate from previously reported global strains. This pattern suggests possible local adaptation or host-associated evolution within *V. planifolia* populations [11]. The Homestead isolate formed a distinct clade with 100% bootstrap support, confirming strong and unambiguous phylogenetic separation from all other global CymMV isolates included in the analysis. The phylogenetic separation was supported by sequence variation in the viral coat protein (CP) gene, where multiple nucleotide substitutions, including both transitions and transversions, were identified [46]. Such mutations may have functional implications, as the CP plays essential roles in virion assembly, host interaction, and systemic viral movement [10]. Structural or amino acid changes in the CP could therefore influence virion stability or host recognition, potentially altering infection dynamics associated with the Homestead isolate [47]. Previous studies have demonstrated that CP mutations can directly affect viral pathogenicity and transmission efficiency [10,46], suggesting that the observed divergence may represent adaptive evolution driven by host–virus interactions and local environmental pressures [48].

VLV was the second most abundant virus detected in this study. Although typically asymptomatic, its high prevalence suggests that it may play a significant role in the overall viral ecology of *V. planifolia*. Phylogenetic analysis places VLV within the genus *Allexivirus* (family Alphaflexiviridae), which includes viruses commonly associated with persistent infections in Allium species but has also been reported in non-Allium hosts, including vanilla [42]. The VLV genome contains genes encoding an RNA-dependent RNA polymerase, triple gene block (TGB) movement proteins, a coat protein, and a small additional open reading frame (~8 kDa) characteristic of a*llexiviruses* [9]. Previous high-throughput sequencing studies have identified VLV in vanilla using leaf-derived datasets but did not fully resolve its distribution across multiple plant organs, particularly reproductive tissues [9]. The detection of VLV in leaf, stem, and pod tissues in this study provides evidence of systemic infection rather than superficial contamination. The frequent co-occurrence of VLV with CymMV suggests that interactions between these viruses may occur in co-infected plants. In Allium crops such as garlic and shallot, *allexivirus* co-infections have been shown to intensify symptom expression, alter host defense responses, and increase viral replication rates [3]. Although comparable studies have not yet been conducted in vanilla, the consistent co-detection of both viruses within the same tissues raises the hypothesis that interactions between them may occur; however, no direct experimental evidence for such an interaction was obtained in the present study, and this possibility requires dedicated future investigation. In contrast to CymMV, which exhibited pronounced genetic divergence and multiple mutation hotspots within the coat protein gene in the Homestead isolate (Figure 4), VLV displayed a highly conserved coat protein sequence across analyzed isolates (Figure 5). This contrast reflects differences observed in the present dataset and highlights distinct patterns of coat protein variation between the two viruses.

 doi:10.20944/preprints202604.1247.v1

In other plant–virus systems, previous studies have shown that elevated coat protein variability in CymMV is associated with ongoing viral adaptation, potentially driven by host immune pressure, environmental selection, or long-term accumulation of mutations [10,11,46]. By contrast, members of the genus *Allexivirus,* including VLV, are known to establish persistent and often asymptomatic infections characterized by reduced sequence diversification and strong structural constraints on viral proteins [9,42]. Together, these observations suggest that CymMV is more likely to function as the dominant pathogenic contributor to disease progression in vanilla, whereas VLV may represent a biologically stable but less aggressive viral component within co-infected plants. Although VLV is generally considered asymptomatic, its widespread systemic presence raises the hypothesis that it could influence host physiology or modulate CymMV accumulation or symptom expression under specific conditions. Evidence for such interactions derives from studies in non-vanilla host systems rather than from the present dataset [3,43].

The observation of VLV in fruit (pod) tissues also raises the possibility of vertical or vegetative transmission, which could contribute to its persistence in commercial propagation systems. Considering its cryptic nature and systemic presence, VLV likely represents an underrecognized but biologically relevant component of the vanilla virome. Further studies focused on its genomic diversity, replication strategy, and potential interactions with other viruses will be essential to determine whether VLV acts as a passive endophyte or an active modulator of disease expression in *V. planifolia.*

These findings underscore the importance of continuous monitoring of viral populations in vanilla crops. As viral populations evolve, new variants may emerge with the potential to overcome existing disease management strategies, including the use of resistant plant varieties. For example, the genetic divergence of the Homestead CymMV isolate raises important implications for disease management. The presence of unique mutations in the coat protein gene could enable this variant to overcome resistance mechanisms in *V. planifolia* or spread more efficiently under local environmental conditions, as observed in other viral gene mutations [46]. The genetic divergence of the Homestead isolate of CymMV, combined with the observed viral diversity in leaf tissues, highlights the complexity of virus-host interactions in *V. planifolia* . The potential for cross-species transmission, as indicated by the presence of other viral species like *Garlic viruses* and *Senna severe yellow mosaic virus*, further complicates the viral landscape. Detection of these low-frequency viruses suggests that *V. planifolia* may encounter a broader range of viral agents than previously recognized, likely introduced through environmental exposure or vector transmission [29]. Understanding the evolutionary dynamics of these viral populations, including the role of genetic mutations in coat proteins, will be critical for developing more effective disease management strategies, particularly for high-value crops like vanilla [49]. It should be noted that viral identifications in this study were based exclusively on high-throughput sequencing data. While CymMV and VLV were detected at consistently high read depths across multiple tissue types and biological replicates, the low-abundance viruses, including Garlic virus A, B, C, and D, and Senna severe yellow mosaic virus, were identified at read depths that cannot fully exclude background sequencing noise. RT-PCR or qPCR confirmation using virus-specific primers is recommended in future studies to validate the presence of these low-abundance taxa in *V. planifolia*. Furthermore, this study was conducted using two *V. planifolia* accessions selected to represent the most severe symptomatic cases in the Homestead germplasm collection. As such, the findings should be interpreted as preliminary data providing the first tissue-level virome characterization of *V. planifolia* in the United States, rather than as population-level conclusions. Expanding the sample set to include a broader range of accessions and symptom severity levels will be an important next step for future studies.

5. Conclusions

This study highlights the significant role of *Cymbidium mosaic virus* (CymMV) and *Vanilla latent virus* (VLV) in the viral ecology of *V. planifolia*, with notable genetic divergence observed in the Homestead CymMV isolate. The observed viral diversity, tissue-specific viral accumulation, and

 doi:10.20944/preprints202604.1247.v1

phylogenetic divergence underscore the complexity of virus-host interactions and the importance of ongoing monitoring and re-search into the evolution of plant viruses in vanilla cultivation. Importantly, this study provides the first integrated tissue-level and phylogenetic evidence demonstrating that VLV is a systemic and persistent component of the vanilla virome, co-existing with CymMV across both vegetative and reproductive organs.

Author Contributions: Conceptualization, J.P., V.M., and X.W; methodology, J.P. and V.M; software, J.P. and V.M; validation, J.P., V.M., and X.W; formal analysis, J.P. and V.M.; investigation, J.P. and V.M.; resources, X.W.; data curation, J.P. and V.M.; writing—original draft preparation, J.P.; writing—review and editing, V.M., and X.W.; visualization, J.P. and V.M.; supervision, X.W.; project administration, X.W.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2022-38640-37488 through the Southern Sustainable Agriculture Research and Education program under subaward number LS23-381. USDA is an equal opportunity employer and service provider.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

Acknowledgments: The authors sincerely appreciate the support from Nicolas Portogues. During the preparation of this manuscript/study, the author(s) used AI tool for the purposes of addressing any grammatic issues. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

#23, #85	*Vanilla planifolia* accessions 23 and 85
ANOVA	Analysis of variance
BME	β-mercaptoethanol
CP	Coat protein
CymMV	*Cymbidium mosaic virus*
DNA	Deoxyribonucleic acid
HTS	High-throughput sequencing
IFAS	Institute of Food and Agricultural Sciences
iTOL	Interactive Tree of Life
ML	Maximum likelihood
RNA	Ribonucleic acid
RdRp	RNA-dependent RNA polymerase
SD	Standard deviation
TGB	Triple gene block
UF/IFAS TREC	University of Florida Institute of Food and Agricultural Sciences, Tropical Research and Education Center
V. planifolia	*Vanilla planifolia*

 doi:10.20944/preprints202604.1247.v1

VLV	*Vanilla latent virus*

References

1. Iftikhar, T.; Majeed, H.; Waheed, M.; Zahra, S.S.; Niaz, M.; AL-Huqail, A.A. Vanilla. In *Essentials of medicinal and aromatic crops*; Springer, 2023; pp. 341–371.
2. Barragán-Ocaña, A.; Silva-Borjas, P.; Cecilio-Ayala, E. Vanilla Production in the World and Mexico: Market Value and Technology. *Social Sciences & Humanities Open* **2024**, *10*, 101076, doi:10.1016/j.ssaho.2024.101076.
3. Villamor, D.; Ho, T.; Al Rwahnih, M.; Martin, R.; Tzanetakis, I. High Throughput Sequencing for Plant Virus Detection and Discovery. *Phytopathology* **2019**, *109*, 716–725.
4. Gu, F.; Xu, F.; Wu, G.; Zhu, H.; Ji, C.; Wang, Y.; Zhao, Q.; Zhang, Z. Annual Accumulation of CymMV May Lead to Loss in Production of Asymptomatic Vanilla Propagated by Cuttings. *Plants* **2024**, *13*, 1505, doi:10.3390/plants13111505.
5. Simmonds, P.; Adams, M.J.; Benk\Ho, M.; Breitbart, M.; Brister, J.R.; Carstens, E.B.; Davison, A.J.; Delwart, E.; Gorbalenya, A.E.; Harrach, B.; et al. Virus Taxonomy in the Age of Metagenomics. *Nature Reviews Microbiology* **2017**, *15*, 161–168.
6. Barba, M.; Czosnek, H.; Hadidi, A. Historical Perspective, Development and Applications of next-Generation Sequencing in Plant Virology. *Viruses* **2014**, *6*, 106–136.
7. Zettler, F.W.; Ko, N.; Wisler, G.; Elliott, M.; Wong, S. Viruses of Orchids and Their Control. **1990**.
8. Straver, J. Vanilla Planifolia HC Andrews. *Plant Resources of South-East Asia* **1999**, 228–233.
9. Grisoni, M.; Marais, A.; Filloux, D.; Saison, A.; Faure, C.; Julian, C.; Theil, S.; Contreras, S.; Teycheney, P.-Y.; Roumagnac, P.; et al. Two Novel Alphaflexiviridae Members Revealed by Deep Sequencing of the Vanilla (Orchidaceae) Virome. *Arch Virol* **2017**, *162*, 3855–3861, doi:10.1007/s00705-017-3540-9.
10. Ajjikuttira, P.; Lim-Ho, C.; Woon, M.; Ryu, K.; Chang, C.; Loh, C.; Wong, S. Genetic Variability in the Coat Protein Genes of Two Orchid Viruses: Cymbidium Mosaic Virus and Odontoglossum Ringspot Virus: Brief Report. *Archives of virology* **2002**, *147*, 1943–1954.
11. Moles, M.; Delatte, H.; Farreyrol, K.; Grisoni, M. Evidence That Cymbidium Mosaic Virus (CymMV) Isolates Divide into Two Subgroups Based on Nucleotide Diversity of Coat Protein and Replicase Genes. *Archives of virology* **2007**, *152*, 705–715.
12. Maina, S.; Donovan, N.J.; Plett, K.; Bogema, D.; Rodoni, B.C. High-Throughput Sequencing for Plant Virology Diagnostics and Its Potential in Plant Health Certification. *Front. Hortic.* **2024**, *3*, 1388028, doi:10.3389/fhort.2024.1388028.
13. Roossinck, M.J. Plant Virus Metagenomics: Biodiversity and Ecology. *Annu. Rev. Genet.* **2012**, *46*, 359–369, doi:10.1146/annurev-genet-110711-155600.
14. Adams, I.P.; Glover, R.H.; Monger, W.A.; Mumford, R.; Jackeviciene, E.; Navalinskiene, M.; Samuitiene, M.; Boonham, N. Next-Generation Sequencing and Metagenomic Analysis: A Universal Diagnostic Tool in Plant Virology. *Molecular plant pathology* **2009**, *10*, 537–545.
15. Al Rwahnih, M.; Daubert, S.; Golino, D.; Islas, C.; Rowhani, A. Comparison of Next-Generation Sequencing versus Biological Indexing for the Optimal Detection of Viral Pathogens in Grapevine. *Phytopathology* **2015**, *105*, 758–763.
16. Kumar, S.; Krishnani, K.K.; Bhushan, B.; Brahmane, M.P. Metagenomics: Retrospect and Prospects in High Throughput Age. *Biotechnology research international* **2015**, *2015*, 121735.
17. Chaudhari, H.G.; Prajapati, S.; Wardah, Z.H.; Raol, G.; Prajapati, V.; Patel, R.; Shati, A.A.; Alfaifi, M.Y.; Elbehairi, S.E.I.; Sayyed, R.Z. Decoding the Microbial Universe with Metagenomics: A Brief Insight. *Front. Genet.* **2023**, *14*, 1119740, doi:10.3389/fgene.2023.1119740.
18. Handelsman, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiol Mol Biol Rev* **2004**, *68*, 669–685, doi:10.1128/MMBR.68.4.669-685.2004.
19. Handelsman, J.; Rondon, M.R.; Brady, S.F.; Clardy, J.; Goodman, R.M. Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products. *Chemistry & Biology* **1998**, *5*, R245–R249, doi:10.1016/S1074-5521(98)90108-9.

 doi:10.20944/preprints202604.1247.v1

20. Quince, C.; Walker, A.W.; Simpson, J.T.; Loman, N.J.; Segata, N. Shotgun Metagenomics, from Sampling to Analysis. *Nature biotechnology* **2017**, *35*, 833–844.
21. Simon, C.; Daniel, R. Metagenomic Analyses: Past and Future Trends. *Applied and environmental microbiology* **2011**, *77*, 1153–1161.
22. Pagán, I.; García-Arenal, F. Tolerance to Plant Pathogens: Theory and Experimental Evidence. *International journal of molecular sciences* **2018**, *19*, 810.
23. Kreuze, J.F.; Perez, A.; Untiveros, M.; Quispe, D.; Fuentes, S.; Barker, I.; Simon, R. Complete Viral Genome Sequence and Discovery of Novel Viruses by Deep Sequencing of Small RNAs: A Generic Method for Diagnosis, Discovery and Sequencing of Viruses. *Virology* **2009**, *388*, 1–7, doi:10.1016/j.virol.2009.03.024.
24. Massart, S.; Olmos, A.; Jijakli, H.; Candresse, T. Current Impact and Future Directions of High Throughput Sequencing in Plant Virus Diagnostics. *Virus Research* **2014**, *188*, 90–96, doi:10.1016/j.virusres.2014.03.029.
25. Wylie, S.J.; Li, H.; Dixon, K.W.; Richards, H.; Jones, M.G.K. Exotic and Indigenous Viruses Infect Wild Populations and Captive Collections of Temperate Terrestrial Orchids (Diuris Species) in Australia. *Virus Research* **2013**, *171*, 22–32, doi:10.1016/j.virusres.2012.10.003.
26. Rott, M.; Xiang, Y.; Boyes, I.; Belton, M.; Saeed, H.; Kesanakurti, P.; Hayes, S.; Lawrence, T.; Birch, C.; Bhagwat, B.; et al. Application of Next Generation Sequencing for Diagnostic Testing of Tree Fruit Viruses and Viroids. *Plant Disease* **2017**, *101*, 1489–1499, doi:10.1094/PDIS-03-17-0306-RE.
27. Roossinck, M.J.; Martin, D.P.; Roumagnac, P. Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology®* **2015**, *105*, 716–727, doi:10.1094/PHYTO-12-14-0356-RVW.
28. Richard, A.; Farreyrol, K.; Rodier, B.; Leoce-Mouk-San, K.; Wong, M.; Pearson, M.; Grisoni, M. Control of Virus Diseases in Intensively Cultivated Vanilla Plots of French Polynesia. *Crop Protection* **2009**, *28*, 870–877, doi:10.1016/j.cropro.2009.06.003.
29. Randriambololona, M.; Rakotonirina, T.J.; Rieux, A.; Grisoni, M. First Report of Odontoglossum Ringspot Virus in Vanilla (Orchidaceae) in Madagascar. *Plant Disease* **2023**, *107*, 1250.
30. Chambers, A.; Cibrián-Jaramillo, A.; Karremans, A.P.; Moreno Martinez, D.; Hernandez-Hernandez, J.; Brym, M.; Resende, M.F.R.; Moloney, R.; Sierra, S.N.; Hasing, T.; et al. Genotyping-By-Sequencing Diversity Analysis of International Vanilla Collections Uncovers Hidden Diversity and Enables Plant Improvement. *Plant Science* **2021**, *311*, 111019, doi:10.1016/j.plantsci.2021.111019.
31. Bushnell, B. BBMap: A Fast, Accurate, Splice-Aware Aligner. **2014**.
32. Piet, Q.; Droc, G.; Marande, W.; Sarah, G.; Bocs, S.; Klopp, C.; Bourge, M.; Siljak-Yakovlev, S.; Bouchez, O.; Lopez-Roques, C.; et al. A Chromosome-Level, Haplotype-Phased Vanilla Planifolia Genome Highlights the Challenge of Partial Endoreplication for Accurate Whole-Genome Assembly. *Plant Communications* **2022**, *3*, 100330, doi:10.1016/j.xplc.2022.100330.
33. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-Based Genome Alignment and Genotyping with HISAT2 and HISAT-Genotype. *Nat Biotechnol* **2019**, *37*, 907–915, doi:10.1038/s41587-019-0201-4.
34. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079, doi:10.1093/bioinformatics/btp352.
35. Alonge, M.; Lebeigle, L.; Kirsche, M.; Jenike, K.; Ou, S.; Aganezov, S.; Wang, X.; Lippman, Z.B.; Schatz, M.C.; Soyk, S. Automated Assembly Scaffolding Using RagTag Elevates a New Tomato System for High-Throughput Genome Editing. *Genome Biol* **2022**, *23*, 258, doi:10.1186/s13059-022-02823-7.
36. Ondov, B.D.; Bergman, N.H.; Phillippy, A.M. Interactive Metagenomic Visualization in a Web Browser. *BMC Bioinformatics* **2011**, *12*, 385, doi:10.1186/1471-2105-12-385.
37. R Core Team *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2023;
38. SnapGene.
39. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; Von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **2020**, *37*, 1530–1534, doi:10.1093/molbev/msaa015.
40. Andriatsitohaina, R.N.N.; Laby, P.; Llopis, J.C.; Martin, D.A. Agroforestry in Madagascar: Past, Present, and Future. *Agroforest Syst* **2024**, *98*, 1659–1680, doi:10.1007/s10457-024-00975-y.

 doi:10.20944/preprints202604.1247.v1

41. Bragard, C.; Caciagli, P.; Lemaire, O.; Lopez-Moya, J.; MacFarlane, S.; Peters, D.; Susi, P.; Torrance, L. Status and Prospects of Plant Virus Control through Interference with Vector Transmission. *Annual review of phytopathology* **2013**, *51*, 177–201.
42. Mansouri, F.; Ryšánek, P. Allexivirus: Review and Perspectives. *Phytopathol. Mediterr.* **2021**, *60*, 389–402, doi:10.36253/phyto-12043.
43. Pagán, I.; García-Arenal, F. Tolerance to Plant Pathogens: Theory and Experimental Evidence. *IJMS* **2018**, *19*, 810, doi:10.3390/ijms19030810.
44. Dunham, J.; Simmons, H.; Holmes, E.; Stephenson, A. Analysis of Viral (Zucchini Yellow Mosaic Virus) Genetic Diversity during Systemic Movement through a Cucurbita Pepo Vine. *Virus research* **2014**, *191*, 172–179.
45. Kondo, H.; Takemoto, S.; Maruyama, K.; Chiba, S.; Andika, I.B.; Suzuki, N. Cymbidium Chlorotic Mosaic Virus, a New Sobemovirus Isolated from a Spring Orchid (Cymbidium Goeringii) in Japan. *Archives of virology* **2015**, *160*, 2099–2104.
46. Sherpa, A.R.; Hallan, V.; Pathak, P.; Zaidi, A.A. Complete Nucleotide Sequence Analysis of Cymbidium Mosaic Virus Indian Isolate: Further Evidence for Natural Recombination among Potexviruses. *Journal of biosciences* **2007**, *32*, 663–669.
47. He, Z.; Jiang, D.; Liu, A.; Sang, L.; Li, W.; Li, S. The Complete Sequence of Cymbidium Mosaic Virus from Vanilla Fragrans in Hainan, China. *Virus genes* **2011**, *42*, 440–443.
48. Gourdel, S.; Leclercq-Le Quillec, F. Coat Protein Gene of Cymbidium Mosaic Virus from Vanilla Fragrans in Reunion Island. *Journal of Phytopathology* **2001**, *149*, 561–564.
49. Roossinck, M.J. Plant Virus Ecology. *PLoS pathogens* **2013**, *9*, e1003304.