Citation:

Welgemoed T, Pierneef R, Read DA, Schulze SE, Pietersen G, Berger DK (2020) **Next generation sequencing reveals past and current widespread occurrence of maize yellow mosaic virus in South Africa** Eur J Plant Pathol 158 (1):237-249. doi:10.1007/s10658-020-02070-1

Next generation sequencing reveals past and current widespread occurrence of maize yellow mosaic virus in South Africa

Tanya Welgemoed [1,2,3], Rian Pierneef [1,3,4], David A. Read [4], Susanna E. Schulze [2,3], Gerhard Pietersen [2,3,5] and Dave K. Berger [2,6*]

[1]Centre for Bioinformatics and Computational Biology, [2]Forestry and Agricultural Biotechnology Institute (FABI), [3]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

[4]Agricultural Research Council – Biotechnology Platform, Pretoria, 0110, South Africa

[5]Department of Genetics, Stellenbosch University, Stellenbosch, 7600, South Africa

[6]Department of Plant and Soil Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

*Correspondence: dave.berger@fabi.up.ac.za (orcid: 0000-0003-0634-1407)

Abstract

Maize yellow mosaic virus (MaYMV) is a single-stranded RNA polerovirus first identified in China. MaYMV was recently reported from West and East Africa, however it had not yet been reported from southern Africa. RNA-seq data from South African field-grown fungal-infected maize was mined for viral sequences by *de novo* assembly of reads that did not map to the maize or fungal genomes. Predicted proteins from the *de novo*-assembled unmapped reads matched MaYMV proteins with regions of 96-100% identity. MaYMV was detected in maize RNAseq data from 2009, 2012 and 2013. Complete South African MaYMV genome sequences (5642 nt) were determined by RT-PCR and Sanger sequencing of samples from two different maize genotypes, years, and sites. Phylogenetic analysis confirmed the species identity as MaYMV, and showed separate clustering of isolates between Africa, Asia and South America. Some MaYMV positive samples had reads matching Potyviridae (Johnson grass mosaic virus and Sugarcane mosaic virus), and mycoviruses (Setosphaeria turcica hypovirus 1, Bipolaris maydis partitivirus 1, and Pleospora typhicola fusarivirus 1). A 2016/2017 RT-PCR survey of maize plants exhibiting virus-like symptoms, such as yellowing and streaking patterns, revealed MaYMV in 39 samples from six provinces in South Africa. This report documents the earliest known MaYMV infection world-wide, and indicates that the virus is widespread throughout Africa.

Keywords

MaYMV; maize yellow mosaic virus; polerovirus; mycovirus; MYDV; luteovirus

Introduction

Maize yellow mosaic virus (MaYMV) is the provisional name of a recently characterised maize RNA virus identified in China and placed in the *Polerovirus* genus (Chen et al. 2016). A parallel study in China discovered and sequenced the same virus, which they named Maize yellow dwarf virus (MYDV)-RMV2 (Wang et al. 2016), and other authors have called it MYDV-like polerovirus or MYDV-RMV (Massawe et al. 2018; Wamaitha et al. 2018). We utilise the commonly used MaYMV designation in this study. The *Polerovirus* genus is in the family *Luteoviridae*, and consists of single stranded RNA positive-strand viruses. Poleroviruses are generally aphid-transmitted and phloem limited (Hogenhout et al. 2008). MaYMV has recently been shown to be vectored by aphids causing leaf reddening symptoms in some maize genotypes and asymptomatic effects in others (Stewart et al. 2020).

MaYMV has been identified using next generation sequencing after extraction of RNA from maize leaves collected on three continents – Asia, Africa and South America (Chen et al. 2016; Palanga et al. 2017; Gonçalves et al. 2017). Samples collected in 2013 in the Yunnan and AnHui provinces were the source of the first reports in China (Chen et al. 2016; Wang et al. 2016). MaYMV was reported in maize fields surrounding the Agricultural Research Centre of Kamboinsé in Burkina Faso, West Africa in 2016 (Palanga et al. 2017). In East Africa, MaYMV was present in maize samples collected from 2012-2014 in Kenya (Wamaitha et al. 2018), from 2013-2016 in Kenya, Tanzania, Rwanda and Ethiopia (Massawe et al. 2018), from 2015 and 2016 in Ethiopia (Guadie et al. 2018) and from Tanzania in 2015 (Read et al. 2019). MaYMV was reported from Sao Paulo state in Brazil (2013 and 2017) and Los Rios region of Ecuador (2016) in South America (Gonçalves et al. 2017; Bernreiter et al. 2017).

MaYMV has also been reported from other grasses, namely sugarcane and the itch weed *Rottboellia cochinchinensis* in the northern Guinea savannah region of Nigeria (Yahaya et al. 2017), and from *Panicum milliaceum* in South Korea, and sorghum in Kenya and South Korea (Lim et al. 2017; Wamaitha et al. 2018). These reports suggest that the occurrence of the virus is relatively widespread, although to our knowledge there have not been any reports of the virus from southern Africa. Currently no data is available on the impact of MaYMV on agricultural yields.

The current study was initiated after the detection of MaYMV in South African grown maize plants subjected to Illumina RNA sequencing. The original aim of the RNA sequencing was to identify novel genes present in near-isogenic lines harbouring a QTL (quantitative trait locus) for resistance to the foliar disease of maize, grey leaf spot (GLS), caused by the fungus *Cercospora zeina* (Welgemoed et al. 2020). We confirmed the presence of MaYMV in these and other samples collected between 2009 and 2017 using RNA-seq and/or with a virus specific reverse transcriptase polymerase chain reaction (RT-PCR) assay, and demonstrate the widespread occurrence of this virus in South African maize.

Methods

Maize material and RNA sequencing

The source material for initial detection of MaYMV was three plants of maize (*Zea mays*) inbred line B73 exposed to natural inoculum pressure of the fungal pathogen *C. zeina* in a field trial held in 2013 at Greytown, Kwa-Zulu Natal province, South Africa. This was the same material, RNA extraction and RNA sequencing (RNA-seq) protocol on an Illumina platform that was described in the Methods S1 file of (Christie et al. 2017). Two samples per plant were collected named BR1A, BR1B, BR2A, BR2B, BR3A and BR3B. Aliquots of RNA from these samples were stored at -80°C, and were used subsequently for RT-PCR confirmation of the presence of MaYMV.

Additional maize RNA-seq data was obtained from parental inbred lines CML444 and SC Malawi, and 14 of their recombinant inbred line (RIL) progeny sampled in 2009 from a field trial at Baynesfield Estate, KwaZulu-Natal (described in (Christie et al. 2017). RNA-seq data was also screened for MaYMV from SC Malawi plants grown at the same site in 2012.

Maize leaf material with putative symptoms of virus infection (e.g. mosaics, mottles, yellow streaking) was collected from maize fields during a survey carried out in the 2016/2017 maize growing season in the following provinces of South Africa: Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North West, and Northern Cape. A total of 125 leaf samples were collected and a selection of 44 representing each province was screened for MaYMV by RT-PCR.

De novo assembly bioinformatics pipeline to identify virus sequences in RNA-seq data

A bioinformatics pipeline was developed to exclude genes that mapped to genomes of maize inbred line B73 or the fungal pathogen C. zeina, and to identify novel genes present in the samples (Online Resource 1). The RNA-seq reads from maize inbred B73 leaf samples were processed using the following protocols. Raw reads were trimmed to remove (i) the first 15 nt containing the random primers and (ii) bases with a quality value less than 20 in a 4 nt sliding window with Trimmomatic (Bolger et al. 2014) based on the FastQC (Andrews 2016) results. The trimmed reads were mapped against the annotated maize B73 genome (Zm-B73-REFERENCE-GRAMENE-4.0; (Jiao et al. 2017)) and the C. zeina genome (Wingfield et al. 2017) simultaneously using TopHat (Kim et al. 2013). Unmapped reads were retained and subjected to strand-specific de novo assembly using a k-mer size of 31 with Trinity (Haas et al. 2013). Protein coding sequences were predicted from the assembled transcripts using TransDecoder with a minimum size of 50 amino acids (Haas et al. 2013). The predicted proteins were then annotated by BLASTP comparison against the NCBI nr (non-redundant) database (NCBI_Resource_Coordinators 2017).

Detection of MaYMV in RNA-seq data by mapping to a MaYMV genome

After initial discovery of MaYMV in the 2013 samples using the pipeline (Online Resource 1), the presence of MaYMV in the 2009, 2012 and 2013 samples was evaluated by mapping the trimmed RNA-seq reads against a MaYMV genome (KU248489) using Bowtie2 with the default settings (Langmead and Salzberg 2012).

MaYMV genome assemblies of South African isolates using RT-PCR, RACE and Sanger sequencing

Two representative MaYMV genomes, named MaYMV RSA BR1A and MaYMV RSA SCM, were assembled and sequenced from RNA samples extracted from a 2013 B73 plant from Greytown and a 2012 SC Malawi plant from Baynesfield, respectively. Direct Sanger sequencing of the MaYMV RSA genomes was carried out by amplifying overlapping RT-PCR segments of the genomes (fragments 1-5), using the primers listed in Online Resource 2. cDNA was synthesized through the addition of 0.75µl of each respective reverse primer (10µM) and 3.75µl nuclease-free water to 1µl of total RNA (100 ng/µl). Denaturation was then performed by heating the mixture to 70°C for 5 minutes, followed by chilling for 5 minutes. 2.5µl of denatured RNA/primer was then added to 7.5µl of reverse-transcriptase (RT) mix, containing 0.5µl GoScript RT, 0.05µl RNasin RNase Inhibitor, 1µl dNTP mix (10mM each), 0.6µl $MgCl_2$ (25mM) and 2µl GoScript 5X RT buffer (Promega, Madison, WI, USA) with nuclease-free water up to 7.5µl. Reactions were held at 42°C for 1 hour. PCR reactions were setup using OneTaq DNA polymerase (New England Biolabs, Ipswich, MA, USA), according to the manufacturers protocol. PCR conditions consisted of initial denaturation at 95°C for 3 min followed by 35 cycles of 95°C for 30s, Tm – 5°C for 30s and 68°C for 1 min, with a final extension step at 68°C for 10 min. PCR products for genome fragments 1 to 5 were enzymatically purified by adding 2µl of Fast alkaline phosphatase and 0.5µl of exonuclease I (Thermo Scientific, Vilnius, Lithuania) to 19 µl of amplicon. The purified products were then used as templates in bi-directional Sanger dye-terminator reactions, performed by Inqaba Biotechnical industries, Pretoria, South Africa.

Rapid amplification of cDNA ends (RACE) confirmation of the 5’ and 3’ terminal nucleotides was performed using the 5’ RACE System for Rapid Amplification of cDNA Ends Version 2.0 and the 3’ RACE System for Rapid Amplification of cDNA Ends (Life Technologies, Carlsbad, CA, USA), according to the manufacturers specifications, using gene-specific primers (GSP) listed in Online Resource 2. Single bands were excised from agarose gels and purified using a NucleoSpin Gel and PCR clean up kit (Macherey-Nagel, Düren, Germany). RACE amplicons were then inserted into pGEM-T Easy (Promega, Madison, WI, USA), prior to Sanger sequencing with the T7/SP6 primer pair. The assembly of the resulting Sanger

sequences was carried out using the CAP contig assembly accessory program in BioEdit 7.2.5 (Hall 1999). The assembled MaYMV genomes were annotated using CLC Main Workbench v8.01 by identification of open reading frames and frameshifts described for the MaYMV Yunnan 11 isolate (KU248489).

Phylogenetic analysis of MaYMV sequences

Whole genome sequences of MaYMV isolates and related Poleroviruses from GenBank, and RNA-dependent RNA polymerase gene sequences from these genomes were aligned using MAFFT (Katoh and Standley 2013). Phylogenetic trees were constructed in RAxML(Stamatakis 2014) using the Maximum Likelihood method (boot-strapped 1000 times). Maize yellow dwarf virus-RMV (KC921392) and Barley virus G isolate Gimje (KT962089) were used as outgroups.

Viromics analysis of maize leaf samples

Metagenomic analysis to determine the virus content in maize leaf sample RNAseq data was carried out with Kaiju (Menzel et al. 2016), by looking for matches in a subset of the NCBI nr database containing all proteins belonging to Archaea, Bacteria and Viruses.

RT-PCR to identify MaYMV in survey samples

RT-PCR was used to confirm the presence of MaYMV in the RNA-seq samples and to survey South African maize growing areas for MaYMV. RNA from the same samples that had been used for RNA-seq and had been stored at -80 °C, as well as RNA extracted from fresh material collected during the 2017 maize virus survey was used. RNA was extracted using the CTAB (cetryltrimethylammonium bromide) method (White et al. 2008).

RT-PCR with the primer pair MaYMV-F and MaYMV-R (Chen et al. 2016) was carried out using the OneTaq® One-Step RT-PCR Kit (New England BioLabs, Ipswich, USA). The RT-PCR reactions were set up in a total volume of 25 µl, which consisted of: 2× Quick-Load One Taq One-Step Reaction Mix, 0.5 µm of each primer, 1 µl of 25x One Taq One-Step Enzyme Mix, 1 µl of the RNA (100 ng/µl), and sterile distilled water. The cycling conditions were as follows: 15 min at 48 °C, 1 min at 94 °C followed by 40 cycles of 15 s at 94 °C, 30 s at 58 °C, and 45 s at 68 °C, and a final extension step of 68 °C for 5 min. PCR products were separated by agarose gel electrophoresis, and visualised after staining with ethidium bromide under ultra-violet light using the Molecular Imager® Gel Doc™ XR System equipped with the Image Lab™ software Ver 6.0 (Bio-rad Laboratories Inc., Hercules, CA, USA).

The positive control for RNA and cDNA quality of MaYMV-negative samples involved RT-PCR with primers actinF and actinR (Van den Berg et al. 2004) that flank an intron in the maize actin gene. The RT-PCR protocol was the same as used for the MaYMV primers.

Selected RT-PCR products were Sanger sequenced using the BigDye® reagents, and processed on an ABI 3500xl Genetic Analyser (Applied Biosystems, Foster City, USA) using the methods described in (Swart et al. 2017). Prior to sequencing, single stranded DNA was removed from RT-PCR reactions as follows: 0.5µl of 10 U/µl exonuclease I (ThermoFisher, Waltham, USA) and 2µl of 2U/µl FastAP® (ThermoFisher, Waltham, USA) was added to 19µl amplification products, and incubated at 37°C for 1h.

Results

Identification of MaYMV from RNAseq data from field-grown maize

A bioinformatics pipeline was developed to identify novel maize protein-coding transcripts from RNA-seq data of maize inbred line B73 introgressed with a QTL for resistance to the foliar fungal pathogen *Cercospora zeina* (Welgemoed et al. 2020).

In this study, the RNA-seq data from the control B73 plants from this 2013 field trial at Greytown in KwaZulu-Natal province of South Africa were processed through the pipeline, which involved (i) recovery of reads that did not map to either the maize or fungal genome, (ii) *de novo* assembly of these reads into transcripts, (iii) conceptual translation of the transcripts into protein sequences, and (iv) BLASTP analysis against the NCBI nr database (Online Resource 1).

Analysis of the data from the B73 leaf samples showed positive BLASTP matches to MaYMV proteins (MaYMV RNA dependent RNA polymerase, P3 coat protein, and P3-P5 readthrough protein) with more than 96% identity and E-values less than $3.1E^{-43}$ (data not shown). Subsequently, the RNA-seq reads from these three plants were mapped to one of the first reported genomes of MaYMV from China (Yunnan 11; KU248489)(Chen et al. 2016). Partial MaYMV genome sequences were assembled from each plant with 5X coverage over more than 70% of the 5,642 nt MaYMV genome (data not shown).

RNA-seq data from maize field trials held in 2009 and 2012 at Baynesfield Estate in KwaZulu-Natal province of South Africa were also assessed. This site is 93 km from the Greytown site where the 2013 samples were collected. Reads corresponding to MaYMV were detected in 11 out of 14 maize plant samples from the 2009 field trial, and all six of the 2012 maize plants sampled (data not shown). Partial MaYMV genome sequences were obtained from most of the plants. The best MaYMV assemblies from individual 2009 and 2012 plants made up 47% and 89% of the 5,642 nt genome, respectively, at more than 5X coverage (data not shown).

Assembly and analysis of MaYMV genome

Full-length representative genome sequences of MaYMV from South Africa were obtained by RT-PCR (including 5'RACE and 3'RACE) and Sanger sequencing. These were from the maize

B73 BR1A sample (2013) and a SC Malawi sample (2012) collected at Greytown and Baynesfield in KwaZulu-Natal Province, respectively. The MaYMV RSA BR1A genome (5,643 nt; MG570476) was 98.5% identical to the MaYMV RSA SCM genome (5,644 nt; MN943641) sequence (Online Resource 3). Both genomes encoded all the expected full-length proteins with more than 95% identity to corresponding proteins from MaYMV Yunnan 11 (P0, P1-P2, P3, P4, P3-P5) (Online Resource 4).

Phylogenetic analysis with whole genome sequences of the two South African isolates RSA BR1A and RSA SCM confirmed their identity as MaYMV (Fig. 1). Whole virus genome alignments used to produce the phylogenetic tree showed that the South African isolates had highest identity to all East African isolates from Kenya, Tanzania and Ethiopia listed in Online Resource 3 (98.3 - 99.6% identity), followed by Nigerian and all listed Chinese isolates (97% identity), and South American isolates from Brazil and Ecuador (93 - 95% identity), and were distinct from the outgroup species Maize Yellow Dwarf Virus-RMV and Barley Virus G isolate Gimje (Fig. 1)(Online Resource 3). Sub-clades corresponding to different continents are evident in both the whole genome and RNA-dependent RNA polymerase gene phylogenetic trees (Fig. 1)(Online Resource 5).

Survey of maize growing regions for MaYMV using an RT-PCR assay

RT-PCR with primers diagnostic for MaYMV (Chen et al. 2016) (within the RNA dependent RNA polymerase and coat protein genes) produced the diagnostic 753 bp MaYMV product from stored RNA corresponding to maize B73 samples BR1A and BR1B (Fig. 2a). Sanger sequencing of the BR1B sample confirmed that the expected MaYMV fragment had been amplified (Online Resource 6).

Maize leaf samples exhibiting virus-like symptoms (mosaic, mottling, yellow streaks; Online Resource 7) were collected during the 2016/2017 season from the main maize producing provinces of South Africa (Fig. 3), and 44 samples were screened for MaYMV by RT-PCR (Table 1). A total of 39 samples were RT-PCR positive for MaYMV (Table 1), representing isolates from all provinces sampled, namely Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, and North West (Fig. 2b).

RT-PCR of MaYMV negative samples targeting a maize actin gene confirmed that the cDNA was intact and that these were true negatives (Online Resource 8). Sanger sequencing of RT-PCR products of samples from different provinces in the survey (17-4010: Mpumalanga, 17-4101: Gauteng, 17-4172: Gauteng, 17-4203: Mpumalanga and 16-3379: North West) confirmed the expected MaYMV product (data not shown).

Viromics analysis of maize leaves containing MaYMV

The maize B73 samples from Baynesfield 2013 in which MaYMV was first identified, including the sample used to obtain the MaYMV RSA BR1A genome by Sanger sequencing, did not show evident virus-like symptoms (data not shown). However, viromics analysis was carried out to determine whether there were other viruses present in these samples. MaYMV was confirmed to be present in the RNAseq reads of these samples (BR1A-BR3A and BR1B-BR3B), although it only made up a small proportion of the RNA virus reads (ranging from 0.01% - 15%)(Fig. 4) (Online Resource 9). RNA virus content ranged from 693 - 439,942 reads in these samples (Online Resource 9). Potyviruses contributed a large proportion of RNA virus reads in some samples: Sugarcane mosaic virus (SCMV) made up 51% of reads in BR1A, and Johnsongrass mosaic virus (JGMV) made up 81 and 83% of reads in samples BR3A and BR3B, respectively. Mycoviruses made up 58% and 62% of reads in the lower leaf fungal-infected samples that did not contain large amounts of potyvirus reads (Fig. 4) (Online Resource 9). Most of the mycovirus reads corresponded to viruses of maize foliar pathogens or endophytes, namely Setosphaeria turcica hypovirus 1, Bipolaris maydis partitivirus 1, and Pleospora typhicola fusarivirus 1 (Online Resource 9).

Discussion

This study presents the first report of MaYMV in maize in South Africa, and demonstrated that it was present in the KwaZulu-Natal province from at least 2009, which represents the oldest report for Africa. Furthermore, the virus currently appears to be widespread throughout the main maize growing regions of the country, since maize leaf samples from all the six tested provinces Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, and North West were RT-PCR positive for MaYMV.

Next generation sequencing followed by a range of bioinformatics tools is increasingly being employed for diagnostics of known viruses and detection of unknown viruses in plant samples (Jones et al. 2017). The “discovery” phase of our study employed a variation from most previous approaches that map nucleic acid sequences (reads or assembled contigs) to databases for virus detection (Massawe et al. 2018; Visser et al. 2016). Our bioinformatics pipeline was designed to identify novel protein-coding genes in maize, with the following key steps: (i) recovery of RNA-seq reads that did not map to the host genome; and (ii) assembly of these reads into RNA contigs which were translated into predicted proteins and searched against the NCBI database. Matches to MaYMV proteins were obtained at high confidence. This approach was similar to modules of the pipeline “VirFind” developed by (Ho and Tzanetakis 2014), which the authors demonstrated by identifying known and novel viruses in more than 30 plant species.

Subsequent to the discovery phase, we further exploited maize RNA-seq data from different South African field trials collected in different seasons (2009, 2012, 2013) to detect

MaYMV by mapping the quality-filtered RNA-seq reads to the first MaYMV genome described from China. Two South African full-length representative genomes, MaYMV RSA BR1A and MaYMV RSA SCM, were determined by RT-PCR and Sanger sequencing. Phylogenetic analysis confirmed the species identity as MaYMV, and annotation confirmed coding capacity for the expected full-length proteins (P0 silencing suppressor; P1-P2 RNA dependent RNA polymerase, P3 coat protein, P4 movement protein, P3-P5 readthrough protein). Prior to this study, Sanger sequencing confirmation has only been conducted for MaYMV isolates from China (Chen et al. 2016; Wang et al. 2016) and one isolate from Kenya (Massawe et al. 2018), whereas all the other MaYMV genome sequences have been constructed from the more error-prone Illumina sequencing data (Wamaitha et al. 2018; Read et al. 2019; Guadie et al. 2018; Gonçalves et al. 2017; Bernreiter et al. 2017).

MaYMV has now been reported from Asia, Africa and South America (Chen et al. 2016; Yahaya et al. 2017; Gonçalves et al. 2017). Phylogenetic analysis of available MaYMV genome sequences including the two South African genomes from this study sheds some light on the possible movement and resultant distribution of the virus. MaYMV isolates for which whole genome sequences are available appear to group in continent-specific clades (Fig. 1). For example, MaYMV from South Africa groups with Kenyan, Tanzanian, and Ethiopian isolates (98-99% similarity), distinct from Chinese isolates (97% similarity), and South American isolates (93-95% similarity)(Online Resource 3). This indicates that MaYMV has been isolated for a sufficient period of time on each continent to develop continent-specific haplotypes, indicating that the virus may have present earlier than the first collection dates (2009 in Africa; this study; 2013 in China and Brazil (Chen et al. 2016; Wang et al. 2016; Gonçalves et al. 2017)).

Several reasons may explain why MaYMV has only recently been identified. Symptoms associated with MaYMV were not known until very recently. (Stewart et al. 2020) used serial aphid transmission to obtain MaYMV infected plants free of other viruses. They showed MaYMV infection resulted in maize leaf reddening symptoms, and that the virus could be transmitted by two aphid species, *Rhopalosiphum padi* and *Rhopalosiphum maidis* (Stewart et al. 2020). Previous “symptoms” reported for plants containing MaYMV genomes were different versions of yellow streaking, which may be due to co-infecting viruses (Palanga et al. 2017; Wamaitha et al. 2018; Chen et al. 2016). We also observed yellow streaking in MaYMV-positive maize in our survey, which may also be due to other viruses (Online Resource 7). However, it remains possible that yellowing of leaves may be a symptom of MaYMV infection in certain maize genotypes, since other poleroviruses can also cause yellowing of leaves (Grisham et al. 2010).

MaYMV has been found together with several other viruses as illustrated in a metagenomics study aimed at characterizing viruses associated with maize lethal necrosis

disease (MLND) (Wamaitha et al. 2018). This was borne out by our viromics analysis of the maize inbred B73 samples that harboured MaYMV, some of which had high levels of potyvirus reads (sugarcane mosaic virus and johnsongrass mosaic virus)(Fig. 4). Another reason for lack of MaYMV identification until recently may be that large scale screening with ELISA using antibodies to poleroviruses (such as BYDV-RMV; (Wang et al. 2016)) or RT-PCR with degenerate universal polerovirus primers (Palanga et al. 2017) may not have been carried out in many of these maize production regions. Finally, the advent of next generation sequencing has afforded a more comprehensive analysis of virus diversity in maize samples. The recent epidemic of MLND in Africa has precipitated several studies where MaYMV has been found by mining data from MLND surveys (Wamaitha et al. 2018; Massawe et al. 2018) (this study).

The results of our RT-PCR survey of 2016/2017 maize samples from six different maize producing provinces in South Africa indicates that MaYMV is widespread in the country in many different agro-ecological zones, indicating that it may have been introduced sometime prior to our first identification in RNA-seq data from a 2009 field trial. MaYMV was also shown to be widely distributed in Ethiopia (Guadie et al. 2018), Kenya (Massawe et al. 2018; Wamaitha et al. 2018), and Tanzania (Read et al. 2019). Despite the widespread distribution of MaYMV in South Africa, it does not seem to be a major threat to maize production locally, based on observations during the 2017 field survey. Manual scouting in maize fields revealed only a few scattered plants with mosaic or yellow streaking symptoms, and there were no cases of large-scale symptoms (data not shown). Leaf reddening symptoms, shown recently to be caused by MaYMV (Stewart et al. 2020) were not evident in our study. The lack of these symptoms in the B73 and SC Malawi plants from which the whole MaYMV genomes were assembled may be due the low virus titre or suppression of leaf reddening in these maize genotypes.

The causal routes of introduction and spread of MaYMV into maize producing countries such as South Africa are not known, however three possible routes could be considered. First, if it is shown to be a seed-borne virus then it could have spread as a result of the worldwide distribution of maize seed (Ranum et al. 2014). If not seed-borne, it could also have reached Africa within leaves of whole plants that were introduced into West Africa from the America's as part of one of the initial introduction routes of maize into Africa (McCann, 2005). Second, MaYMV is likely to be spread locally by aphid vectors since this has now been demonstrated for two *Rhapalosiphum spp.* (Stewart et al. 2020). *R. maidis* and other aphids are widespread in maize production areas throughout Africa (Hatting et al. 2000; Stewart et al. 2020). Third, MaYMV has been reported from alternative hosts, such as sugarcane and itchgrass in Nigeria (Yahaya et al. 2017), and sorghum in Kenya (Wamaitha et al. 2018) and South Korea (Lim et al. 2017). Sugarcane and sorghum are also known to be hosts of aphids (Singh et al. 2004) and therefore could serve as reservoirs of the virus. The host range of the sugarcane aphid,

379 *Melanaphis sacchari,* includes other commercial Poaceae such as maize and sorghum (Singh
380 et al. 2004), which makes it another potential vector.
381 Our study did not reveal any relationship between MaYMV presence and the fungal
382 pathogen *C. zeina* in maize, however an intriguing finding was the presence of mycovirus
383 reads in the B73 leaf samples. These corresponded to a ssRNA fusarivirus from *Pleospora*
384 *typhicola*, a dsRNA partitivirus from *Bipolaris maydis*; and a ssRNA hypovirus from
385 *Setosphaeria turcica*. All three of these hosts are members of the Dothidiomycete class of
386 fungi which contains many plant pathogens, including *C. zeina*. Consistent with higher levels
387 of *C. zeina* infection in the lower leaf samples of B73 plants (BR1B, BR2B, BR3B)(see
388 Methods S1 file of (Christie et al. 2017)), there were more mycovirus reads in these samples
389 than upper leaves (BR1A, BR2A, BR3A). In addition, symptoms of northern corn leaf blight
390 (caused by *Setosphaeria turcica*, also known as *Exserohilium turcicum*) were evident in the
391 lower leaves of the B73 plants sampled (data not shown). *Bipolaris maydis* causes southern
392 corn leaf blight, a disease not known to occur on maize in South Africa but it has been reported
393 from West Africa (Aregbesola et al. 2020). The presence of these mycovirus reads may
394 indicate the presence of these fungi or they could represent hits to unknown mycoviruses of
395 Dothidiomycete pathogens or endophytes present in the maize samples. Interestingly, a
396 recent paper that describes data mining of the RNAseq data from a range of fungi reported
397 the presence of Setosphaeria turcica hypovirus in cultures of this fungus grown *in vitro* (Gilbert
398 et al. 2019).
399 Several unanswered questions remain, such as (i) can MaYMV be transferred in the seed;
400 (ii) are field symptoms observed on MaYMV-positive maize caused by MaYMV alone or co-
401 infecting viruses; (iii) is there interaction between MaYMV and other viruses found in infected
402 maize; (iv) what is the influence of MaYMV on the MLND disease complex; (v) what is the role
403 of maize host factors and maize genotype in the expression of symptoms caused by MaYMV;
404 (vi) which aphid vectors are most important in different production regions of Africa; and (vii)
405 what is the impact of alternate hosts?
406 Several of these questions could be addressed by building on progress with testing
407 transmission by different aphid species using maize plants free of MaYMV and other viruses
408 (Stewart et al. 2020). This could be further enhanced by developing a MaYMV infectious clone
409 that is functional in maize. To this end, infectious clones have been developed and tested in
410 *Nicotiana benthamiana*, and although symptoms were not seen in this alternate host, the
411 researchers demonstrated that the MaYMV P0 protein confers silencing suppression activity
412 (Chen et al. 2016; Wang et al. 2018). This suggests that the presence of MaYMV in mixed
413 infections may compromise maize host silencing of other co-infecting viruses, thus enhancing
414 viral disease in the field.
415

Conclusion

This study has illustrated the power of next generation sequencing with libraries derived from total RNA for the identification of RNA viruses from crop field samples. We established that MaYMV is widespread in South Africa in addition to its presence in West Africa (Palanga et al. 2017; Yahaya et al. 2017), and widespread distribution in East Africa (Massawe et al. 2018; Wamaitha et al. 2018; Guadie et al. 2018). Considering the importance of maize in other southern Africa countries such as Zambia, Zimbabwe, Botswana and Mozambique the monitoring for MaYMV in these countries should be an immediate priority.

Acknowledgements and funding

This work was funded by the National Research Foundation, South Africa (grant # 98977), the Department of Agriculture Forestry and Fisheries Research and Technology Fund, South Africa (grant # 98617), and the University of Pretoria, South Africa (Genomics Research Institute grant # A0U945). The funding agencies provided financial support only and had no role in the design, execution, interpretation or documentation of this study. The authors thank P. Tongoona and F. Middleton for assistance with maize field trials, local farmers for access to maize field sites, and M.Greve for assistance with Fig. 3.

Compliance with Ethical Standards

The authors declare no conflicts of interest. All authors consent to the submission of this manuscript. The manuscript has been prepared following principles of ethical and professional conduct. The research did not involve human participants or animal subjects, therefore neither statements concerning informed consent nor welfare of animals is applicable.

Availability of data

All data is available in the manuscript and Electronic Supplementary Material. The MaYMV RSA BR1A and MaYMV RSA SCM genome sequences have been deposited in Genbank (Accessions MG570476; MN943641, respectively). RNA-seq data has been deposited at the NCBI GEO (Gene Expression Omnibus) repository (Accessions GSE94442, GSE99005).

Authors' contributions

Conceived and designed experiments: DKB, GP. Bioinformatics analysis: TW, RP. Sample collection: DKB, SES, GP. Laboratory analyses: TW, SES, DAR. Wrote the manuscript: TW, DKB. Edited the manuscript: GP, RP, SES, DAR.

References

Andrews, S. (2016). FastQC: A quality control application for FastQ data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 09-2018.

Aregbesola, E., Ortega-Beltran, A., Falade, T., Jonathan, G., Hearne, S., & Bandyopadhyay, R. (2020). A detached leaf assay to rapidly screen for resistance of maize to *Bipolaris maydis*, the causal agent of southern corn leaf blight. [journal article]. *European Journal of Plant Pathology, 156*(1), 133-145, doi:10.1007/s10658-019-01870-4.

Bernreiter, A., Garcia Teijeiro, R., Jarrin, D., Garrido, P., & Ramos, L. (2017). First report of Maize yellow mosaic virus infecting maize in Ecuador. *New Disease Reports, 31*, 1, doi:10.5197/j.2044-0588.2017.036.011.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114-2120.

Chen, S., Jiang, G., Wu, J., Liu, Y., Qian, Y., & Zhou, X. (2016). Characterization of a Novel Polerovirus Infecting Maize in China. *Viruses, 8*(5), 120, doi:10.3390/v8050120.

Christie, N., Myburg, A. A., Joubert, F., Murray, S. L., Carstens, M., Lin, Y. C., et al. (2017). Systems genetics reveals a transcriptional network associated with susceptibility in the maize–grey leaf spot pathosystem. *The Plant Journal, 89*(4), 746-763.

Gilbert, K. B., Holcomb, E. E., Allscheid, R. L., & Carrington, J. C. (2019). Hiding in plain sight: New virus genomes discovered via a systematic analysis of fungal public transcriptomes. *PLOS ONE, 14*(7), e0219207, doi:10.1371/journal.pone.0219207.

Gonçalves, M. C., Godinho, M., Alves-Freitas, D. M. T., Varsani, A., & Ribeiro, S. G. (2017). First Report of Maize yellow mosaic virus Infecting Maize in Brazil. *Plant Disease, 101*(12), 2156, doi:10.1094/PDIS-04-17-0569-PDN.

Grisham, M. P., Johnson, R. M., & Zimba, P. V. (2010). Detecting Sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes. *Journal of Virological Methods, 167*(2), 140-145, doi:10.1016/j.jviromet.2010.03.024.

Guadie, D., Abraham, A., Tesfaye, K., Winter, S., Menzel, W., & Knierim, D. (2018). First Report of Maize yellow mosaic virus Infecting Maize (*Zea mays*) in Ethiopia. *Plant Disease, 102*(5), 1044, doi:10.1094/PDIS-08-17-1290-PDN.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols, 8*(8), 1494-1512, doi:10.1038/nprot.2013.084.

Hall (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series, 41*, 95-98.

Hatting, J., Poprawski, T., & Miller, R. (2000). Prevalences of fungal pathogens and other natural enemies of cereal aphids (*Homoptera: Aphididae*) in wheat under dryland and irrigated conditions in South Africa. *BioControl, 45*(2), 179-199.

Ho, T., & Tzanetakis, I. E. (2014). Development of a virus detection and discovery pipeline using next generation sequencing. *Virology, 471*, 54-60.

Hogenhout, S. A., Ammar, E.-D., Whitfield, A. E., & Redinbaugh, M. G. (2008). Insect vector interactions with persistently transmitted viruses. *Annu. Rev. Phytopathol., 46*, 327-359.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017). Improved maize reference genome with single-molecule technologies. [Letter]. *Nature, 546*, 524-527.

Jones, S., Baizan-Edge, A., MacFarlane, S., & Torrance, L. (2017). Viral diagnostics in plants using next generation sequencing: computational analysis in practice. *Frontiers in Plant Science, 8*, 1770, doi:10.3389/fpls.2017.01770.

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular biology and evolution, 30*(4), 772-780, doi:10.1093/molbev/mst010.

503 Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2:
504 accurate alignment of transcriptomes in the presence of insertions, deletions and
505 gene fusions. *Genome biology, 14*(4), R36, doi:10.1186/gb-2013-14-4-r36.
506 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
507 *methods, 9*(4), 357-359, doi:10.1038/nmeth.1923.
508 Lim, S., Yoon, Y., Jang, Y. W., Bae, D. H., Kim, B. S., Maharjan, R., et al. (2017). First
509 Report of Maize yellow mosaic virus Infecting *Panicum miliaceum* and *Sorghum*
510 *bicolor* in South Korea. *Plant Disease, 102*(3), 689, doi:10.1094/PDIS-08-17-1261-
511 PDN.
512 Massawe, D. P., Stewart, L. R., Kamatenesi, J., Asiimwe, T., & Redinbaugh, M. G. (2018).
513 Complete sequence and diversity of a maize-associated *Polerovirus* in East Africa.
514 [journal article]. *Virus Genes, 54*(3), 432-437, doi:10.1007/s11262-018-1560-5.
515 Menzel, P., Ng, K. L., & Krogh, A. (2016). Fast and sensitive taxonomic classification for
516 metagenomics with Kaiju. *Nature Communications, 7*(1), 11257,
517 doi:10.1038/ncomms11257.
518 NCBI_Resource_Coordinators (2017). Database resources of the national center for
519 biotechnology information. *Nucleic acids research, 45*(Database issue), D12.
520 Palanga, E., Longué, R., Koala, M., Néya, J., Traoré, O., Martin, D. P., et al. (2017). First
521 report of maize yellow mosaic virus infecting maize in Burkina Faso. *New Disease*
522 *Reports, 35*, 26, doi:10.5197/j.2044-0588.2017.035.026.
523 Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production,
524 utilization, and consumption. *Annals of the New York Academy of Sciences, 1312*(1),
525 105-112, doi:10.1111/nyas.12396.
526 Read, D. A., Featherstone, J., Rees, D. J. G., Thompson, G. D., Roberts, R., Flett, B. C., et
527 al. (2019). First report of maize yellow mosaic virus (MaYMV) on maize (*Zea mays*)
528 in Tanzania. *Journal of Plant Pathology, 101*(1), 203, doi:10.1007/s42161-018-0152-
529 5.
530 Singh, B., Padmaja, P., & Seetharama, N. (2004). Biology and management of the
531 sugarcane aphid, *Melanaphis sacchari* (Zehntner)(*Homoptera: Aphididae*), in
532 sorghum: a review. *Crop Protection, 23*(9), 739-755.
533 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis
534 of large phylogenies. *Bioinformatics, 30*(9), 1312-1313,
535 doi:10.1093/bioinformatics/btu033.
536 Stewart, L. R., Todd, J. C., Willie, K., Massawe, D. P., & Khatri, N. (2020). A recently
537 discovered maize polerovirus causes leaf reddening symptoms in several maize
538 genotypes and is transmitted by both the corn leaf aphid (*Rhopalosiphum maidis*)
539 and the bird cherry-oat aphid (*Rhopalosiphum padi*). *Plant Disease*, Online First,
540 10.1094/pdis-1009-1019-2054-sc.
541 Swart, V., Crampton, B. G., Ridenour, J. B., Bluhm, B. H., Olivier, N. A., Meyer, J. J. M., et
542 al. (2017). Complementation of CTB7 in the Maize Pathogen *Cercospora zeina*
543 Overcomes the Lack of *In Vitro* Cercosporin Production. *Mol Plant Microbe Interact,*
544 *30*(9), 710-724, doi:10.1094/MPMI-03-17-0054-R.
545 Van den Berg, N., Crampton, B. G., Hein, I., Birch, P. R., & Berger, D. K. (2004). High-
546 throughput screening of suppression subtractive hybridization cDNA libraries using
547 DNA microarray analysis. *Biotechniques, 37*(5), 818-824.
548 Visser, M., Burger, J. T., & Maree, H. J. (2016). Targeted virus detection in next-generation
549 sequencing data using an automated e-probe based approach. *Virology, 495*, 122-
550 128.
551 Wamaitha, M. J., Nigam, D., Maina, S., Stomeo, F., Wangai, A., Njuguna, J. N., et al. (2018).
552 Metagenomic analysis of viruses associated with maize lethal necrosis in Kenya.
553 *Virology Journal, 15*(1), 90, doi:10.1186/s12985-018-0999-2.
554 Wang, F., Zhao, X., Dong, Q., Zhou, B., & Gao, Z. (2018). Characterization of an RNA
555 silencing suppressor encoded by maize yellow dwarf virus-RMV2. [journal article].
556 *Virus Genes, 54*(4), 570-577, doi:10.1007/s11262-018-1565-0.

Wang, F., Zhou, B. G., Gao, Z. L., & Xu, D. F. (2016). A New Species of the Genus *Polerovirus* Causing Symptoms Similar to Maize yellow dwarf virus-RMV of Maize in China. *Plant Disease, 100*(7), 1508-1508, doi:10.1094/PDIS-11-15-1259-PDN.

Welgemoed, T., Pierneef, R., Sterck, L., Van de Peer, Y., Swart, V., Scheepers, K. D., et al. (2020). De novo Assembly of Transcriptomes From a B73 Maize Line Introgressed With a QTL for Resistance to Gray Leaf Spot Disease Reveals a Candidate Allele of a Lectin Receptor-Like Kinase. *Frontiers in Plant Science, 11*(191), doi:10.3389/fpls.2020.00191.

White, E. J., Venter, M., Hiten, N. F., & Burger, J. T. (2008). Modified Cetyltrimethylammonium bromide method improves robustness and versatility: The benchmark for plant RNA extraction. *Biotechnology Journal, 3*(11), 1424-1428.

Wingfield, B. D., Berger, D. K., Steenkamp, E. T., Lim, H.-J., Duong, T. A., Bluhm, B. H., et al. (2017). IMA Genome-F 8 Draft genome of *Cercospora zeina*, *Fusarium pininemorale*, *Hawksworthiomyces lignivorus*, *Huntiella decipiens* and *Ophiostoma ips*. *IMA Fungus, 8*(2), 385–396, doi:10.5598/imafungus.2017.08.02.10.

Yahaya, A., Al Rwahnih, M., Dangora, D., Gregg, L., Alegbejo, M., Lava Kumar, P., et al. (2017). First Report of Maize yellow mosaic virus Infecting Sugarcane (*Saccharum* spp.) and Itch Grass (*Rottboellia cochinchinensis*) in Nigeria. *Plant Disease, 101*(7), 1335, doi:10.1094/PDIS-03-17-0315-PDN.

Figure Legends

Fig. 1 Phylogenetic analysis of maize yellow mosaic virus and related Poleroviruses from Africa, South America and China. The evolutionary history was inferred from whole genome sequences by using the Maximum Likelihood method based on the GTRGAMMA model. Bootstrap consensus values are shown at the nodes (1,000 repetitions). The whole genome sequences are labelled by the country of origin with the following NCBI accession numbers: RSA BR1A (MG570476), RSA SCM (MN943641), Kenya KALRO (MH205607), Tanzania 76 (MG664790), Ethiopia (MF684369), Nigeria (KY684356), China Y11 (KU248489), China Y1 (KU179221), China MYDV-RMV2 (KT992824), China SC (MK652149), Brazil (KY940544), Ecuador (KY052793), BVG Gimje (KT962089) and MYDV-RMV (KC921392). Barley virus G isolate Gimje and Maize yellow dwarf virus-RMV were used as outgroups. The scale bar indicates the number of nucleotide substitutions

Fig. 2 Maize yellow mosaic virus diagnostic RT-PCR from maize samples**.** RT-PCR products amplified using the primer pair MaYMV-F and MaYMV-R were visualized by agarose gel electrophoresis, with an expected 753 bp product from positive samples. **a**. RT-PCR templates from RNA of maize B73 RNA-seq samples positive for MaYMV. Lane M, Fast DNA Ladder (New England Biolabs, Ipswich, USA); lane 1, BR1A; lane 2, BR1B; lane 3, PCR no template control. **b**. RT-PCR templates were from a selection of maize leaf samples collected in South Africa: Lane M, O'GeneRuler 100 bp DNA Ladder (Thermo-Fischer, Waltham, USA); lane 1, 16-3379; lane 2, 17-4114; lane 3, 16-3308; lane 4, 16-3328; lane 5, 17-4610; lane 6, 17-4638; lane 7, 17-4641; lane 8, 17-4647; lane 9,16-3224; lane 10, 16-3252; lane 11, 16-3256; lane

12, 17-4263; lane 13, 17-4207; lane 14, 17-4187; lane 15, 17-4179; lane 16, 17-4109; lane 17, BR1B positive control; lane 18, PCR no template control

Fig. 3 Map of South Africa to show sites surveyed for maize yellow mosaic virus including fields were samples were collected for RNA-seq. The names of provinces are shown: FS: Free State, GT: Gauteng, KZN: KwaZulu-Natal, LP: Limpopo, MP: Mpumalanga, and NW: North West. Sites that were RT-PCR positive or negative for MaYMV are shown with plus signs or open circles, respectively. The MaYMV positive sites at Greytown and Baynesfield are marked as a closed circle and a closed triangle, respectively. The inset shows the position of the surveyed area within South Africa.

Fig. 4. Maize yellow mosaic virus-positive B73 plants also contain potyviruses and mycoviruses. RNAseq reads from B73 maize leaves collected from the 2013 field trial at Greytown were searched against the NCBI nr database (micro-organism) using Kaiju software, and reads that matched Riboviria (RNA viruses) are shown. The total amount of Riboviria reads in each sample was set at 100%. Luteoviridae, which include Poleroviridae corresponding to MaYMV, were present at 3, 15, 9, 12, 0.01 and 0.1% in BR1A, BR1B, BR2A, BR2B, BR3A and BR3B, respectively. The Potyviridae Sugarcane mosaic virus (SCMV) made up 51% of reads in BR1A, and Johnsongrass mosaic virus (JGMV) made up 81 and 83% of Riboviria in samples BR3A and BR3B, respectively. The total amount of Riboviria reads were 15475, 2357, 693, 2151, 439942 and 170518 in BR1A, BR1B, BR2A, BR2B, BR3A, & BR3B, respectively.

Electronic Supplementary Material

Online Resource 1 (ESM_1.pdf) Bioinformatics pipeline based on *de novo* assembly of unmapped reads used for discovery of maize yellow mosaic virus in maize RNA-seq data.

Online Resource 2 (ESM_2.pdf) Oligonucleotide primers used to amplify and Sanger sequence the complete genomes of maize yellow mosaic virus isolates RSA BR1A and RSA SCM.

Online Resource 3 (ESM_3.pdf) Percent nucleotide identity for complete genomes between maize yellow mosaic virus from South Africa and other worldwide isolates.

Online Resource 4 (ESM_4.pdf) Comparison of predicted maize yellow mosaic virus proteins of RSA BR1A and RSA SCM from South Africa and MaYMV Yunnan 11 from China.

Online Resource 5 (ESM_5.pdf) RNA-dependent RNA polymerase-based phylogenetic analysis of maize yellow mosaic virus and related Poleroviruses.

Online Resource 6 (ESM_6.pdf) Confirmation by Sanger sequencing that the expected MaYMV RT-PCR product was amplified from maize inbred B73 sample BR1B.

Online Resource 7 (ESM_7.pdf) Representative maize leaf symptoms of samples that were positive for maize yellow mosaic virus with the RT-PCR assay using MaYMV-F and MaYMV-R primers.

Online Resource 8 (ESM_8.pdf) Actin RT-PCR to confirm RNA and cDNA integrity for samples that were RT-PCR negative for maize yellow mosaic virus.

Online Resource 9 (ESM_9.xlsx) Maize B73 reads corresponding to RNA viruses.

652 **Table 1** Maize samples from 2016-2017 survey screened for maize yellow mosaic virus by RT-PCR

Sample number	Presence of MaYMV	Province	Site	Latitude	Longitude	Altitude (masl)	Month-Year collected
17-4016	positive	Free State	Kestell	-28.273869	28.906955	1704	January 2017
17-4276	positive	Gauteng	Bronkhorstspruit	-25.797133	28.570144	1511	March 2017
17-4245	positive	Gauteng	Bronkhorstspruit	-25.80011	28.654494	1442	March 2017
17-4246	positive	Gauteng	Bronkhorstspruit	-25.80011	28.654494	1442	March 2017
17-4101	positive	Gauteng	Krugersdorp	-26.087766	27.639682	1582	March 2017
17-4107	positive	Gauteng	Krugersdorp	-26.087766	27.639682	1582	March 2017
17-4109	positive	Gauteng	Krugersdorp	-26.087766	27.639682	1582	March 2017
17-4172	positive	Gauteng	UP* Experimental farm	-25.750395	28.259805	1364	March 2017
17-4179	positive	Gauteng	UP* Experimental farm	-25.750395	28.259805	1364	March 2017
17-4180	positive	Gauteng	UP* Experimental farm	-25.750395	28.259805	1364	March 2017
17-4187	positive	Gauteng	UP* Experimental farm	-25.750395	28.259805	1364	March 2017
17-4660	positive	KwaZulu-Natal	Ntabamhlophe (Estcourt)	-29.098092	29.706567	1494	April 2017
16-3283	positive-faint band	Limpopo	Groblersdal	-25.150167	29.393495	903	November 2016
16-3224	negative	Limpopo	Groblersdal	-25.220835	29.265568	953	November 2016
16-3252	negative	Limpopo	Groblersdal	-25.305412	29.409676	945	November 2016
16-3256	negative	Limpopo	Groblersdal	-25.305412	29.409676	945	November 2016
17-4661	positive	Limpopo	Modimolle	-24.565655	28.159019	1241	April 2017
17-4610	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4612	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4627	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4638	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4641	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4647	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4659	positive	Limpopo	Ofcolaco	-24.096708	30.406472	626	March 2017
17-4240	positive	Mpumalanga	Balmoral	-25.870148	29.015475	1446	March 2017
17-4228	positive	Mpumalanga	Malelane	-25.485229	31.517655	311	February 2017
17-4275	positive	Mpumalanga	Malelane	-25.524097	31.414203	336	February 2017
17-4269	positive	Mpumalanga	Malelane	-25.617295	31.656612	238	February 2017

Sample number	Presence of MaYMV	Province	Site	Latitude	Longitude	Altitude (masl)	Month-Year collected
17-4261	positive	Mpumalanga	Malelane	-25.618222	31.645087	245	February 2017
17-4263	positive	Mpumalanga	Malelane	-25.618222	31.645087	245	February 2017
17-4284	positive	Mpumalanga	Middelburg	-25.85413	29.38483	1480	March 2017
17-4290	positive	Mpumalanga	Middelburg	-25.85413	29.38483	1480	March 2017
17-4298	positive	Mpumalanga	Middelburg	-25.85413	29.38483	1480	March 2017
17-4135	positive-faint band	Mpumalanga	Nelspruit	-25.437556	30.649663	848	February 2017
17-4203	positive	Mpumalanga	Nelspruit	-25.439789	30.995983	674	February 2017
17-4207	positive	Mpumalanga	Nelspruit	-25.439789	30.995983	674	February 2017
17-4010	positive-faint band	Mpumalanga	Secunda	-26.612128	29.151277	1610	January 2017
16-3328	negative	North West	Brits	-25.238817	27.550659	965	November 2016
16-3379	positive	North West	Brits	-25.440832	27.657748	1021	November 2016
16-3304	positive-faint band	North West	Brits	-25.693796	27.784004	1151	November 2016
16-3308	negative	North West	Brits	-25.693796	27.784004	1151	November 2016
17-4114	positive	North West	Wolmaranstad	-27.477863	25.750998	1268	March 2017
17-4151	positive	North West	Wolmaranstad	-27.477863	25.750998	1268	March 2017
17-4164	positive	North West	Wolmaranstad	-27.477863	25.750998	1268	March 2017

*UP (University of Pretoria)

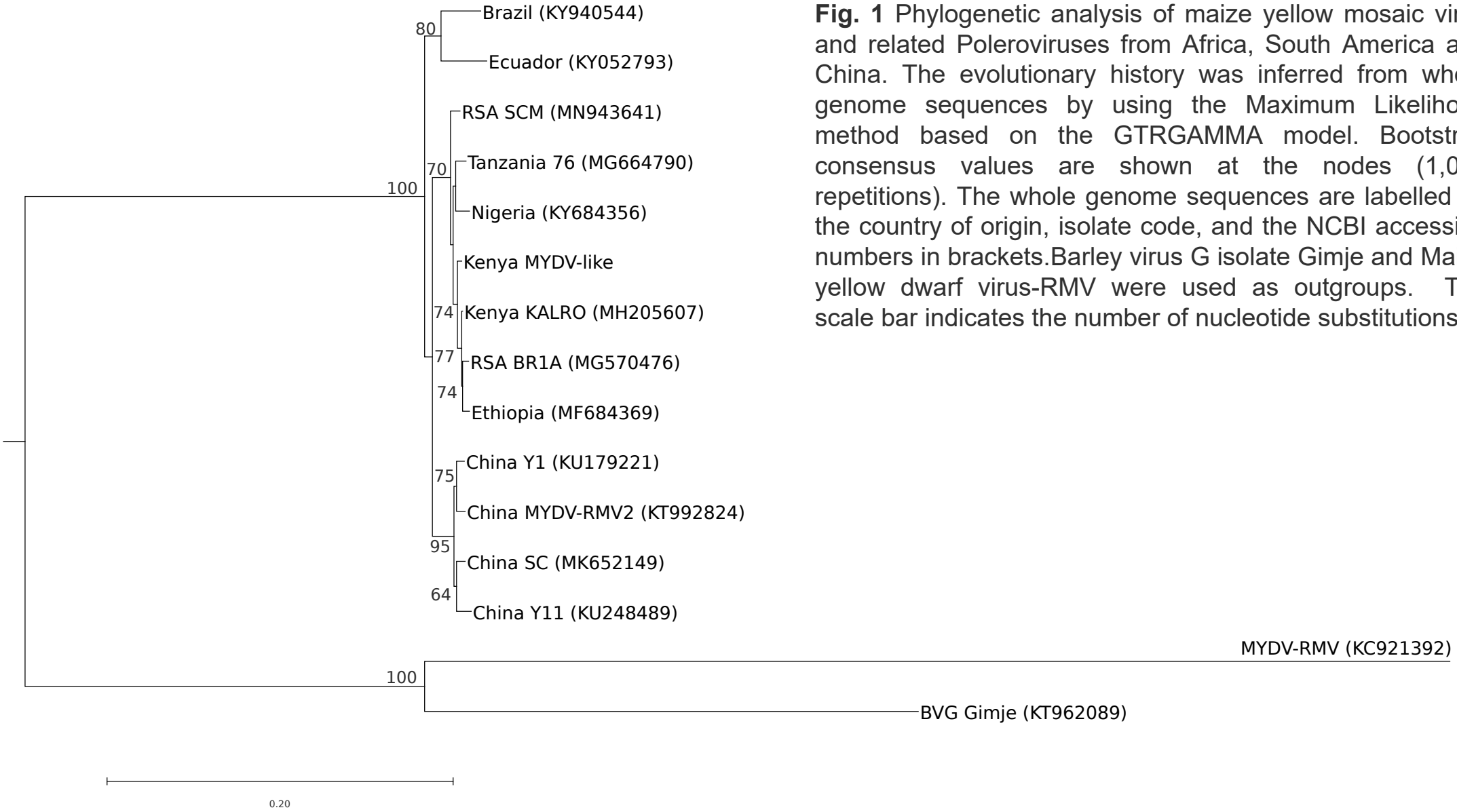


Fig. 1 Phylogenetic analysis of maize yellow mosaic virus and related Poleroviruses from Africa, South America and China. The evolutionary history was inferred from whole genome sequences by using the Maximum Likelihood method based on the GTRGAMMA model. Bootstrap consensus values are shown at the nodes (1,000 repetitions). The whole genome sequences are labelled by the country of origin, isolate code, and the NCBI accession numbers in brackets.Barley virus G isolate Gimje and Maize yellow dwarf virus-RMV were used as outgroups. The scale bar indicates the number of nucleotide substitutions

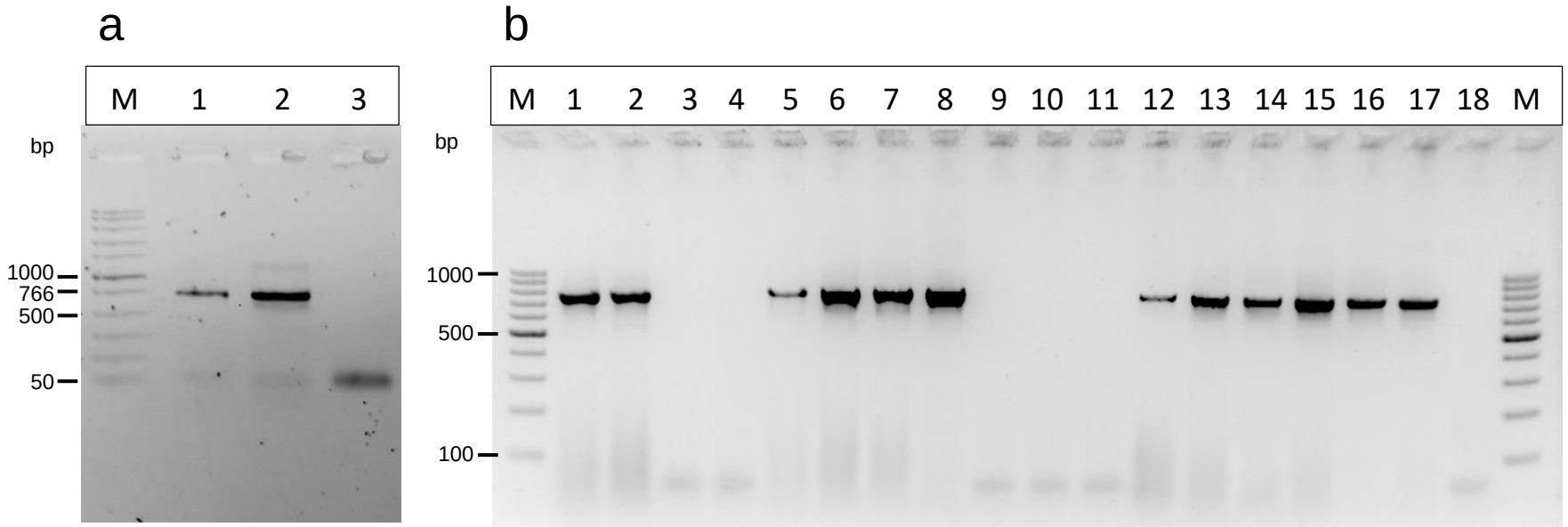


Fig. 2 Maize yellow mosaic virus diagnostic RT-PCR from maize samples. RT-PCR products amplified using the primer pair MaYMV-F and MaYMV-R were visualized by agarose gel electrophoresis, with an expected 753 bp product from positive samples. a. RT-PCR templates from RNA of maize B73 RNA seq samples positive for MaYMV. Lane M, Fast DNA Ladder (New England Biolabs, Ipswich, USA); lane 1, BR1A; lane 2, BR1B; lane 3, PCR no template control. b. RT-PCR templates were from a selection of maize leaf samples collected in South Africa: Lane M, O'GeneRuler 100 bp DNA Ladder (Thermo-Fischer, Waltham, USA); lane 1, 16-3379; lane 2, 17-4114; lane 3, 16-3308; lane 4, 16-3328; lane 5, 17-4610; lane 6, 17-4638; lane 7, 17-4641; lane 8, 17-4647; lane 9,16-3224; lane 10, 16-3252; lane 11, 16-3256; lane 12, 17-4263; lane 13, 17-4207; lane 14, 17-4187; lane 15, 17-4179; lane 16, 17-4109; lane 17, BR1B positive control; lane 18, PCR no template control

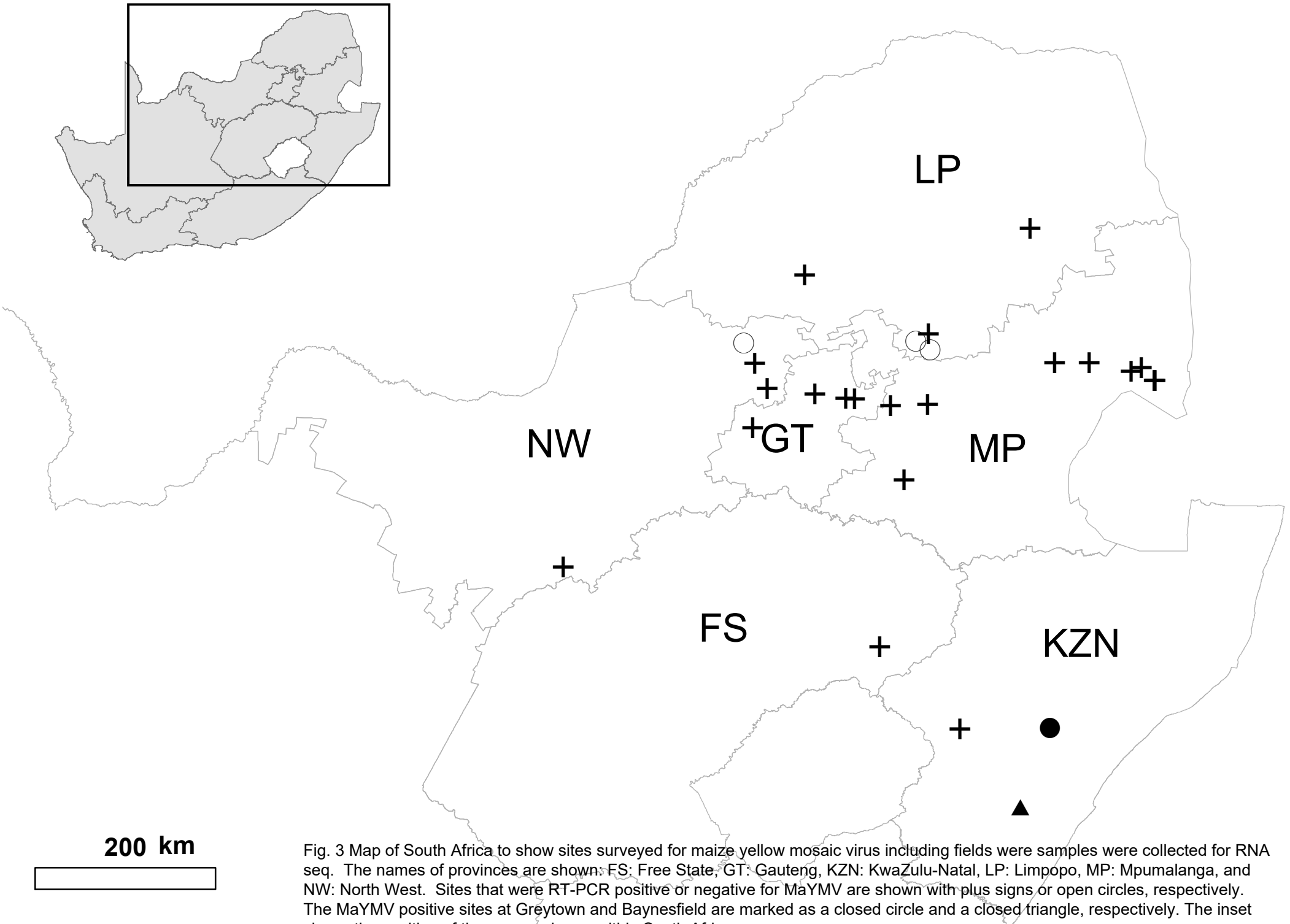


Fig. 3 Map of South Africa to show sites surveyed for maize yellow mosaic virus including fields were samples were collected for RNA seq. The names of provinces are shown: FS: Free State, GT: Gauteng, KZN: KwaZulu-Natal, LP: Limpopo, MP: Mpumalanga, and NW: North West. Sites that were RT-PCR positive or negative for MaYMV are shown with plus signs or open circles, respectively. The MaYMV positive sites at Greytown and Baynesfield are marked as a closed circle and a closed triangle, respectively. The inset shows the position of the surveyed area within South Africa

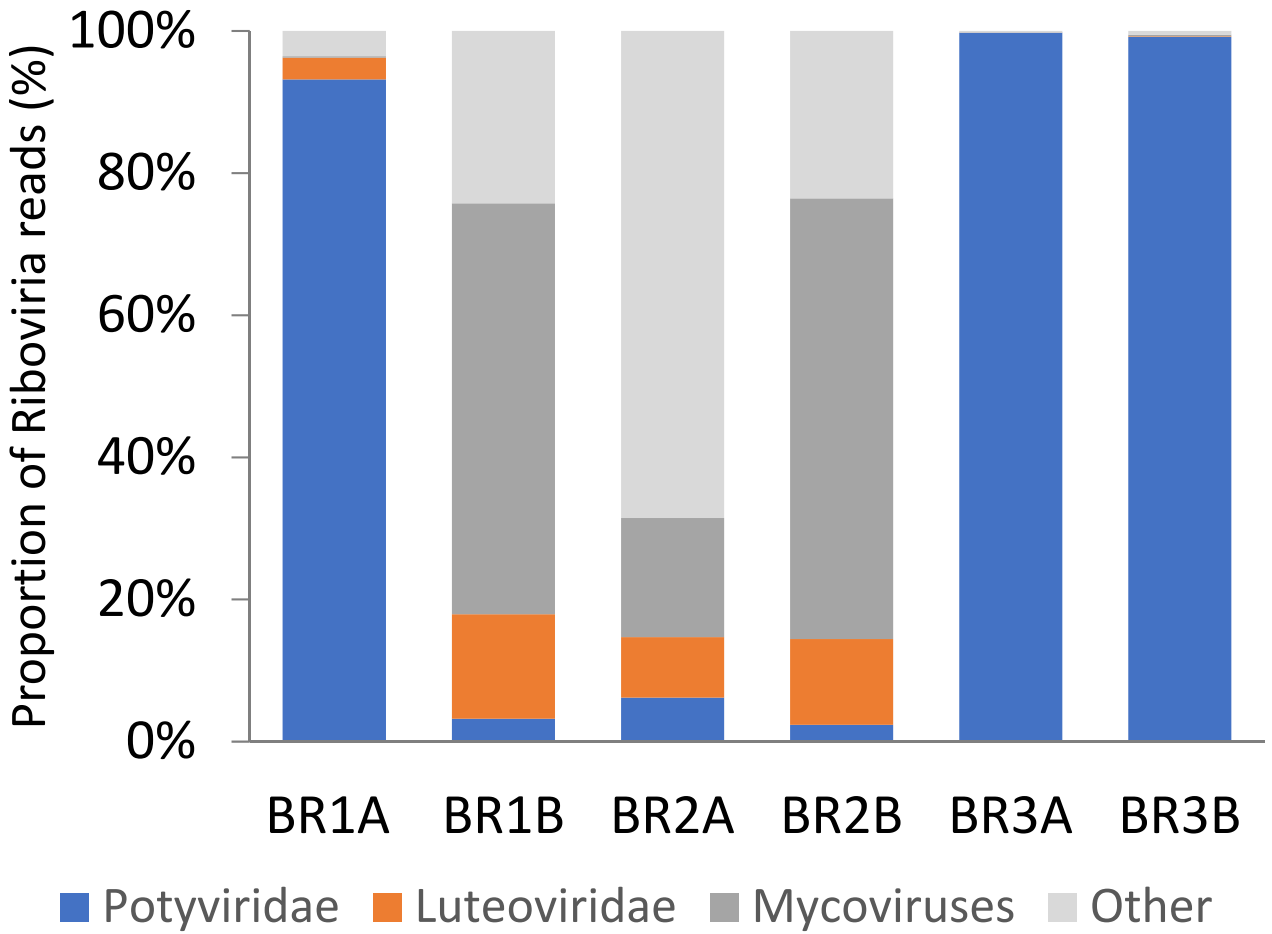
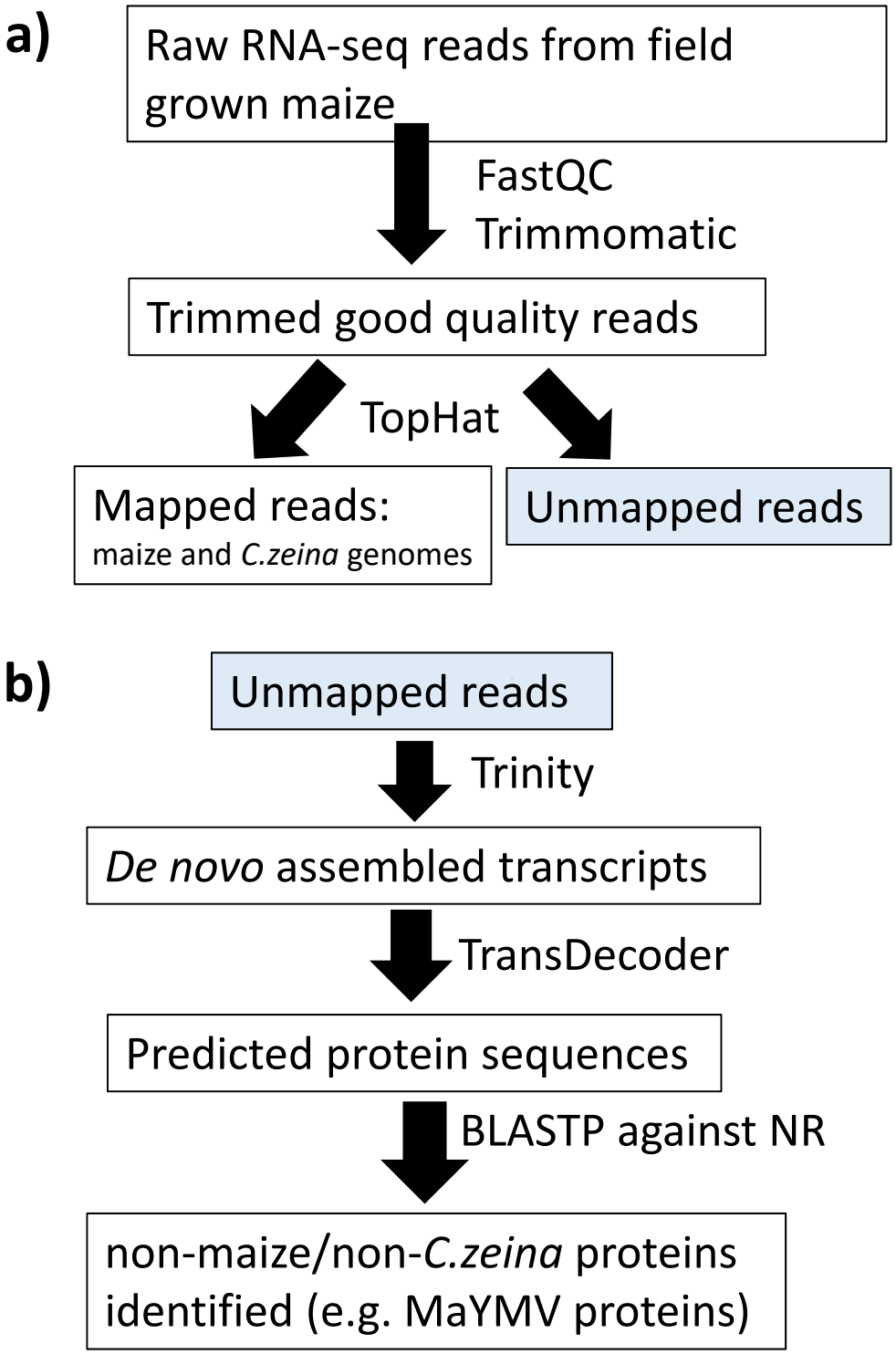


Fig. 4. Maize yellow mosaic virus-positive B73 plants also contain potyviruses and mycoviruses. RNAseq reads from B73 maize leaves collected from the 2013 field trial at Greytown were searched against the NCBI nr database (micro-organism) using Kaiju software, and reads that matched Riboviria (RNA viruses) are shown. The total amount of Riboviria reads in each sample was set at 100%. Luteoviridae, which include Poleroviridae corresponding to MaYMV, were present at 3, 15, 9, 12, 0.01 and 0.1% in BR1A, BR1B, BR2A, BR2B, BR3A and BR3B, respectively. The Potyviridae Sugarcane mosaic virus (SCMV) made up 51% of reads in BR1A, and Johnsongrass mosaic virus (JGMV) made up 81 and 83% of Riboviria in samples BR3A and BR3B, respectively. The total amount of Riboviria reads were 15475, 2357, 693, 2151, 439942 and 170518 in BR1A, BR1B, BR2A, BR2B, BR3A, & BR3B, respectively.



Online Resource 1. Bioinformatics pipeline based on *de novo* assembly of unmapped reads used for discovery of maize yellow mosaic virus in maize RNA-seq data. (a) Raw reads were assessed with FastQC, trimmed with Trimmomatic based on the FastQC results, and aligned to the reference genomes of maize and *C. zeina* to collect unmapped reads. (b) Unmapped reads were assembled using Trinity, their protein sequences predicted with TransDecoder and compared against the NCBI nr database using BLASTP.

Online Resource 2 Oligonucleotide primers used to amplify and Sanger sequence the complete genomes of Maize yellow mosaic virus isolates RSA BR1A and RSA SCM.

Primer name	Primer sequence 5' - 3'	Genome position*	Primer length (nt)	Tm °C
MaYMV_5'_RACE_GSP1	CGAGGAAATCGAATGGAGAT	409-428	20	n/a
MaYMV_5'_RACE_GSP2	GCTATCTCCCGTGATAATGCGCTGTA	366-391	26	n/a
MaYMV_3'_RACE_GSP	CGTCTCTTCAGACGCCAGCTAATTCA	4969-4997	26	n/a
MaYMV_frag_1F	CGCAATTCGGACTGGAGGAAACAT	271-294	24	62
MaYMV_frag_1R	CGTCGGAATTCGTGGTTGTCAAGA	1066-1089	24	62
MaYMV_frag_2F	CTGGTGGTTACGCTTCATGTGTTC	922-945	24	61
MaYMV_frag_2R	GACAAGCAGGTTGCTTTGAGAAGTCT	2163-2188	26	61
MaYMV_frag_3F	GTATGTGACACCCGCTCTGAGA	1850-1871	22	61
MaYMV_frag_3R	GATTTCTGAACTCCAGCCACTCTCT	2880-2904	25	61
MaYMV_frag_4F	CGGAATGGATGCTTCAAGACGATAT	2736-2760	25	59
MaYMV_frag_4R	GCAAGGGATGAGAGCTTGTTGT	3880-3901	22	60
MaYMV_frag_5F	CGGGCCGTCTTTATCAGAGTGT	3734-3755	22	62
MaYMV_frag_5R	CACGCAACGACTTTCGCTGAA	5018-5038	21	61

*Primer position on the MaYMV RSA BR1a genome sequence

Online Resource 3 Percent nucleotide identity for complete genomes between maize yellow mosaic virus from South Africa and other worldwide isolates

Region	Virus species name given (Genbank/Reference)	Virus isolate country and name	NCBI Accession #	Short name (Figure 1)	%nucleotide identity to MaYMV RSA BR1A (MG570476)	%nucleotide identity to MaYMV RSA SCM (MN943641)	Host Plant(s)	Region (district)	Year collected	Reference
Africa	Maize yellow mosaic virus (MaYMV)	South Africa (RSA BR1A)	MG570476	RSA BR1A	100%	98.50%	Maize	South Africa (KwaZulu-Natal province)	2013	This study
Africa	MaYMV	South Africa (RSA SCM)	MN943641	RSA SCM	98.50%	100%	Maize	South Africa (KwaZulu-Natal province)	2012	This study
Africa	Maize yellow dwarf virus-RMV	Kenya (MYDV-RMV KARLO)	MH205607	Kenya KALRO	99.60%	98.80%	Maize	Kenya	2012-2014	1
Africa	MYDV-like polerovirus	Kenya (MYDV-like 97_EA)	MF974579	Kenya MYDV-like	99.10%	98.80%	Maize	Kenya	2013-2016	2
Africa	MaYMV	Tanzania 16/0076	MG664790	Tanzania 76	98.30%	98.30%	Maize	Tanzania	2015	3
Africa	MaYMV	Ethiopia MV115	MF684369	Ethiopia	99.30%	98.50%	Maize	Ethiopia (Benishangul-Gumuz, Oromia province)	2015 + 2016	4
Africa	MaYMV	Nigeria KDHTS	KY684356	Nigeria	97.00%	97.30%	Sugar Cane/Itch Grass	Nigeria (Northern Guinea savannah region)	2015	5
China	MaYMV	China Yunnan 11	KU248489	China Y11	96.90%	96.90%	Maize	China (Yunnan province)	2013 + 2014	6
China	MaYMV	China Yunnan 1	KU179221	China Y1	96.80%	96.90%	Maize	China (Yunnan province)	2013 + 2014	6
China	MYDV-RMV2	China MYDV-RMV2	KT992824	China MYDV-RMV2	96.70%	96.70%	Maize	China (Anhui province)	2013	7
China	MaYMV	China SC-2	MK652149	China SC	96.90%	96.90%	Sugar Cane	China (Neijiang, Sichuan province)	2018	8
America	MaYMV	Brazil SP-1	KY940544	Brazil	94.90%	95.10%	Maize	Brazil (Sao Paulo province)	2013 + 2017	9
America	MaYMV	Ecuador M04	KY052793	Ecuador	92.80%	92.80%	Maize	Ecuador (Los Rios province)	2016	10
Outgroup	Barley virus G	Barley virus G (Gimje)	KT962089	BVG Gimje	77.00%	76.90%	Barley	South Korea	2015	11
Outgroup	MYDV-RMV	Maize yellow dwarf virus-RMV	KC921392	MYDV-RMV	72.40%	72.40%	Wheat, Maize	USA	1987	12

1.Wamaitha, M.J., et al., Virology Journal, 2018. **15**(1): 90.

2.Massawe, D.P., et al., Virus Genes, 2018. **54**(3): p. 432-437.

3.Read, D.A., et al., Journal of Plant Pathology, 2019.**101**(1): p. 203.

4.Guadie, D., et al. Plant Disease, 2018.**102**(5): p. 1044.

5.Yahaya, A., et al. Plant Disease, 2017.**101**(7): p. 1335.

6.Chen, S., et al. Viruses, 2016.**8**(5): 120.

7.Wang, F., et al. Plant Disease, 2016. **100**(7): p. 1508.

8.Sun, S.-R., et al, Plant Disease, 2019. **103**(9): p. 2482.

9.Gonçalves, M.C. Plant Disease, 2017.**101**(12): p. 2156.

10.Bernreiter, A., et al. New Disease Reports, 2017**31**: p. 1.

11.Zhao,F., et al. 2015 (KT962089)

12.Krueger, E.N., et al. . Frontiers in Microbiology, 2013**4**: 205.

Online Resource 4 Comparison of predicted maize yellow mosaic virus proteins of RSA BR1A and RSA SCM from South Africa and MaYMV Yunnan 11 from China

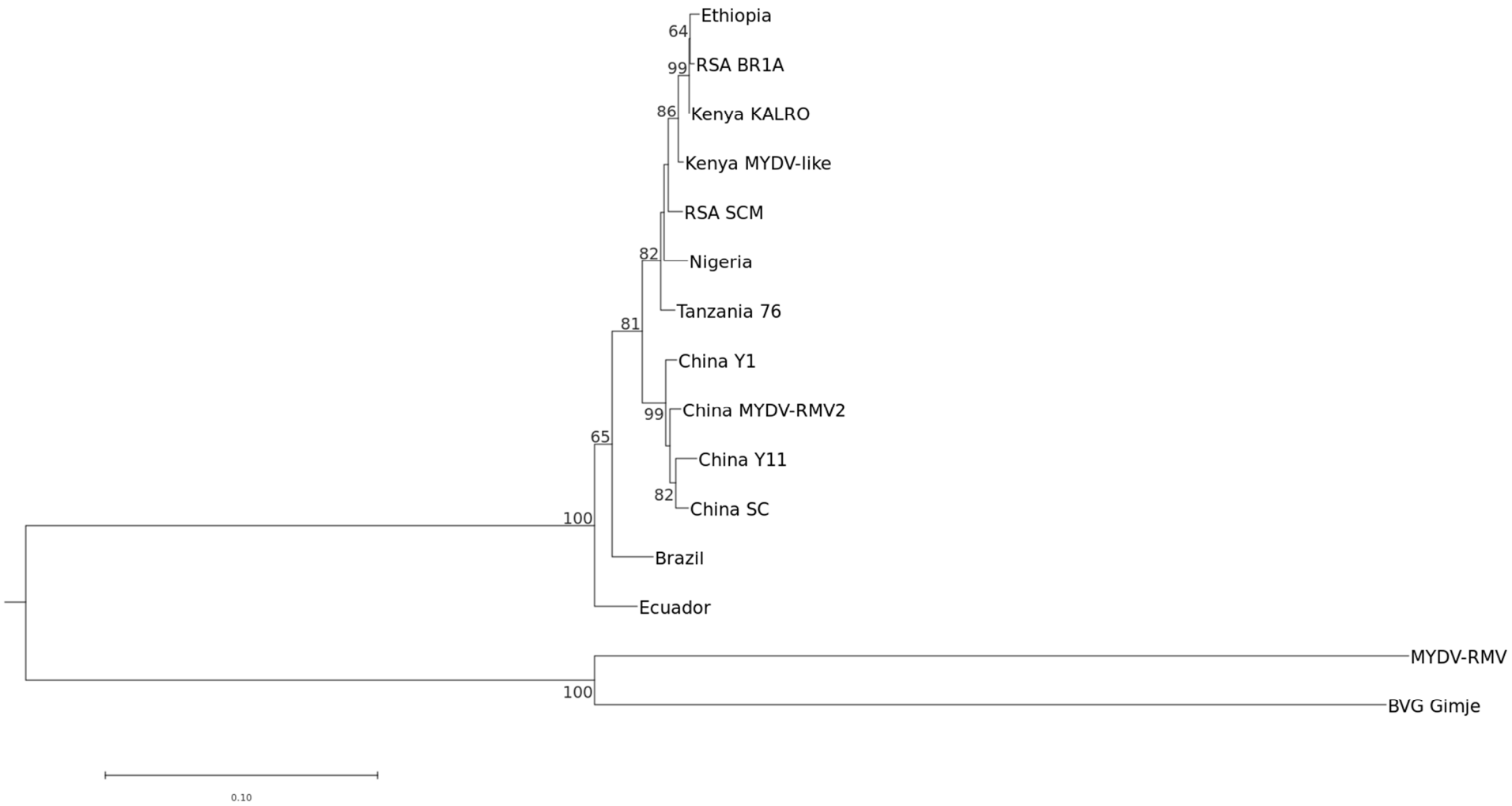
		RSA BR1A* vs Yunnan 11#		**RSA SCM& vs Yunnan 11#**		**RSA BR1A* vs RSA SCM&**	
Protein name	**Protein abbreviation**	% AA identity	% AA similarity^	% AA identity	% AA similarity^	% AA identity	% AA similarity^
Suppressor of Silencing	P0	95.8%	96.6%	95.1%	97.0%	94.7%	96.2%
	P1	97.1%	97.7%	96.3%	97.1%	97.9%	98.6%
RNA-dependant RNA polymerase	P1-P2	98.1%	98.7%	98.2%	98.7%	98.8%	99.1%
Coat protein	P3	100%	100%	100%	100%	100%	100%
Movement protein	P4	98.4%	98.4%	99.0%	99.0%	99.5%	99.5%
Readthrough domain protein	P3-P5	96.5%	97.9%	96.5%	97.7%	98.5%	99.1%

* Amino acid sequences from MaYMV RSA BR1A (Genbank Accession MG570477)

Amino acid sequences from MaYMV Yunnan 11 (Genbank Accession KU248489)

& Amino acid sequences from MaYMV RSA SCM (Genbank Accession MN943641)

^Amino acid similarity based on BLOSUM62



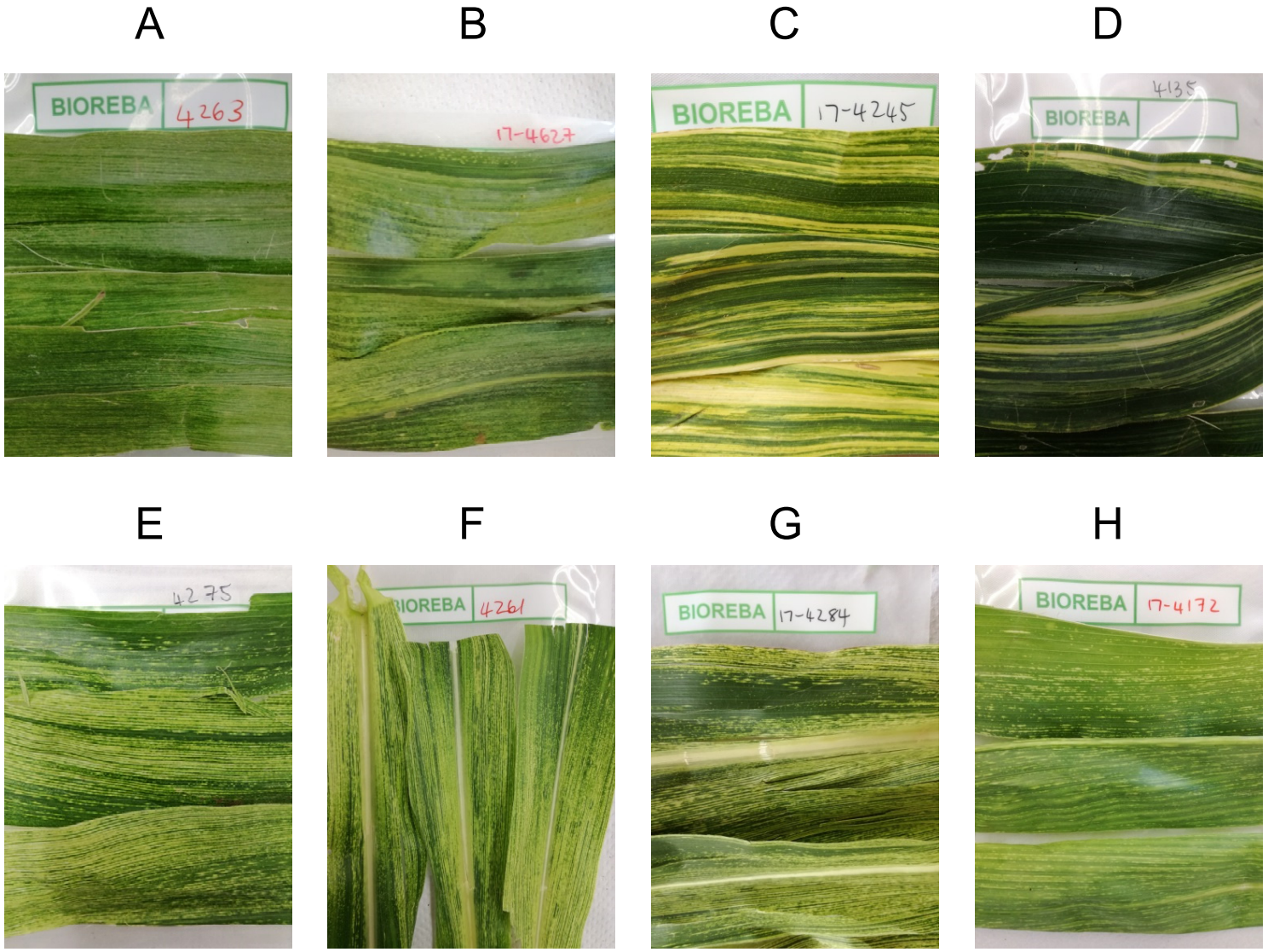
Online Resource 5. RNA-dependent RNA polymerase-based phylogenetic analysis of maize yellow mosaic virus and related Poleroviruses. The evolutionary history was inferred from the RNA-dependent RNA polymerase nucleotide sequences by using the Maximum Likelihood method with the GTRGAMMA model. Bootstrap consensus values are shown at the nodes. The sequences were extracted from complete genome sequences with the following NCBI accession numbers: RSA BR1A (MG570476), RSA SCM (MN943641), Kenya KALRO (MH205607), Tanzania 76 (MG664790.1), Ethiopia (MF684369), Nigeria (KY684356.1), China Y11 (KU248489.1), China Y1 (KU179221.1), China MYDV-RMV2 (KT992824.1), China SC (MK652149), Brazil (KY940544.1), Ecuador (KY052793), BVG Gimje (KT962089.1) and MYDV-RMV (KC921392.1). The latter two were used as outgroups. The scale bar indicates the number of nucleotide substitutions.

```
                          5′ TGCTCACATGTCTTCGAGAAGGAA 3′ MaYMV-F primer
                             1                                                  50
BR1A (MG570476)              TGCTCACATGTCTTCGAGAAGGAATACCTCGCACGCCCGGTTAACCAAAA
KU248489                     TGCTCACATGTCTTCGAGAAGGAACACCTCGCACGCCCGGTTAACCAAAA
BR1B.MaYMV.F                 --------------------------------------------------
BR1B.MaYMV.R                 ------------TTCGAGAAGGAATACCTCGCACGCCCGGTTAACCAAAA

                             51                                                 100
BR1A (MG570476)              CAAAATGATTTACAAATTGGTTTATGGTTACAACCCGGCGAACGGTTCAT
KU248489                     CAAAATGATCTATAAATTGGTTTATGGTTATAACCCGGCGAACGGTTCAT
BR1B.MaYMV.F                 --AAATGATTTACAAATTGGTTTATGGTTACAACCCGGCGAACGGTTCAT
BR1B.MaYMV.R                 CAAAATGATTTACAAATTGGTTTATGGTTACAACCCGGCGAACGGTTCAT
                                      *  *                 *

                             101                                                150
BR1A (MG570476)              CGGAGGTCTTGCAGCGTTATCTTGACGCCTGTATGAGCGTGCTTCACGAG
KU248489                     CGGAGGTCTTGCAGCGCTATCTTGACGCCTGCATGAGCGTGCTTCACGAG
BR1B.MaYMV.F                 CGGAGGTCTTGCAGCGTTATCTTGACGCCTGTATGAGCGTGCTTCACGAG
BR1B.MaYMV.R                 CGGAGGTCTTGCAGCGTTATCTTGACGCCTGTATGAGCGTGCTTCACGAG
                                           *             *

                             151                                                200
BR1A (MG570476)              CTACGCCATGACCCCGAAACCGTGGAGCTTCTGTACAAGTGGCTGGTATC
KU248489                     CTACGCCATGACCCCGAAACCGTGGAGCTTCTGTACAAGTGGCTGGTATC
BR1B.MaYMV.F                 CTACGCCATGACCCCGAAACCGTGGAGCTTCTGTACAAGTGGCTGGTATC
BR1B.MaYMV.R                 CTACGCCATGACCCCGAAACCGTGGAGCTTCTGTACAAGTGGCTGGTATC

                             201                                                250
BR1A (MG570476)              TCCAGTCCAGCAACAAAAGGTTTGAAAACAGAAGTTTCAAAGGTAGCCAG
KU248489                     TCCAGTCCAGCAACAAAAGGTTTGAACACAGAAGCTTCGAAGATAGCCAG
BR1B.MaYMV.F                 TCCAGTCCAGCAACAAAAGGTTTGAAAACAGAAGCTTCAAAGGTAGCCAG
BR1B.MaYMV.R                 TCCAGTCCAGCAACAAAAGGTTTGAAAACAGAAGYTTCAAAGGTAGCCAG
                                                 *       *   *   *

                             251                                                300
BR1A (MG570476)              ACACACGTGAGTTGCAAGTGCTGGAATCCTAGTCTCACACATAAGCAGCC
KU248489                     ACACACGTGAGTTGCAAGTGCTGGAATCTTAGTCTCACACACAAGCAGCC
BR1B.MaYMV.F                 ACACACGTGAGTTGCAAGTGCTGGAATCCTAGTCTCACACATAAGCAGCC
BR1B.MaYMV.R                 ACACACGTGAGTTGCAAGTGCTGGAATCCTAGTCTCACACATAAGCAGCC
                                                    *           *

                             301                                                350
BR1A (MG570476)              ATAGATTGGAAACTCTTTTGCGGGGTTCTTATAGGGATCCTCGTTGCTGT
KU248489                     ATAGATTGGAAACTCTTTTGCGGGGTTCTCATAGGGATCCTCGTTGCTGT
BR1B.MaYMV.F                 ATAGATTGGAAACTCTTTTGCGGGGTTCTTATAGGGATCCTCGTTGCTGT
BR1B.MaYMV.R                 ATAGATTGGAAACTCTTTTGCGGGGTTCTTATAGGGATCCTCGTTGCTGT
                                                     *

                             351                                                400
BR1A (MG570476)              CCCCGTAACCATCTTTGGCTTGTACAAGATCTACCTATCTATCTCCTCGA
KU248489                     CCCCGTAACCATCTTTGGCTTGTACAAGATCTACCTATCTATCTCCTCGA
BR1B.MaYMV.F                 CCCCGTAACCATCTTTGGCTTGTACAAGATCTACCTATCTATCTCCTCGA
BR1B.MaYMV.R                 CCCCGTAACCATCTTTGGCTTGTACAAGATCTACCTATCTATCTCCTCGA

                             401                                                450
BR1A (MG570476)              ACGTGCGTTCAATTGTGAATGAATACGGGAGGTAGAAATGGACGCAGAGC
KU248489                     ACGTGCGTTCAATTGTGAATGAATACGGGAGGTAGAAATGGACGCAGAGC
BR1B.MaYMV.F                 ACGTGCGTTCAATTGTGAATGAATACGGGAGGTAGAAATGGACGCAGAGC
BR1B.MaYMV.R                 ACGTGCGTTCAATTGTGAATGAATACGGGAGGTAGAAATGGACGCAGAGC

                             451                                                500
BR1A (MG570476)              TAGGAACCGCCGACGCGCTCGCAATAATAACCGGGCCCAGCCAGTGGTTG
KU248489                     TAGGAACCGCCGACGCGCTCGCAATAATAACCGGGCCCAGCCAGTGGTTG
BR1B.MaYMV.F                 TAGGAACCGCCGACGCGCTCGCAATAATAACCGGGCCCAGCCAGTGGTTG
BR1B.MaYMV.R                 TAGGAACCGCCGACGCGCTCGCAATAATAACCGGGCCCAGCCAGTGGTTG
```

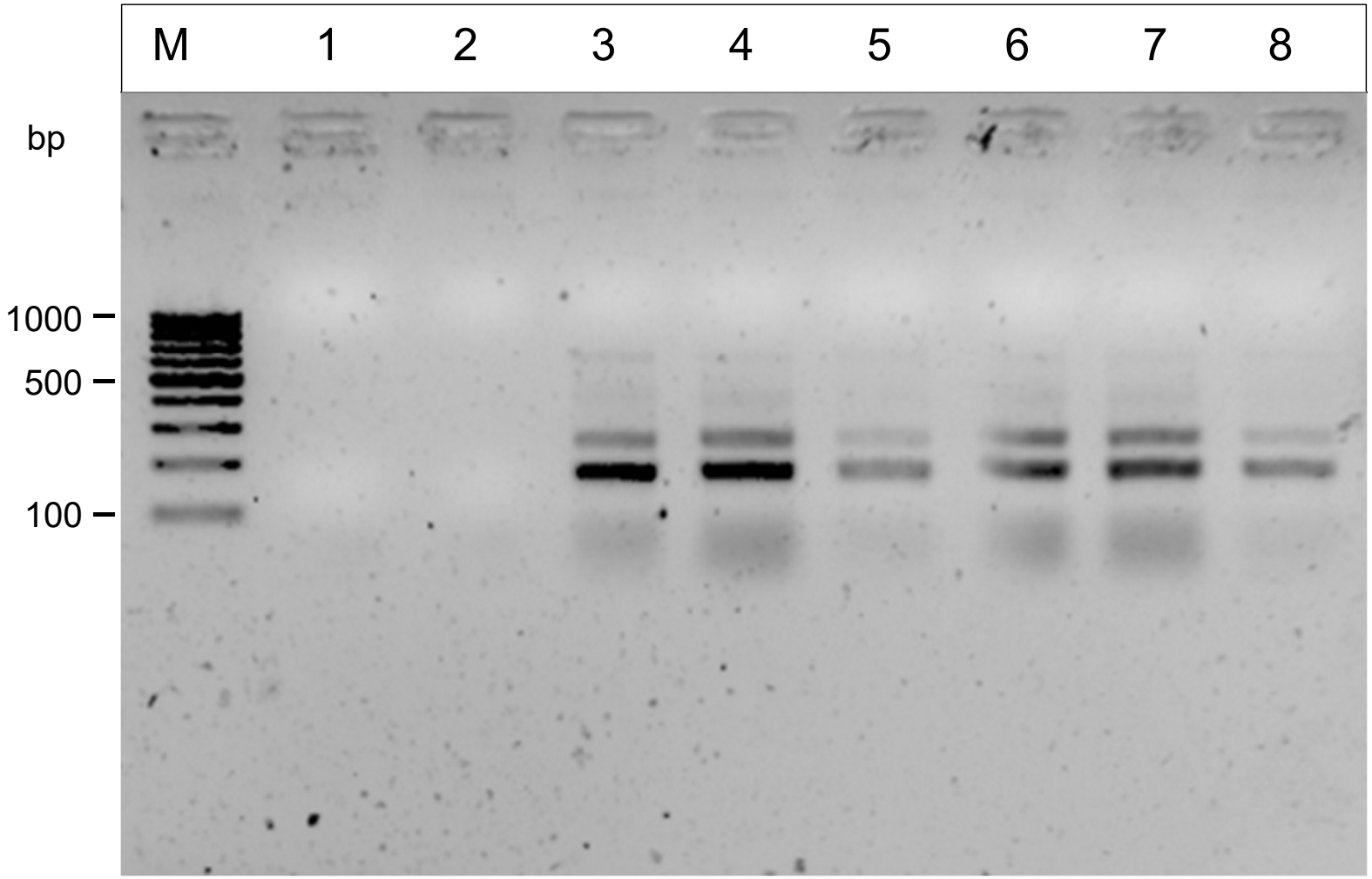
```
                          501                                                 550
BR1A (MG570476)           TTGTCGCGGCAAATCCGCGTCGAGGACGCCCTCGAAGACGAAGACGACCA
KU248489                  TTGTCGCGGCAAATCCGCGTCGAGGACGCCCTCGAAGACGAAGACGACCA
BR1B.MaYMV.F              TTGTCGCGGCAAATCCGCGTCGAGGACGCCCTCGAAGACGAAGACGACCA
BR1B.MaYMV.R              TTGTCGCGGCAAATCCGCGTCGAGGACGCCCTCGAAGACGAAGACGACCA

                          551                                                 600
BR1A (MG570476)           AGTGGAAACACTGCAGGAAGACCTGGAGTCAGACGAGGCTCGCGGGAGAC
KU248489                  AGTGGAAACACTGCAGGAAGACCTGGAGTCAGACGAGGCTCGCGGGAGAC
BR1B.MaYMV.F              AGTGGAAACACTGCAGGAAGACCTGGAGTCAGACGAGGCTCGCGGGAGAC
BR1B.MaYMV.R              AGTGGAAACACTGCAGGAAGACCTGGAGTCAGACGAGGCTCGCGGGAGAC

                          601                                                 650
BR1A (MG570476)           TTTTGTATTTTCAAAGGACTCTCTCACGGGCAATGCCTCCGGAAAAGTCA
KU248489                  TTTTGTATTTTCAAAGGACTCTCTCACGGGCAATGCCTCCGGAAAAGTCA
BR1B.MaYMV.F              TTTTGTATTTTCAAAGGACTCTCTCACGGGCAATGCCTCCGGAAAAGTCA
BR1B.MaYMV.R              TTTTGTATTTTCAAAGGACTCTCTCACGGGCAATGCCTCCGGAAAAGTCA

                          651                                                 700
BR1A (MG570476)           CCTTCGGGCCGTCTTTATCAGAGTGTGCAGCATTCAGTGGCGGAATTCTC
KU248489                  CCTTCGGGCCGTCTTTATCAGAGTGTGCAGCATTCAGTGGCGGAATTCTC
BR1B.MaYMV.F              CCTTCGGGCCGTCTTTATCAGAGTGTGCAGCATTCAGTGGCGGAATTCTC
BR1B.MaYMV.R              CCTTCGGGCCGTCTTTATCAGAGTGTGCAGCATTCAGTGGCGGAAT-CTC
                                                                          *

                          701                                                   753
BR1A (MG570476)           AAGGCCTACCATGAGTATAAGATCTCAAAGATCATACTGGAGTTCATCTCCGA
KU248489                  AAGGCCTACCATGAGTATAAGATCTCAAAGATCATACTGGAGTTCATCTCCGA
BR1B.MaYMV.F              AAGGCCTACCATGAGTATAAGATCTCAAAGATCATACTGGA---------
BR1B.MaYMV.R              AAGGCCTACC----------------------------------------
                                    MaYMV-R primer 3′ TCTAGTATGACCTCAAGTAGAGGCT 5′
```

Online Resource 6. Confirmation by Sanger sequencing that the expected maize yellow mosaic virus RT-PCR product was amplified from maize inbred B73 sample BR1B. Sequences derived from sequencing the 753bp RT-PCR product with the MaYMV-F or MaYMV-R primers were named BR1B.MaYMV.F or BR1B.MaYMV.R, respectively. These sequences were aligned to the corresponding sequence from the assembled MaYMV RSA BR1A genome sequence determined by 5'RACE, 3'RACE, RT-PCR and Sanger sequencing [BR1A (MG570476)], and the reference sequence MaYMV Yunnan 11 from China [KU248489]. The MaYMV-F and MaYMV-R primer sequences are reported in Chen *et al.* (2016). Non-consensus sites are shown by an asterisk.

Chen S, Jiang G, Wu J, Liu Y, Qian Y, Zhou X (2016) Characterization of a Novel Polerovirus Infecting Maize in China. Viruses 8 (5):120



Online Resource 7. Representative maize leaf symptoms of samples that were positive for maize yellow mosaic virus with the RT-PCR assay using MaYMV-F and MaYMV-R primers. A, B: Mosaic symptoms; C-D: Yellow streaks; E-H: Narrow yellow streaks. A-H samples: 17-4263, 17-4267, 17-4245, 17-4135, 17-4275, 17-4261, 17-4284, 17-4172, respectively. Maize genotypes are not known.



Online Resource 8. Actin RT-PCR to confirm RNA and cDNA integrity for samples that were RT-PCR negative for maize yellow mosaic virus. RT-PCR products amplified using the primer pair actinF and actinR were visualized by agarose gel electrophoresis, with an expected 169 bp cDNA actin product. Lane M, O'GeneRuler 100 bp DNA Ladder (Thermo-Fischer, Waltham, USA); lane 1, RT no template control; lane 2, PCR no template control; lane 3, maize healthy control/RNA positive control; lane 4, 16-3308; lane 5, 16-3328; lane 6, 16-3224; lane 7, 16-3252; lane 8, 16-3256. The additional 277 bp product in lanes 3-8 is the actin gDNA product, since the primers flank an intron, indicating presence of some gDNA in the samples.

Please see the online published version of the paper for Online Resource 9.xlsx. It is open access