Phytopathology® • 2021 • 111:1051-1058 • https://doi.org/10.1094/PHYTO-09-20-0420-R

Virology

e-Xtra*

Sixty Years from the First Disease Description, a Novel Badnavirus Associated with Chestnut Mosaic Disease

Armelle Marais,[1,†] Sergio Murolo,[2] Chantal Faure,[1] Yoann Brans,[3] Clément Larue,[4,5] François Maclot,[6] Sébastien Massart,[6] Michela Chiumenti,[7] Angelantonio Minafra,[7] Gianfranco Romanazzi,[2] Marie Lefebvre,[1] Teresa Barreneche,[1] Cécile Robin,[4] Rémy J. Petit,[4] and Thierry Candresse[1]

[1] University of Bordeaux, INRAE, UMR BFP, Villenave d'Ornon, France
[2] Department Agricultural, Food and Environmental Sciences, Università Politecnica delle Marche, Ancona, Italy
[3] Laboratoire de Virologie et de Biologie Moléculaire, Centre Technique Interprofessionnel des Fruits et Légumes, Prigonrieux, France
[4] University of Bordeaux, INRAE, UMR Biogeco, Cestas, France
[5] INVENIO, Maison Jeannette, Douville, France
[6] Plant Pathology Laboratory, TERRA-Gembloux Agro-BioTech, University of Liège, Belgium
[7] National Research Council of Italy Institute for Sustainable Plant Protection, Bari, Italy
Accepted for publication 19 October 2020.

ABSTRACT

Although chestnut mosaic disease (ChMD) was described several decades ago, its etiology is still not clear. Using classical approaches and high-throughput sequencing (HTS) techniques, we identified a novel *Badnavirus* that is a strong etiological candidate for ChMD. Two disease sources from Italy and France were submitted to HTS-based viral indexing. Total RNAs were extracted, ribodepleted, and sequenced on an Illumina NextSeq500 (2 × 150 nt or 2 × 75 nt). In each source, we identified a single contig of ≈7.2 kb that corresponds to a complete circular viral genome and shares homologies with various badnaviruses. The genomes of the two isolates have an average nucleotide identity of 90.5%, with a typical badnaviral genome organization comprising three open reading frames. Phylogenetic analyses and sequence comparisons showed that this virus is a novel species; we propose the name *Chestnut mosaic virus* (ChMV). Using a newly developed molecular detection test, we systematically detected the virus in symptomatic graft-inoculated indicator plants (chestnut and American oak) as well in chestnut trees presenting typical ChMD symptoms in the field (100 and 87% in France and Italy surveys, respectively). Datamining of publicly available chestnut sequence read archive transcriptomic data allowed the reconstruction of two additional complete ChMV genomes from two *Castanea mollissima* sources from the United States as well as ChMV detection in *C. dentata* from the United States. Preliminary epidemiological studies performed in France and central eastern Italy showed that ChMV has a high incidence in some commercial orchards and low within-orchard genetic diversity.

Keywords: *Badnavirus*, *Castanea* sp., *Caulimoviridae*, etiology, high-throughput sequencing, viral disease

European chestnut (*Castanea sativa* Mill.) has a long-standing tradition of cultivation in many European countries. It is an important species economically, as a source of timber and fruit, and ecologically, through the multiple ecosystemic services it provides. In Europe, chestnut covers ≈2.5 million hectares, mainly concentrated in France, Italy, Spain, Portugal, Switzerland, the Balkan regions, and southern England (Conedera et al. 2016). Chestnut (*Castanea* spp.) can be heavily affected by various pathogens. The most detrimental are caused by fungal-like organisms (Oomycetes) and fungi such as *Phytophthora cambivora* Petri and *P. cinnamomic* Rands., the agents of ink disease, or *Cryphonectria parasitica*, which is the causal agent of chestnut blight, and all provoke disorders that can lead to tree mortality (Prospero et al. 2012; Rigling and Prospero 2018). In Italy, Gualaccini (1958) described a chestnut disease associated with viral symptoms (mosaic, shoots with asymmetric leaf blade deformation) that was reported in Campania during the 1980s (Ragozzino and Lahoz 1986) and in the Marche region (central eastern Italy) in 2000 (Antonaroli and Perna 2000). In France, the disease was first identified circa 1987 on cultivars of *C. sativa* × *C. crenata* hybrids from commercial orchards located in the southwest of the country. Desvignes (1999b) provided a more detailed description of the symptoms, which include necrotic lesions in the bark and wood that turn into cankers, chlorotic lesions and yellow stripes on leaf veins, and partial limb atrophy, and called this disease chestnut mosaic disease (ChMD). This disease can heavily impact the production of both young and secular trees (Antonaroli and Perna 2000). It has also been reported in Japan and Hungary (Horvath et al. 1975; Shimada 1962). Even though its etiology has remained unknown, researchers hypothesized that the causal agent of ChMD could be a virus introduced in Europe between 1940 and 1960, when a number of *C. crenata* cultivars were imported from Japan for breeding purposes. Investigations in France and Italy established that the causal agent can be eliminated by thermotherapy, is aphid-transmissible, and is graft-transmissible to *Castanea* and *Quercus* species, in which it may elicit symptoms (Desvignes 1999b; Desvignes and Lecocq 1995; Vettraino et al. 2005). The susceptibility to the ChMD agent of *Castanea* species/cultivars has been evaluated in several studies (Desvignes 1992, 1999b; Desvignes and Lecocq 1995). Three categories of cultivars could be defined: tolerant, moderately susceptible, and fully susceptible. Graft incompatibility was also observed when cultivars of

†Corresponding author: A. Marais; armelle.marais-colombel@inrae.fr

Funding: Part of this work was based on COST Action FA1407 (DIVAS) and supported by COST (European Cooperation in Science and Technology).

*The *e*-Xtra logo stands for "electronic extra" and indicates there are supplementary materials published online.

First and second authors contributed equally to this work.

The nucleotide sequences reported here have been deposited in GenBank under accession numbers MT261366, MT269853, MT270664 to MT270682, and MT339503 to MT339590.

The author(s) declare no conflict of interest.

different susceptibilities were assembled by grafting. Most of the *C. sativa* cultivars and hybrids are tolerant to ChMD, although some well-known French hybrids like 'Maraval' (Ca 74) are fully susceptible and used for indexing purposes to detect the ChMD agent in tolerant cultivars (Desvignes and Lecocq 1995).

In the past decade, a number of studies have highlighted the potential of nontargeted molecular diagnostics based on high-throughput sequencing (HTS) to elucidate the etiology of viral plant diseases and to provide viral sequence data from which rapid diagnostic molecular assays can be developed (Martin et al. 2016; Villamor et al. 2019). Since 2009, HTS and bioinformatics have been used for the discovery, characterization, and de novo assembly of the genome of known and novel plant viruses and viroids (Kreuze et al. 2009; Rott et al. 2017). This has accelerated the application of HTS technologies in the field of disease diagnostics (Massart et al. 2014) and in quarantine regulations (Martin et al. 2016; Massart et al. 2017).

Badnaviruses are plant pararetroviruses that belong to the family *Caulimoviridae* and have emerged as serious pathogens causing severe yield losses in a wide range of economically important crops worldwide (Bhat et al. 2016). The genome of badnaviruses is composed of a noncovalently closed, circular, double-stranded DNA (range, 7.2 to 9.2 kbp) and is encapsidated in bacilliform virions. This genome typically harbors three open reading frames (ORFs) encoding, respectively, a protein of unknown function, the virion-associated protein, and a polyprotein containing functional and structural domains (movement protein [MP], coat protein, aspartic protease, reverse-transcriptase [RT], and RNase H) (Bhat et al. 2016; Hohn and Rothnie 2013). Badnaviruses can also be present as integrated sequences in some host plant genomes (endogenous badnaviruses) (Bhat et al. 2016; Staginnus et al. 2009). The contributions of these integrated sequences to host and virus evolution are still poorly understood (Geering et al. 2014).

Because of the very limited knowledge of the etiology of ChMD, and based on previously published studies (Desvignes 1992, 1999a, b; Desvignes and Cornaggia 1996; Desvignes and Lecocq 1995), we investigated the hypothesis that a virus might be involved in this disease. After combining HTS-based viral indexing and classical approaches, we report the complete genome sequence of a novel badnavirus species that we propose calling *Chestnut mosaic virus* (ChMV). We show that there is a strict correlation between the presence of the virus and the appearance of typical ChMD symptoms in various graft-inoculated indicator plants. Preliminary epidemiological studies performed in Italy and in France revealed that the virus can have a high incidence in some orchards and, as expected, can be associated with symptomatic or asymptomatic infections.

MATERIALS AND METHODS

Plant samples and virus isolates. Virus isolates included in this study are listed in Supplementary Table S1. Isolate LC1224H is originated from a red oak (*Q. rubra*) artificially inoculated in 1992 with a chestnut mosaic source from a hybrid *C. sativa* × *C. crenata* included in a French breeding program. Leaves of grafted oaks displayed typical symptoms including chlorotic mottle, yellow veins, and mosaic (Desvignes and Lecocq 1995) (Fig. 1A). Isolate FRlc1224A was derived from the same source and is the result of a back-inoculation by the grafting of LC1224H to the natural chestnut hybrid Maraval (Ca 74; *C. crenata* × *C. sativa*) indicator (Desvignes 1992). Isolate LC1224F originated from a Maraval indicator inoculated by aphid transmission from an initial ChMD source in a *C. crenata* × *C. sativa* French hybrid (Desvignes and Cornaggia 1996). The LCA552 and LCA584 isolates were collected from *C. sativa* trees in France in 2009 and 2018, and the T32018 disease source was isolated from a French hybrid *C. crenata* × *C. sativa* in 2018. All of these isolates have been held and propagated on 'Maraval' indicator plants at the Centre Technique Interprofessionnel des Fruits et Légumes (CTIFL) virology laboratory (Lanxade, France).

In the framework of a survey performed in Italian chestnut orchards to monitor chestnut blight (Acquasanta Terme [AP], Locality Umito, Italy) (Murolo et al. 2018), typical leaf symptoms of ChMD were recorded in 2016. Six symptomatic plants were collected, pooled (10 to 15 symptomatic shoots), and included in the HTS analysis (ITumito39 source).

To evaluate the incidence of ChMV, chestnut trees from INRAE chestnut biological resource center (https://www6.bordeaux-aquitaine.inrae.fr/biogeco/Ressources) were sampled. This orchard is located at the Villenave d'Ornon INRAE center (France), with trees distributed in three plots (A, E, and Port) (Supplementary Table S1). A total of 43 *C. sativa*, 14 *C. mollissima*, 6 *C. crenata*, and 32 hybrid chestnut trees were sampled and corresponded to a total of 38 symptomatic trees with typical ChMD symptoms, 47 asymptomatic trees, and 10 trees with dubious or atypical symptoms. In addition, in the central eastern Italy Marche region, leaves from 60 symptomatic and 10 asymptomatic grafted *C. sativa* 'Marrone' trees of different ages were collected at a commercial chestnut orchard (Plot I) (Supplementary Table S1).

Isolates FRlc1224A and ITumito39 were used for the HTS analysis. All other samples were included either in the incidence analysis or in the causal relationship analysis (Supplementary Table S1).

Total RNA extraction and RNA-Seq analysis. Symptomatic leaves from a 'Maraval' indicator (FRlc1224A) were collected and used to extract total RNAs according to the protocol described by Reid et al. (2006). For the Italian material, total RNAs were extracted from symptomatic leaves according to the protocol described by Gambino et al. (2008). Total RNAs were then submitted to a DNAse treatment following the manufacturer's recommendations (Fisher Scientific, Illkirch, France). Ribosomal RNAs were removed using a RiboMinus Plant Kit for RNA-Seq (Invitrogen, Fisher Scientific, Illkirch, France) before cDNA library synthesis with the Illumina TruSeq Stranded RNA library prep kit (Illumina

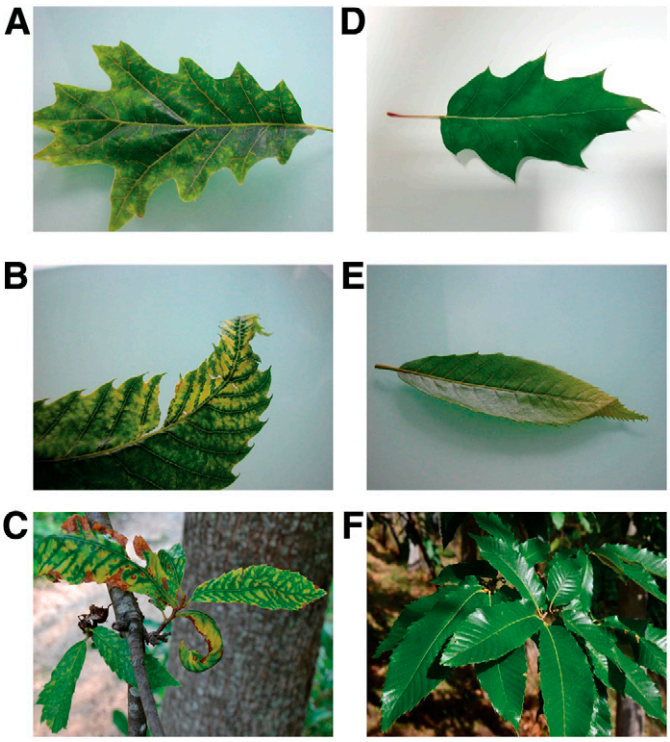

Fig. 1. Symptoms of chestnut mosaic disease on various hosts. **A,** Isolate LC1224H: Red oak (*Quercus rubra*) graft-inoculated with a diseased source. **B,** Isolate FRlc1224A: 'Maraval' Ca 74 graft-inoculated with LC1224H. **C,** Isolate ITumito39: symptomatic leaves from 'Marrone' grafted onto *Castanea sativa*. **D,** Noninoculated *Q. rubra*. **E,** Noninoculated 'Maraval' Ca 74. **F,** Asymptomatic leaves from 'Marrone' grafted onto *C. sativa*.

Inc., San Diego, CA) and sequenced on an Illumina NextSeq500 (2 × 150 nt or 2 × 75 nt) in a multiplexed format (GIGA-Genomics Facility, Université de Liège, Liège, Belgium).

Bioinformatic analysis. Primary quality analyses were performed using Geneious Prime 2019.2.1 software (https://www.geneious.com). De novo assemblies of quality-filtered reads were performed using Velvet (Zerbino and Birney 2008), Geneious R 11 (https://www.geneious.com), and Spades (Bankevich et al. 2012), or by using the CLC genomics workbench 8.0 (https://www.clcbio.com). Contigs were annotated by BlastN and BlastX comparisons with nucleotide and nonredundant protein GenBank databases, respectively. Blast results were screened using e-value thresholds of 10^{-6} and 10^{-4} for BlastN and BlastX, respectively. Publicly available chestnut RNA-Seq transcriptomic data were retrieved from the National Center for Biotechnology Information (NCBI) Sequence Read Archive, and downloaded reads were mapped against the sequence of the FRlc1224A isolate using CLC Genomics Workbench 11.0. When needed, de novo assembly and contig annotations were also performed as described.

Total DNA extraction and PCR confirmation of genome completeness and circularity. To verify the completeness of the assembled genome sequences and genome circularity, pairs of specific outward-facing primers were designed for each isolate (Ch-Bad-6976F/Ch-Bad-252R for the isolate FRlc1224A and Bad-Ch-6481F/Bad-Ch-325R for the isolate ITumito39) (Supplementary Table S2). Leaf tissues (0.5 g) were pulverized in liquid nitrogen and total DNAs were extracted in CTAB buffer (2% cetyl trimethylammonium bromide, 100 mM Tris-HCl, 1.4 M NaCl, and 20 mM EDTA) by adding 3% polyvinyl pyrrolidone 40 and 0.5% sodium metabisulfite (Doyle and Doyle 1990). Finally, the DNA pellets were resuspended in 50 µl of sterile water. Polymerase chain reactions (PCR) were performed in a 50-µl reaction volume containing 10 mM Tris-HCl (pH 8.5), 2 mM $MgCl_2$, 50 mM KCl, 0.2 mM dNTPs, forward and reverse primers at 1 µM each, and 1.25 U of Dream Taq (ThermoFisher) using 50 ng of the template. After an initial denaturation step at 95°C for 4 min, 40 cycles (Ch-Bad-6976F/Ch-Bad-252R) and 35 cycles (Bad-Ch-6481F/325R) were set at 94°C for 30 s, 60°C (Ch-Bad-6976F/Ch-Bad-252R) or 55°C (Bad-Ch-6481F/325R) for 30 s, and 72°C for 90 s, followed by a final extension step of 10 min at 72°C. PCR amplification products were sequenced on both strands (GATC; Eurofins, Ebersberg, Germany).

ChMV molecular detection and variant analysis by PCR. For the molecular detection of ChMV, two sets of primers were designed in conserved regions of ORF3 designed using the sequences of isolates FRlc1224A and ITumito39. One primer pair (Ch-Bad-1466F/Ch-Bad-1800R) (Supplementary Table S2) allows the amplification of a genomic region (335 nt) in the MP domain (Fig. 2), whereas the second pair (Ch-Bad-5860F/Ch-Bad-6109R) (Supplementary Table S2) amplifies a 232-nt fragment in the RH domain (Fig. 2). An aliquot of 25 ng of total DNA was used for the PCR assays in a 50-µl volume containing 10 mM Tris-HCl (pH 8.5), 2 mM $MgCl_2$, 50 mM KCl, 0.2 mM dNTPs, forward and reverse primers at 1 µM each, and either 1.25 units of DreamTaq or 1 unit of GoTaq. After an initial denaturation step at 95°C for 4 min, 35 cycles were set at 94°C for 30 s, 56°C for 30 s, and 72°C for 90 s, followed by a final extension step of 10 min at 72°C. Amplicons were analyzed by electrophoresis on 1.5% agarose gel and directly sequenced on both strands (GATC). Possible phytoplasma infection was evaluated using primer pair P1/P7 (Deng and Hiruki 1991; Smart et al. 1996) and, in nested PCR, primers R16F2n/R2 (Gundersen and Lee 1996).

Sequence and phylogenetic analyses. The full-length genomes were analyzed by ORF Finder (https://www.ncbi.nlm.nih.gov/projects/gorf/) to identify putative ORFs in the viral genome. Deduced amino acid (aa) sequences were analyzed for conserved protein domains gathered in the Conserved Domains Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), and theoretical molecular weights were calculated using ExPASy (https://web.expasy.org/compute_pi/). Multiple alignments of nucleotide (nt) or amino acid (aa) sequences were performed using the ClustalW program (Thompson et al. 1994) implemented in MEGA version 7.0 (Kumar et al. 2016). Genetic distances (p-distances using a strict nucleotide or amino acid identity) were calculated using MEGA 7.0. Phylogenetic trees were reconstructed using the neighbor-joining method implemented in MEGA 7.0, and the robustness of nodes was assessed from 1,000 bootstrap resamplings.

RESULTS

Determination of the complete genome sequence of a novel badnavirus from two chestnut disease sources. Two ChMD sources were included in the HTS analysis. The French source (FRlc1224A) showed typical ChMD symptoms, including leaf deformation, yellow veins, and chlorotic diffuse mottling (Fig. 1B). The Italian source (ITumito39) is a mixture of six plants that showed intensive vein banding and leaf blade deformation (Fig. 1C). The HTS of ribodepleted RNAs extracted from the sources FRlc1224A and ITumito39 yielded a total of 10,737,052 and 4,135,330 reads, respectively. De novo assembly and Blast annotation allowed for the identification of a single long contig with significant homology with badnaviruses. These contigs were, respectively, 7,264 and 7,214 bp long and showed short terminal redundancies, consistent with the structure of the long RNA transcript involved in the replication of badnaviruses (Teycheney et al. 2020) and suggesting they represented the full coverage of a circular badnaviral genome. A total of 39,657 reads were integrated in the FRlc1224A contig, representing 0.37% of total reads, with a mean coverage depth of 795×; 611 reads (0.015% of total reads) were integrated in the ITumito39 contig, with a mean coverage depth of 14.4×. The circularity and completion of the DNA genome sequence of each isolate were validated by PCR of purified DNA extracted from the host plants and using specific outward-facing primers designed from the contig sequences. The respective 436- and 1,007-nt fragments were amplified and sequenced, confirming DNA genome completeness and circularity (data not shown). The assembled sequences have been deposited in GenBank under accession numbers MT269853 (for the FRlc1224A contig) and MT261366 (for the ITumito39 contig). No other plant

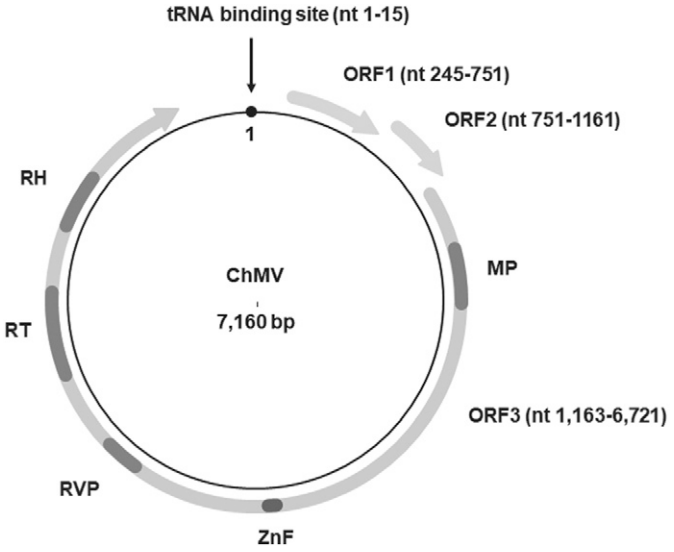

Fig. 2. Schematic representation of the genomic organization of the chestnut mosaic virus. The tRNA binding site is indicated and defines position 1 on the genome. The three open reading frames (ORFs) are shown as gray arrows, and their positions are shown in parentheses. Five conserved motifs are identified in the ORF3 polyprotein: MP, viral movement protein (pfam01107); ZnF, zinc finger (pfam00098); RVP, retroviral aspartyl protease (pfam00077); RT, reverse transcription (cd01647); and RH, ribonuclease H (cd09274).

virus was detected in the two datasets during the Blast annotation of contigs.

Genome organization of ChMV and determination of its phylogenetic relationships. The badnaviral genomes characterized independently from the French and Italian ChMD sources are within the range of badnavirus genome sizes (7,160 and 7,161 bp long, respectively) (Teycheney et al. 2020). The genomic organization is the same for both isolates; it comprises three ORFs encoded on the positive strand (Fig. 2), which is typical for badnaviruses (Teycheney et al. 2020). The ORF1 (nt 245 to 751, numbering according to the isolate FRlc1224A sequence) encodes a protein of 169 aa (19.8 kDa), the ORF2 (nt 751 to 1161) encodes a 137-aa protein (15 kDa), and the third ORF (nt 1,163 to 6,721) encodes a polyprotein of 1,853 aa (211.7 kDa) with five conserved protein domains (Fig. 2): a viral movement protein (MP; cl03100), a zinc-binding motif (ZnF; pfam00098), a retroviral aspartyl protease domain (RVP; pfam00077), an RT domain (cd01647), and a ribonuclease H domain (RH; cl14782). The two "Cys" motives (C-X_2-C-X_4-H-X_4-C and C-X_2-C-X_{11}-C-X_2-C-X_4-C-X_2-C) usually found in the coat protein of badnaviruses (Bhat et al. 2016) were also detected in the ORF3-deduced protein between amino acid positions 777 and 790 and 902 and 928.

Both isolates are closely related, with an overall 90.5% nt identity. The three indels observed between the two sequences are located in the intergenic region; the isolate ITumito39 ended up being one nucleotide longer. The three ORFs have the same sizes and are strictly colinear; the encoded proteins share, respectively, 95.2% (ORF1), 95.5% (ORF2), and 94.8% (ORF3) amino acid identity.

To characterize the phylogenetic relationships and taxonomic position of the chestnut badnavirus, a phylogenetic tree was reconstructed using an alignment of full genome nucleotide sequences of members of the genus *Badnavirus*, with the rice tungro bacilliform virus used as an outgroup (Fig. 3). Both isolates cluster in group 3, defined by Wang et al. (2014), together with gooseberry vein banding virus, rubus yellow net virus, grapevine vein-clearing virus, birch leafroll-associated virus, wisteria badnavirus 1 (WBV1), and pagoda yellow mosaic-associated virus (Fig. 3). Nevertheless, they are clearly distant from all of these species, defining a novel branch supported by a 99% bootstrap value (Fig. 3). Tree topology was similar when using an alignment of representative badnaviral ORF3 protein sequences (Supplementary Fig. S1). To confirm these analyses, pairwise comparisons of genome sequences showed that the isolate FRlc1224A has only weak identity levels with representative members of the genus *Badnavirus*, comprising between 42.1% nt identity (sugarcane bacilliform IM virus; 42.5% for the isolate ITumito39) and 50.9% nt identity (WBV1; 50.8% for the isolate ITumito39). The same tendency is observed when considering the genome proteins. The ORF1-encoded protein shows only weak homology with the corresponding proteins of WBV1 (27.8% aa identity) and pagoda yellow mosaic-associated virus (26.1% aa identity), and the ORF2-encoded protein shares only 33.1% aa identity with the corresponding protein of the most closely related virus, WBV1. The polyprotein encoded by ORF3 shares 49.5% aa identity with the corresponding protein of the closest relative, pagoda yellow mosaic-associated virus. Using the ORF3 region (RT and RH domain) used for taxonomical discrimination in the family *Caulimoviridae* (Teycheney et al. 2020), the FRlc1224A isolate shows between 64% (with gooseberry vein banding virus) and 68.4% nt (with birch leafroll-associated virus) identity (Table 1), which is less than the 80% nt identity value used as the species demarcation threshold in the family. Therefore, this virus represents a novel species in the family *Caulimoviridae*. In the same taxonomically informative region, the isolates FRlc1224A and ITumito39 share 91.9% nt identity (97.8% aa identity), indicating that they belong to the same viral species (Table 1).

Identification of ChMV in publicly available chestnut HTS data. Datamining of chestnut HTS data from various chestnut sources publicly available at GenBank (EST sequences, whole genome assembly, RNA-Seq, and genotyping-by-sequencing reads available in the Sequence Read Archive) allowed the identification of ChMV in several of those datasets (Supplementary Table S3). In particular, two complete genomes were obtained from datasets involving *C. mollissima* 'Vanuxem' in the United States, one from the whole genome assembly (JRKL01079565) and the other from de novo assembly of RNA-Seq data (SRX4015368), with 99.2 and 97.4% nt identity, respectively, with the FRIc1224A isolate over the whole genome (Fig. 4). In addition, partial ChMV genome assemblies larger than kilobase pairs could be obtained from a range of other datasets generated in the United States or China from *C. mollissima* (Supplementary Table S3); all of these showed significant relatedness with the FRlc1224A sequence, as shown by a phylogenetic tree reconstructed using nucleotide alignments of concatenated ChMV sequences retrieved from the various datasets (Fig. 4). In addition, partial ChMV genomes could be reconstructed from two datasets obtained from *C. dentata* in the United States. Interestingly, one of these two *C. dentata* isolate sequences shows the closest relationship with the ITumito39 sequence (Fig. 4), with only 89.2% nt identity with the isolate FRlc1224A compared with 93.9% nt identity with ITumito39. The second isolate of *C. dentata* appears to be equally related to the FRlc1224A and ITumito39 isolates, with 90.9 and 90.6% nt identity, respectively.

Incidence and genetic variability of ChMV in France and Italy. The incidence and genetic variability of ChMV were investigated by analyzing two genomic regions of ORF3, one 335-nt-long located in the MP domain amplified using primer pair Ch-Bad-1466F/Ch-Bad-1800R and the other 232-nt-long in the RNase H domain and amplified with primer pair Ch-Bad-5860F/Ch-Bad-6109R Supplementary Table S2 and Fig. 2). The two primer pairs were designed to be able to detect isolates FRlc1224A and ITumito39. In Italy, a total of 70 *C. sativa* 'Marrone' samples were collected in the same location. In France, 95 chestnut accessions belonging to three different *Castanea* species or hybrids were sampled in three plots (A, Port, E). Both symptomatic and asymptomatic samples were collected, along with some samples with atypical or dubious symptoms. Globally, ChMV was frequent in the surveyed plots, with 57 of 70 (81.5%) infected *C. sativa* samples in Italy and 65 of 95 trees (68%) in France (Table 2). In the Italian orchard, half of the asymptomatic trees and 87% of the symptomatic trees were found to be infected by ChMV (Table 2). None of the analyzed samples was found positive using a phytoplasma-specific PCR assay. In the French collection, the virus was detected in 100% (38/38) of the trees showing typical ChMD symptoms and in 49% (23/47) of the asymptomatic trees, including two trees that were symptomless but showed strong symptoms on rootstock off-shoots (Supplementary Fig. S2). ChMV was also detected in 4 of the 10 trees showing atypical or suspicious symptoms.

The genetic variability of ChMV was evaluated by analyzing the sequences of the two PCR amplicons generated for the incidence survey. Considering the relative homogeneity of the origin of the Italian samples, the number of samples included in this analysis was limited to 13 (four from asymptomatic trees and nine symptomatic ones) (Supplementary Table S1). The final dataset consisted of a total of 53 isolates for which the sequences of the two genomic regions were available (49 from the incidence survey and four from independent ChMD sources held in collection at CTIFL). As illustrated by the unrooted neighbor-joining tree reconstructed from the alignments of RT-RNase H domain nucleotide sequences (Supplementary Fig. S3A), ChMV diversity is structured into two clusters defined by the geographical origin of the samples (Italy and France). The sequences determined from the four independent French disease sources (FRlc1224A, T30218, LCA552, and LCA584) belong to the same French cluster. Overall, the level of genetic diversity is very low in this genomic region, with an average pairwise nucleotide divergence (diversity) of 2.2 ± 0.5%. This value is even lower when considering the intragroup diversity of 0.2 ± 0.1% within the French cluster and 0.1 ± 0.1% within the Italian ones. In contrast, the intergroup

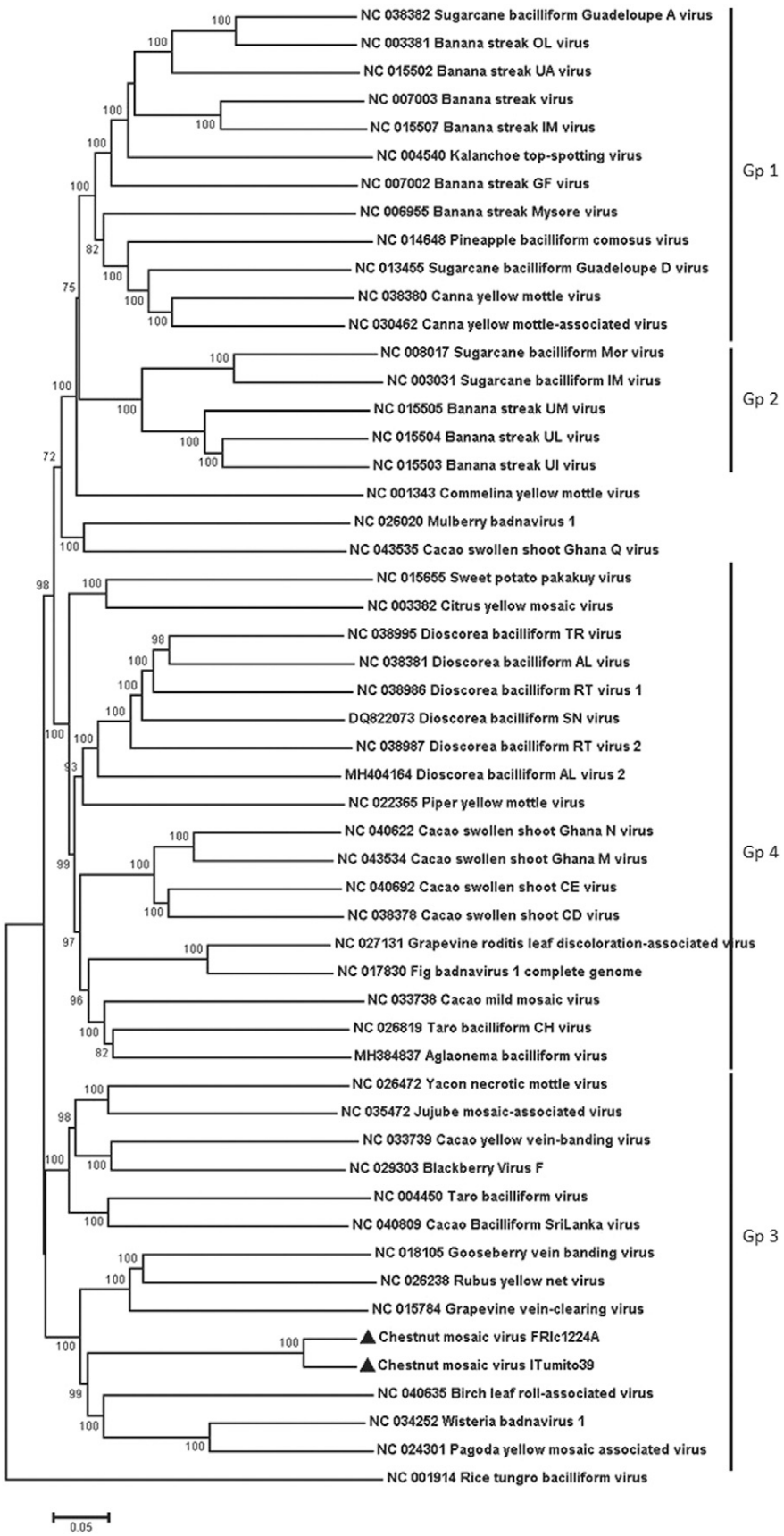

Fig. 3. Phylogenetic tree reconstructed using the complete genome sequences of badnavirus members. Virus names as well as GenBank accession numbers are indicated. The tree was reconstructed using the neighbor-joining method, and randomized bootstrapping was performed to evaluate the statistical significance of branches (1,000 replicates). Bootstrap values more than 70% are shown. The scale bar represents 5% nucleotide divergence between sequences. The groups, as defined by Wang et al. (2014), are indicated. Chestnut mosaic virus isolates determined in this work are indicated by black triangles. Rice tungro bacilliform virus was used as the outgroup.

diversity reaches 6.3 ± 1.5%, confirming the existence of two geographical clusters. The same trends are observed when analyzing the genomic region located in the MP domain (Supplementary Fig. S3B). The same geographical clustering could be observed, with the exceptions of three French isolates that seem to be more closely related to the Italian cluster (Supplementary Fig. S3B). Another French isolate, 20971-E, remains isolated and does not fit in either group. The average nucleotide divergence in this region is slightly higher than that in the RT-RNase H region (5.8 ± 0.7%), and the intergroup diversity reaches a value of 13.4 ± 1.8%; however, the value for the other region is 6.3%.

DISCUSSION

Since the seminal work of Desvignes and collaborators in the 1990s, it has been acknowledged that the agent responsible for ChMD is most likely a thermosensitive, graft-transmissible virus that can be transmitted experimentally and probably naturally by the aphid *Myzocallis castanicola* (Desvignes 1992, 1999a, b; Desvignes and Cornaggia 1996; Desvignes and Lecocq 1995). Depending on the chestnut genotype, this infection can be asymptomatic or result in the expression of severe and conspicuous ChMD symptoms. In chestnut orchards in the Marche region (Italy), both young and mature plants were affected, thus significantly decreasing chestnut production. Symptoms have also been observed in some *Quercus* species following experimental graft inoculation. To date, however, the causal agent remains to be identified.

By using HTS-based viral indexing, we were able to identify and characterize, in two independent ChMD sources, two isolates of the same novel virus. Phylogenetic and sequence analyses showed that this virus belongs to the genus *Badnavirus*, in the family *Caulimoviridae*, and could be considered a new species (proposed as ChMV). Interestingly, this new virus clusters with a group of badnaviruses including rubus yellow net virus, GVBaV, and grapevine vein-clearing virus.

TABLE 1. Percentage of identity between the open reading frame 3 (ORF3) region encoding the reverse transcription RNase H[a] of chestnut mosaic virus (ChMV) isolate FRlc1224A and the corresponding genomic regions of the isolate ITumito39 and of the most closely related members of the genus *Badnavirus*

Virus[b]	Nucleotide identity	Amino acid identity
ChMV ITumito39	91.9%	97.8%
RYNV	65.1%	71.6%
GVBV	64.0%	68.9%
GVCV	66.8%	72.6%
BLRaV	68.4%	72.6%
WBV1	68.2%	71.6%
PYMaV	67.7%	71.9%

[a] This region is typically used for taxonomic discrimination in the family *Caulimoviridae* (Teycheney et al. 2020).
[b] RYNV, rubus yellow net virus; GVBV, gooseberry vein banding virus; GVCV, grapevine vein-clearing virus; BLRaV, birch leafroll-associated virus; WBV1, wisteria badnavirus 1; and PYMaV, pagoda yellow mosaic-associated virus.

There is unambiguous evidence that ChMV, as reported here, is an episomal virus. It was detected in graft-inoculated indicators, but not in noninoculated control plants of the same variety, demonstrating its graft-transmissibility, which is a property of episomal viruses. This line of evidence is further reinforced by the detection of ChMV in symptomatic, graft-inoculated indicator *Quercus* plants, but, again, not in the corresponding control plants. In parallel, the HTS detection of ChMV in DNAse-treated RNAs, the failure to detect ChMV in a range of the surveyed chestnut trees, and the sequence diversity identified in ChMV all rule out a scenario in which an endogenous ChMV genome integrated in the chestnut genome could be responsible for the HTS and PCR results reported here. There was, in fact, no indication of ChMV in the chestnut genome assembly (JRKL01079565) because no integration borders could be identified and a single contig, representing a complete unintegrated viral genome transcript, was identified. Therefore, integration of ChMV as an endogenous viral element (Bhat et al. 2016) does not appear to be a general genomic feature of chestnut.

According to the simplified hierarchical approach proposed by Fox (2020) for assessing causal relationships in plant virology, ChMV appears to be a good candidate for, if not the causative agent of, ChMD. There are several arguments and experimental evidence supporting this idea. After HTS analyses, ChMV was the sole virus detected in the French source FRlc1224A from a ChMD source initially involving a *C. sativa* × *C. crenata* hybrid. It was also the sole virus detected in the Italian ChMD source analyzed by HTS. Using molecular detection tests developed during this work, the virus was consistently found in other symptomatic accessions derived from the same diseased source (LC1224H, a *Q. rubra* artificially inoculated, and LC1224F, an indicator plant inoculated by aphid transmission)

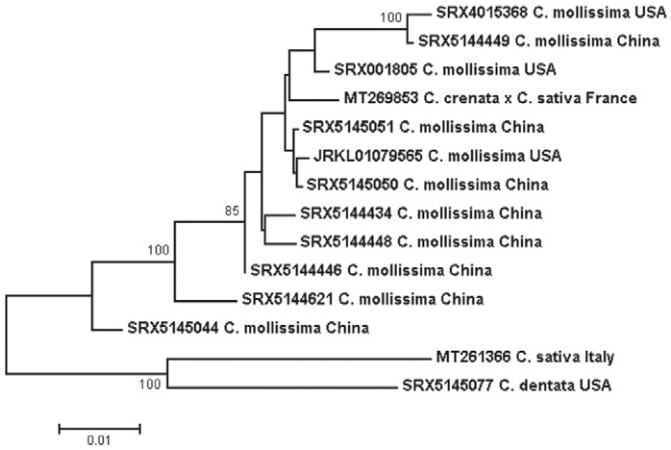

Fig. 4. Unrooted neighbor-joining phylogenetic tree reconstructed from the alignment of concatenated nucleotide sequences related to chestnut mosaic virus detected by datamining of publicly available transcriptomic chestnut data. Randomized bootstrapping was performed to evaluate the statistical significance of branches (1,000 replicates). Bootstrap values more than 70% are shown. The scale bar represents 10% nucleotide divergence between sequences.

TABLE 2. Number and percentage of chestnut mosaic virus-infected plants of the plot, the *Castanea* species sampled, and symptomatology

Origin of the sampled plants	Infected/total plants	Infected/asymptomatic plants	Infected/symptomatic plants	Infected/plants with atypical symptoms
France (overall)	65/95 (68.4%)	23/47 (48.9%)	38/38 (100%)	4/10 (40%)
Plot A	13/23 (56.5%)	6/15 (40%)	6/6 (100%)	1/2 (50%)
Plot E	48/66 (72.7%)	15/28 (53.6%)	30/30 (100%)	3/8 (37.5%)
Plot Port	4/6 (66.6%)	2/4 (50%)	2/2 (100%)	na[a]
C. crenata	5/6 (83%)	1/1 (100%)	4/4 (100%)	0/1
C. sativa	30/43 (70%)	11/20 (55%)	18/18 (100%)	1/5 (20%)
C. mollissima	13/14 (93%)	1/2 (50%)	10/10 (100%)	2/2 (100%)
Hybrid[b]	17/32 (53.1%)	10/24 (41.6%)	6/6 (100%)	1/2 (50%)
Italy *C. sativa*	57/70 (81.5%)	5/10 (50%)	52/60 (87%)	na[a]

[a] na, not applicable.
[b] Interspecific hybrids among *C. crenata*, *C. mollisima*, and *C. sativa.*

(Fig. 5). Additionally, three other independent chestnut sources shown by biological indexing on the 'Maraval' indicator to be affected by ChMD were found to be infected by ChMV (LCA552, LCA584, and T32018 in Fig. 5). Therefore, there is a correlation between the appearance of ChMD symptoms and the presence of ChMV in the graft-inoculated indicators, supporting the hypothesis of a causal relationship between ChMV infection and ChMD. A total of five independent ChMD sources collected between 1990 and 2018 in two countries (Italy and France) were ChMV-positive, satisfying the experimental and consistency criteria (Bradford Hill 1965; Fox 2020).

Preliminary studies indicate that ChMV is highly prevalent in the analyzed orchards in France and Italy, confirming the earlier results of Desvignes (1999a). In parallel, the identification of ChMV sequences in publicly available HTS data provides a strong indication of the presence of ChMV in *C. mollissima* in the United States and China and in *C. dentata* in the United States. In the surveys, ChMV was not systematically associated with symptomatic infections, although its frequency was systematically higher in symptomatic plants. This result was expected because previous grafting experiments had demonstrated that not all chestnut varieties/species are susceptible to ChMD and develop symptomatic infections (Desvignes 1992, 1999b; Desvignes and Lecocq 1995). Biological indexing on the susceptible 'Maraval' indicator, in particular, has identified latent ChMV infections in many symptomless *C. sativa* varieties or *C. sativa* × *C crenata* hybrids (Desvignes 1992, 1999b; Desvignes and Lecocq 1995). However, all surveyed symptomatic plants in France were found to harbor the virus, whereas it was detected in 52 of 60 (87%) tested symptomatic Italian trees. The failure to detect ChMV in eight symptomatic Italian trees might reflect sequence variability and an incomplete inclusiveness of the PCR primers used or low or uneven virus accumulation. Using biological indexing, an uneven distribution of the ChMD agent in infected trees has been found, leading to the failure to detect it in parts of some infected trees (Desvignes 1999b; Desvignes and Lecocq 1995).

Even though Koch's postulates were not fully verified, the experiments reported here make a very strong case for the role of ChMV as the causal agent of chestnut mosaic disease. The low ChMV diversity observed in France and Italy is consistent with the scenario of its recent introduction in Europe (Desvignes and Lecocq 1995), whereas the genetic separation of the Italian and French clusters is suggestive of separate introduction events. These results and the associated development of molecular tools for the detection of ChMV will help speed the selection of virus-free mother plants and mitigate the virus spread in new chestnut orchards and layerings. However, many questions remain regarding the variability of symptom intensity in relation to cultivar susceptibility, ChMV-induced graft incompatibility, the impact of pedoclimatic conditions and of synergic and competitive interferences with other chestnut pathogens, and silvicultural management.

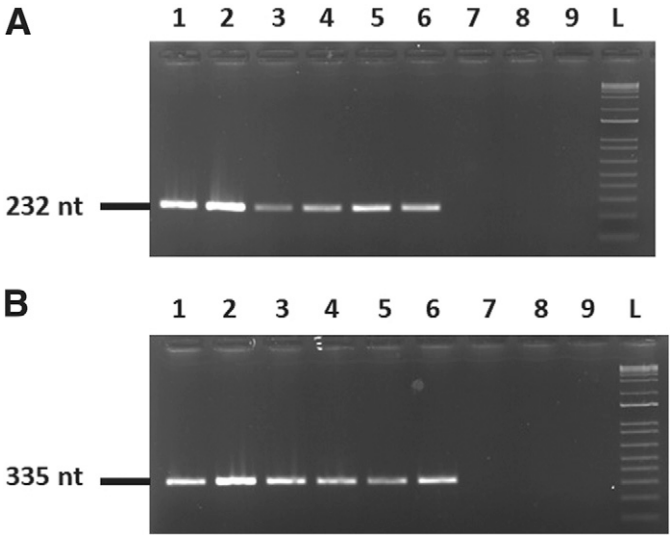

Fig. 5. Detection of chestnut mosaic virus in various samples by polymerase chain reaction (PCR) using primer pairs **A,** Ch-Bad1466F/1800R and **B,** Ch-Bad5860F/6109R. Lane 1: LC1224F. Lane 2: LC1224H. Lane 3: FRlc1224A. Lane 4: T32018. Lane 5: LCA552. Lane 6: LCA584. Lane 7: 'Maraval' Ca 74 noninoculated plant. Lane 8: *Quercus rubra* noninoculated plant. Lane 9: no template. Lane L: molecular weight marker. Horizontal bars on the left of the figure indicate the size of the amplification products. The isolates are listed in Supplementary Table S1.

ACKNOWLEDGMENTS

We thank the INRAE Vine Experimental Unit of Bordeaux (UEVB) and X. Capdevielle for taking care of chestnut orchards; Ascenzio Santini (Azienda Agricola Laga Nord, Umito, Aquasanta Terme, Italy) for the essential support during phytosanitary monitoring in chestnut orchards; D. Cornaggia and P. Gentit for the information regarding the past work of CTIFL, the characterization of the symptoms of the disease on biological indicators, and the conservation of the isolates; and Eden Darnige for proofreading the manuscript.

LITERATURE CITED

Antonaroli, R., and Perna, M. R. 2000. Una fitopatia ad eziologia ancora incerta: il giallume dl castagno in Emilia Romagna e nelle Marche. Sherwood 6:43-46.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 9:455-477.

Bhat, A. I., Hohn, T., and Selvarajan, R. 2016. Badnaviruses: The current global scenario. Viruses 8:177.

Bradford Hill, A. 1965. The environment and disease: Association or causation? Proc. R. Soc. Med. 58:295-300.

Conedera, M., Tinner, W., Krebs, P., de Rigo, D., and Caudullo, G. 2016. *Castanea sativa* in Europe: distribution, habitat, usage and threats. Pages 78-79 in: European Atlas of Forest Tree Species. J. San-Miguel-Ayanz, D. de Rigo, G. Caudullo, T. Houston Durrant, and A. Mauri, eds. Publ. Off. Eu, Luxembourg.

Deng, S., and Hiruki, C. 1991. Amplification of 16S rRNA genes from culturable and non-culturable mollicutes. J. Microbiol. Methods 14:53-61.

Desvignes, J. C. 1992. Characterization of the chestnut mosaic. Acta Hortic. 309:353-358.

Desvignes, J. C. 1999a. Mosaïque du chataignier. Pages 179-181 in: Maladies à Virus des Arbres Fruitiers. CTIFL, France.

Desvignes, J. C. 1999b. Sweet chestnut incompatibility and mosaics caused by the chestnut mosaic virus (ChMV). Acta Hortic. 494:451-454.

Desvignes, J. C., and Cornaggia, D. 1996. Mosaïque du chataignier: Transmission par le puceron *Myzocallis castanicola.* Phytoma 481:39-41.

Desvignes, J. C., and Lecocq, G. 1995. New knowledges on the chestnut mosaic virus disease. Acta Hortic. 386:578-584.

Doyle, J. J., and Doyle, J. L. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Fox, A. 2020. Reconsidering causal association in plant virology. Plant Pathol. 69:956-961.

Gambino, G., Perrone, I., and Gribaudo, I. 2008. A rapid and effective method for RNA extraction from different tissues of grapevine and other woody plants. Phytochem. Anal. 19:520-525.

Geering, A. D. W., Maumus, F., Copetti, D., Choisne, N., Zwickl, D. J., Zytnicki, M., McTaggart, A. R., Scalabrin, S., Vezzulli, S., Wing, R. A., Quesneville, H., and Teycheney, P. Y. 2014. Endogenous florendoviruses are major components of plant genomes and hallmarks of virus evolution. Nat. Commun. 5:5269.

Gualaccini, G. 1958. Una virosi nuova del castagno. Boll. Staz Patol. Veg. Roma 16:67-75.

Gundersen, D. E., and Lee, I. M. 1996. Ultrasensitive detection of phytoplasmas by nested-PCR assays using two universal primer pairs. Phytopathol. Mediterr. 35:144-151.

Hohn, T., and Rothnie, H. 2013. Plant pararetroviruses: Replication and expression. Curr. Opin. Virol. 3:621-628.

Horvath, J., Ecke, I., Gal, T., and Dezcery, M. 1975. Demonstration of virus-like particles in sweet chestnut and oak with leaf deformations in Hungary. Z. Pflkrankh. Pfsdrntz. 82:498-502.
Kreuze, J. F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., and Barker, I. 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. Virology 388:1-7.
Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA 7: Molecular Evolutionary Genetics Analysis version 7 for bigger datasets. Mol. Biol. Evol. 33:1870-1874.
Martin, R. R., Constable, F., and Tzanetakis, I. E. 2016. Quarantine regulations and the impact of modern detection methods. Annu. Rev. Phytopathol. 54:189-205.
Massart, S., Candresse, T., Gil, J., Lacomme, C., Predajna, L., Ravnikar, M., Reynard, J. S., Rumbou, A., Saldarelli, P., Škorić, D., Vainio, E. J., Valkonen, J. P. T., Vanderschuren, H., Varveri, C., and Wetzel, T. 2017. A framework for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. Front. Microbiol. 8:45.
Massart, S., Olmos, A., Jijakli, H., and Candresse, T. 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Res. 188:90-96.
Murolo, S., De Miccolis Angelini, R. M., Faretra, F., and Romanazzi, G. 2018. Phenotypic and molecular investigations on hypovirulent *Cryphonectria parasitica* in Italy. Plant Dis. 102:540-545.
Prospero, S., Vannini, A., and Vettraino, A. M. 2012. *Phytophthora* on *Castanea sativa* Mill. (sweet chestnut). JKI Data Sheets 6.
Ragozzino, A., and Lahoz, E. 1986. Una malattia virus-simile del castagno in provincia di Avellino. Giornate di studio sul castagno. Soc. Orti. Italiana 307-311.
Reid, K. E., Olsson, N., Schlosser, J., Peng, F., and Lund, S. T. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol. 6:27-37.
Rigling, D., and Prospero, S. 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: invasion history, population biology and disease control. Mol. Plant Pathol. 19:7-20.
Rott, M., Xiang, Y., Boyes, I., Belton, M., Saeed, H., Kesanakurti, P., Hayes, S., Lawrence, T., Birch, C., Bhagwat, B., and Rast, H. 2017. Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. Plant Dis. 101:1489-1499.
Shimada, S. 1962. Chestnut yellows. Plant Prot. Tokyo 16:253-254.
Smart, C. D., Schneider, B., Blomquist, C. L., Guerra, L. J., Harrison, N. A., Ahrens, U., Lorenz, K. H., Seemuller, E., and Kirkpatrick, B. C. 1996. Phytoplasma-specific PCR primers based on sequences of 16s-23s rRNA spacer region. Appl. Environ. Microbiol. 62:2988-2993.
Staginnus, C., Iskra-Caruana, M. L., Lockhart, B., Hohn, T., and Richert-Pöggeler, K. R. 2009. Suggestions for a nomenclature of endogenous pararetroviral sequences in plants. Arch. Virol. 154:1189-1193.
Teycheney, P. Y., Geering, A. D. W., Dasgupta, I., Hull, R., Kreuze, J. F., Lockhart, B., Muller, E., Olszewski, N., Pappu, H., Pooggin, M. M., Richert-Pöggeler, K. R., Schoelz, J. E., Seal, S., Stavolone, L., and Umber, M., and ICTV Report Consortium. 2020. ICTV virus taxonomy profile: *Caulimoviridae*. J. Gen. Virol. 101:1025-1026.
Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap, penalties and weight matrix choice. Nucleic Acids Res. 22:4673-4680.
Vettraino, M., Vannini, A., Flamini, L., Lagnese, R., Pizzichini, L., Talevi, S., and Fulbright, D. W. 2005. A new transmissible symptomology on sweet chestnut in Italy. Acta Hortic. 693:547-550.
Villamor, D. E. V., Ho, T., Al Rwahnih, M., Martin, R. R., and Tzanetakis, I. E. 2019. High throughput sequencing for plant virus detection and discovery. Phytopathology 109:716-725.
Wang, Y., Cheng, X., Wu, X., Wang, A., and Wu, X. 2014. Characterization of complete genome and small RNA profile of pagoda yellow mosaic associated virus, a novel badnavirus in China. Virus Res. 188:103-108.
Zerbino, D. R., and Birney, E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.