South African Journal of Botany 117 (2018) 256–267

Contents lists available at ScienceDirect

South African Journal of Botany

journal homepage: www.elsevier.com/locate/sajb

Viral metagenomics reveals sweet potato virus diversity in the Eastern and Western Cape provinces of South Africa

T.F. Nhlapo [a,e,*], D.J.G. Rees [b], D.A. Odeny [c], J.M. Mulabisana [d], M.E.C. Rey [e]

[a] Biotechnology Platform, Agricultural Research Council, Private Bag X5, Onderstepoort 0110, South Africa
[b] Department of Life and Consumer Sciences, University of South Africa, Private Bag X6, Florida 1710, South Africa
[c] ICRISAT (ICRAF Campus), PO Box 39063-00623, Nairobi, Kenya
[d] Vegetable and Ornamental Plant Institute, Agricultural Research Council, Private Bag X293, Pretoria 0001, South Africa
[e] School of Molecular and Cell Biology, University of the Witwatersrand, Private Bag 3, Johannesburg PO Wits 2050, South Africa

ARTICLE INFO

Article history:
Received 7 April 2018
Accepted 29 May 2018
Available online xxxx

Edited by E Balázs

Keywords:
Viral metagenomics
Ipomoea batatas
Virus detection
Next generation sequencing (NGS)
Mixed infections

ABSTRACT

Limited studies have been undertaken with regard to virus complexes contributing to the aetiology of sweet potato virus disease (SPVD) in South Africa (SA). In this study, a metagenomic approach was adopted to reveal the genetic diversity of viruses infecting sweet potato. In order to undertake a comprehensive analysis of viral sequences, total RNA was isolated from 17 asymptomatic and symptomatic sweet potato plants that were collected from the Eastern (EC) and Western Cape (WC) provinces of SA. DNase-treated total RNA was depleted of ribosomal RNA (rRNA) and deep-sequenced using the Illumina MiSeq platform. Genomic DNA, isolated from the same plants, underwent rolling circle amplification (RCA) and deep sequencing. Sequence reads were analysed with the CLC Bio Genomics Workbench. Both *de novo* and reference-guided assemblies were performed resulting in four near full-length RNA virus genomes. BLAST searches using *de novo* assembled sequences against published virus genomes confirmed the presence of previously detected begomoviruses in the Western Cape (WC) province, namely Sweet potato mosaic virus (SPMaV) and Sweet potato leaf curl Sao Paulo virus (SPLCSPV). The begomoviruses were detected in mixed infections with two major disease-causing RNA viruses, Sweet potato feathery mottle virus (SPFMV) and Sweet potato chlorotic stunt virus (SPCSV). The sequence data further demonstrated mixed infections of RNA and DNA viruses from 11 of the 17 sequenced samples. Metagenomics is a reliable diagnostic tool for virus diversity detection, in particular virus-complexes and synergies affecting disease aetiology.

1. Introduction

Sweet potato (*Ipomoea batatas* (L.) Lam.) is a dicotyledonous perennial plant belonging to the morning glory family *Convolvulaceae*. It produces edible, highly nutritious tubers and is ranked as the 3rd most important root crop and the 7th most important staple crop in the world (Valverde et al., 2007; Clark et al., 2012). Sweet potato is an attractive crop to resource-poor farmers as it is easy to grow, high yielding, is drought and heat tolerant, and crowds out weeds (Kays, 2004). In countries such as Zambia, South Africa (SA), Uganda, Kenya, Nigeria, and Tanzania women are the primary growers of sweet potato and they use it to generate income. Sweet potatoes have a high content of carbohydrates and dietary fibre; they are rich in vitamin A, vitamin C, vitamin B6 and because of these nutritional benefits they are used for poverty alleviation (Van Jaarsveld et al., 2005). A number of DNA and RNA viruses accumulate in the crop as a result of vegetative propagation (Valverde et al., 2007), and pose a serious threat to sweet potato production (Valverde et al., 2007; Kreuze et al., 2009; Tesfaye et al., 2011). It has been recorded that viral diseases can decrease yield and quality of sweet potato storage roots by 30–100% in countries such as the United States (Valverde et al., 2007), Peru (Cuellar et al., 2008), SA (Domola et al., 2008), Kenya (Ateka et al., 2004) and Ethiopia (Tesfaye et al., 2011).

The most prominent viral disease complex known to affect sweet potato worldwide is sweet potato virus disease (SPVD), which was observed for the first time in Uganda in 1940 (Ateka et al., 2004). Although two viruses, Sweet potato chlorotic stunt virus (SPCSV) (family *Closteroviridae*) (Kreuze et al., 2002) and Sweet potato feathery mottle virus (SPFMV) (family *Potyviridae*) (Cuellar et al., 2008; Clark et al., 2012) are known to cause SPVD, there are reports in SA suggesting the involvement of several other viruses as role-players in sweet potato disease complexes (Domola et al., 2008; Tesfaye et al., 2011). The last comprehensive survey in SA documenting viral prevalence was carried out in 2003 and employed serology (ELISA assays), reverse transcription PCR (RT-PCR), electron

* Corresponding author.
E-mail address: thulilenhlapo@gmail.com (T.F. Nhlapo).

https://doi.org/10.1016/j.sajb.2018.05.024
0254-6299/

microscopy and indexing methods for virus detection (Domola et al., 2008). Using these traditional methods, a previous survey detected 9 viruses including SPFMV, SPCSV, Sweet potato mild mottle ipomovirus (SPMMV) (family *Potyviridae*), Sweet potato chlorotic fleck virus (SPCFV) (family *Flexiviridae*), Sweet potato caulimo-like virus (SPCV) (family *Caulimoviridae*), Sweet potato virus G (SPVG) (family *Potyviridae*), Sweet potato virus 2 (SPV2) (family *Potyviridae*), Sweet potato latent (SPLV) (family *Potyviridae*), and Sweet potato mild speckling virus (SPMSV) (family *Potyviridae*) (Domola et al., 2008). More recently, two begomoviruses, Sweet potato leaf curl Sao Paulo virus (SPLCSPV) and Sweet potato mosaic virus (SPMaV) were detected for the first time in the Limpopo province of SA using rolling circle amplification (RCA), cloning and sequencing (Esterhuizen et al., 2012). These sweet potato viruses, previously classified as sweepoviruses have recently been classified under begomoviruses (Brown et al., 2015). Begomoviruses belong to the family *Geminiviridae* and genus *Begomovirus* and consist of circular ssDNA viruses that can be bipartite or monopartite (Fauquet et al., 2003). The Infection of sweet potato with begomoviruses has been reported in Japan, Israel, Peru, Italy, Spain, China, Taiwan, Korea, Kenya, United States of America, Puerto Rico, Costa Rico, South Africa and Brazil (Kreuze and Fuentes, 2008; Paprotka et al., 2010; Albuquerque et al., 2011, 2012; Clark et al., 2012; Esterhuizen et al., 2012).

In the last two decades, virus detection methods have shifted from traditional techniques to metagenomic approaches coupled with high throughput sequencing (Boonham et al., 2014). Viral metagenomics has been used to identify novel viruses in plants (Kreuze and Fuentes, 2008; Idris et al., 2014). This approach is considered an unbiased one for viral detection since no prior knowledge of the virus is necessary, and neither virus-specific primers, nor antibodies are required. Consequently novel viruses, if present, can be detected, identified and quantified in a single experiment (Studholme et al., 2011). However, in the absence of reference sequences, the use of high throughput sequencing for virus detection requires *de novo* genome assembly of new virus sequences, which can be a challenge. A metagenomic approach also means that the entire microbial community within a sample can be described, even in mixed viral interactions, thus simplifying diagnostics (Idris et al., 2014). In most cases, the virus sequences generated in a metagenomic study would form a small proportion of the total nucleic acids making the removal of host sequences critical prior to, or after sequencing (Stobbe and Roossinck, 2014). For this reason, enrichment methods such as isolation of dsRNA (Clark et al., 2012) and small interfering RNA (siRNAs) (Kreuze et al., 2009) have been employed to detect DNA and RNA viruses from different hosts (Kashif et al., 2012). The availability of NGS platforms such as those supplied by Illumina (Illumina Inc., San Diego, CA, USA) has further revolutionized viral metagenomics studies. NGS technologies generate large amounts of data rapidly at reduced costs and many bioinformatic tools have been developed to handle data analysis (Massart et al., 2014). This study was carried out with the objective of establishing the current status of sweet potato viruses in two South African provinces (Western Cape and Eastern Cape). NGS of a symptomatic sweet potato field sample revealed a mixed infection of six viruses [SPMaV, SPLCSPV, SPFMV, SPCSV, SPVG and Sweet potato virus C (SPVC)], for the first time in SA.

2. Materials and methods

2.1. Sources of plant material

Sweet potato cuttings were collected from four smallholder farms in the EC province of South Africa: Alice (32°47′13.6″S 26°50′56.8″E), Zwide (33°52′12.8″S 25°34′24.0″E), Kwazakhele (33°53′11.0″S 25°36′00.7″E), and Motherwell (33°48′08.3″S 25°35′47.0″E). In the WC province, material was collected from four commercial farms, the locations included Paarl (33°40′12.0”S 18°58′08.0″E), Klawer (31°46′59.0”S 18°37′00.0″E), Franschhoek (33°55′00.1”S 19°07′59.9″E) and Lutzville (31°33′11.0”S 18°12′57.0″E). The cuttings were transplanted to potting soil in 20 cm diameter pots and grown in a greenhouse at the Agricultural Research Council – Vegetable and Ornamental Plant Institute (ARC-VOPI) in Pretoria, SA (25°40′51.67″S 28°17′10.25″E). Plants were grown at optimum temperatures of 25 °C for 16 h (day cycle) and 15 °C for 8 h (night cycle) (Domola et al., 2008). Plants were watered once a day and soluble nutrient fertilisation (Multifeed P, Plaaskem, Pty, LTD) was applied on a weekly basis. Insect pests were also monitored and controlled by spraying with recommended insecticides as required. A list of samples, abbreviations and symptoms are depicted in Table 2.

2.2. Description of symptoms

Sweet potato plants collected from the field were maintained in the glasshouse and observed for symptom development over a period of 6 months. Plants exhibiting symptoms typical of viral infection, such as upward curling of the leaves, chlorotic spots, vein clearing, and purple ring spots, were selected for analysis by sequencing. Symptom severity was scored using a 1–5 scoring scale (Mwanga et al., 2001; Domola et al., 2008) where, 1 = no virus symptoms, 2 = mild symptoms (chlorotic spots), 3 = moderate symptoms (chlorosis, chlorotic spots and vein clearing), 4 = severe symptoms (chlorotic spots, leaf curl, and leaf puckering/necrosis) and 5 = very severe symptoms (chlorotic spots, leaf curl, mottling, and stunting).

2.3. RNA library preparation and sequencing

After a period of 6 months, 10 symptomatic and 7 asymptomatic plants were randomly selected for further analysis (Table 2). Total RNA was isolated from the 17 samples using the QIAGEN RNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. To remove DNA, total RNA underwent DNase treatment using the QIAGEN RNase-free DNase Set (QIAGEN, Valencia, CA, USA). The integrity of the extracted RNA was analysed by agarose gel electrophoresis and quantified using the Qubit™ RNA BR Assay Kit (Invitrogen, Life Technologies). The total RNA was stored at −80 °C until further use. Total RNA was treated with the Ribo-Zero™ Magnetic Kit (Plant Leaf) (Epicentre, Madison, WI, USA) to deplete ribosomal RNA (rRNA). RNA paired-end libraries were prepared using the Illumina TruSeq™ Stranded Total RNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) following the manufacturer's instructions. The libraries were quantified using the Qubit™ dsDNA HS Assay Kit (Invitrogen, Life Technologies). The libraries were sequenced on the Illumina MiSeq (Illumina Inc., San Diego, CA, USA).

2.4. Nucleic acid isolation, rolling circle amplification (RCA), library preparation and sequencing

Genomic DNA (gDNA) was isolated from plant leaf material using the QIAGEN DNeasy Plant Mini Kit (*QIAGEN* Inc., Valencia, CA, USA) following manufacturer's instructions. The integrity of the extracted DNA was visualised by electrophoresis and quantified using the Qubit™ dsDNA BR Assay Kit (Life Technologies, Thermo Fisher Scientific Inc.). The gDNA was stored at −80 °C until further use. Prior to sequencing, genomic DNA underwent rolling circle amplification (RCA) using the Illustra™ TempliPhi™ 100 Amplification Kit (GE Healthcare, Amersham, UK) following the manufacturer's instructions. The RCA products were subjected to the Nextera DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) to generate 17 RCA libraries. Samples were sequenced using the Illumina MiSeq (Illumina Inc., San Diego, CA, USA) platform.

Table 1
Oligonucleotide sequences used for confirming the presence of RNA viruses.

Virus name	Primer name	Sequence (5′-3′)	Amplicon size (bp)
Sweet potato feathery mottle virus	SPFMV-F	TCCAGACCCTAAGTCCAAGAT	587
	SPFMV-R	AGTGCGTCATAATCTGCCTAAA	
Sweet potato chlorotic stunt virus	SPCSV-F	CTCTGACTCCGATGTAGGTTTC	590
	SPCSV-R	CGGTTGCAAGATGCCAATAC	
Sweet potato virus G	SPVG-F	GGAAACACAGGAAGAGGAAGAG	689
	SPVG-R	GGGACAGCATGATCCAATAGAG	
Sweet potato virus C	SPVC-F	GGCCATATACAGCACCAGAAA	338
	SPVC-R	TTCCTGAGTTGAGCGTGTATTC	
Sweet potato mosaic virus	SPMaV-F	CCGAAGCTATGTCCCGATTT	314
	SPMaV-R	GGTCCTTATTGGGCCTTCTATC	
Sweet potato leaf curl Sao Paulo virus	SPLCSPV-F	TCGAGATAGGAGGCCCAATAA	322
	SPLCSPV-R	GCAACGCAGAGTCTGATACA	

2.5. Sequence assembly for RNA libraries

Adaptor removal and quality trimming was performed for each of the 17 RNA libraries using Fastq-mcf (Aronesty, 2013) and the quality of the sequence reads was analysed using FastQC. A quality threshold of 30 and a Phred score of 33 were selected for trimming options. Sequence reads below the length of 50 base pairs (bp) and greater than 180 bp were discarded. The trimmed reads were aligned to the sweet potato chloroplast sequence and sweet potato expressed sequence tags (ESTs) to subtract host sequences. The unmapped sequence reads were assembled into contigs using the CLC Genomics Workbench *de novo* assembly tool, with default parameters. The contigs were subjected to BLASTn and BLASTx searches against viral sequences downloaded from the NCBI database. To generate consensus sequences for phylogenetic analysis, sequence reads and contigs matching sweet potato viruses were mapped to the full-length reference genomes, using alignment settings: length fraction = 0.5 and similarity = 0.9. The newly assembled sequences for SPVG, SPFMV, SPVC, SPCSV RNA1 and SPCSV RNA2 were deposited in GenBank under accession numbers KT069224, KT069222, KT069223, KX932096 and KT069221 respectively.

2.6. Data analysis for 17 RCA libraries

The libraries underwent quality trimming using Fastq-mcf (Aronesty, 2013). A quality threshold of 30 and a Phred score of 33 were selected for trimming options. Sequence reads below the length of 50 bp and greater than 150 bp were treated as low quality and discarded. The trimmed sequence reads generated from the RCA libraries were aligned to the partial sweet potato chloroplast genome (accession number KF242475) and the sweet potato mitochondrial DNA (accession number FN421476) using the CLC Bio Genomics Workbench (version 7.5.1) (CLC bio, Aarhus, Denmark) in order to filter out host sequences. The unmapped sequence reads were collected and then assembled into contigs using the CLC Bio Genomics Workbench *de novo* assembly tool. The contigs generated from the *de novo* assembly were subjected to BLASTn and BLASTx searches against the viral databases downloaded from the NCBI database. The full-length reference sequences of the viruses detected in the BLAST searches were retrieved and used in subsequent reference-guided assemblies. The sequence reads and contigs matching sweet potato viruses were mapped to the full genomes of the closest hits, using mapping settings: length fraction = 0.7 and similarity = 0.9 in the CLC Bio Genomics Workbench. The newly assembled sequences for SPLCSPV and SPMaV were deposited in the GenBank under accession numbers KX859238 and KX859239 respectively.

2.7. Phylogenetic analysis

The complete genome sequences of SPFMV, SPVG, SPVC, SPCSV, SPLCSPV and SPMaV were retrieved from the NCBI database and used for multiple sequence alignments (MSA) and phylogenetic analysis in MEGA 6.06 (Tamura et al., 2013). The neighbour-joining method was used to generate phylogenetic trees. The bootstrap tests were conducted using 1000 replicates and the evolutionary distances were computed using the Jukes-Cantor method.

2.8. RT-PCR and PCR confirmation of viruses identified by NGS

Full-length genome sequences of closely related isolates of SPVG, SPFMV, SPVC, SPCSV, SPLCSPV and SPMaV were downloaded and used

Table 2
Symptoms and disease scoring of field plants collected from the Eastern and Western Cape provinces of South Africa.

Sample ID	Location	Province	Symptoms	Scoring
KT10[a]	Klawer	Western Cape	Leaf curl, purple ring spots	3
F11	Franschoek	Western Cape	Purple-edged vein feathering, purple ring spots	3
KF1[a]	Klawer	Western Cape	Leaf curl, purple ring spots, stunting	4
L18	Lutzville	Western Cape	Purple ring spots	2
KT6[a]	Klawer	Western Cape	Chlorotic spots	2
P2	Paarl	Western Cape	Chlorotic spots	2
L9	Lutzville	Western Cape	Leaf curl, purple ring spots	3
F4	Franschoek	Western Cape	Necrotic spots	2
K10	Klawer	Western Cape	Asymptomatic	1
P14	Paarl	Western Cape	Asymptomatic	1
L11	Lutzville	Western Cape	Asymptomatic	1
Z24	Zwide	Eastern Cape	Chlorotic spots, vein clearing	3
KZ17	KwaZakhele	Eastern Cape	Chlorotic spots	2
M19[b]	Motherwell	Eastern Cape	Asymptomatic	1
FH22	Fort Hare	Eastern Cape	Asymptomatic	1
M11[b]	Motherwell	Eastern Cape	Asymptomatic	1
FH14	Fort Hare	Eastern Cape	Asymptomatic	1

[a] During collection plants were infested with unidentified whitefly species.
[b] Insect damage was observed on sweet potato leaves.

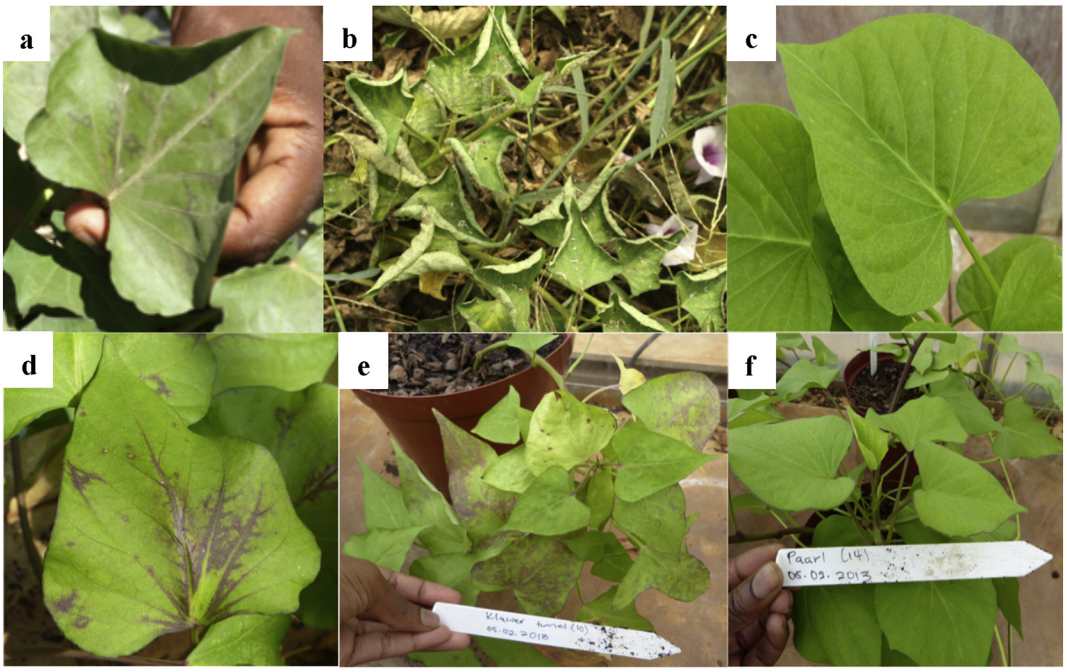


Fig. 1. Chlorotic spots with purple rings observed on sweet potato plants during field collection in KwaZakhele (Eastern Cape) (a). Sweet potato plants from Klawer (Western Cape) exhibited upward curling of the leaves (b). In the glasshouse KZ17 displayed mild chlorotic spots (c). F11 displayed purple ring spots and purple edged vein feathering (d), KT10 was characterised by leaf curling and purple ring spots (e), and P14 showing no virus symptoms (f). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in multiple sequence alignments (MSA). Alignments were submitted to the IDT PrimerQuest Tool (http://eu.idtdna.com/primerquest/home/index) for primer design. Conserved regions such as the coat protein gene were targeted for primer design. The primer sequences used in the study are listed in Table 1. Since DNA and RNA viruses were identified in sample KT10, genomic DNA (gDNA) was isolated from sample KT10 using the QIAGEN DNeasy Plant Mini Kit (*QIAGEN* Inc., Valencia, CA, USA) following manufacturer's instructions. Total RNA was also isolated from sample KT10 using the RNeasy Plant Mini Kit (QIAGEN, Valencia, CA, USA). The total RNA was converted to cDNA using the TaKaRa PrimeScipt 1st strand cDNA synthesis kit (TaKaRa, Japan). The polymerase chain reactions (PCRs) were performed using the TaKaRa EmeraldAmp® GT PCR Master Mix (TaKaRa, Japan) following the manufacturers instructions. PCR's consisted of 12.5 μl of the TaKaRa Emerald GT PCR Master Mix (TaKaRa, Japan), 0.5 μM of each primer (10 mM), 2 μl of the template DNA and 9 μl of nuclease-free water. The PCRs were performed using the recommended thermal cycling conditions: Initial denaturation at 94 °C for 5 min, 35 cycles of 94 °C for 30 s, annealing temperature (57–63 °C) for 45 s, extension at 72 °C for 2 min, and the final extension for 10 min at 72 °C. The amplicons were visualised on a 2% agarose gel by electrophoresis. The PCR amplicons were sent to Inqaba Biotechnical Industries (Pty) Ltd. service provider for direct

Table 3
Summary of the *de novo* assembly analysis.

Sample ID	Sequence reads before filtering	Sequence reads after QC analysis	Sequence reads mapped to host sequences	Unmapped sequence reads	*De novo* assembly		
					Number of contigs	Maximum contig length	N50
KT10	4605084	1474192	929232	544960	3374	10474[a]	336
F11	1400366	1295800	914635	381165	4836	6291[a]	347
KF1	835460	537401	308419	228982	3370	10436[a]	343
L18	541556	521024	341882	179142	4928	10540[a]	430
KT6	931874	551084	399984	151100	2439	5383	337
P2	777294	452674	267809	184865	2390	5646	340
L9	656312	357033	244743	112290	1202	5491	324
F4	724812	334552	220941	113611	1558	3373	319
K10	888036	570397	410811	159586	2118	2826	344
P14	557134	535138	369592	165546	6142	5386	422
L11	584506	566361	447747	118614	3846	5382	389
Z24	1602288	820006	584570	235436	4365	7101	355
KZ17	2015640	1594713	1106418	490295	11207	10442[a]	394
M19	789620	403739	210279	193460	3817	5386	331
FH22	398602	380539	265215	1115324	4767	5386	486
M11	653108	379272	257294	1211978	2708	5403	347
FH14	797028	448469	323391	125078	1814	5519	344

De novo assembly analysis was performed for each individual sample (library) to identify infecting viruses.
[a] Contigs resembling full virus genomes were obtained after host sequence filtration. More detail (virus name, number of sequence reads and coverage) of the identified viruses is given in Tables 4a & 4b.

Table 4a
Summary statitics of sweet potato viruses detected in symptomatic and asymptomatic plants using *de novo* assembly.

Sampled ID	Data after QC (Mb)	Virus detected	Total number of virus contigs[a]	Total number of sequence reads[b]	Maximum contig length	Number of sequence reads[c]	Sequence depth of maximum contig
KT10	303.8	SPFMV	2	27,129	10,474	26,638	225X
		SPCSV-RNA1	4	3649	7675	3365	41X
		SPCSV-RNA2	3	9725	4935	6467	123X
		SPMaV	4	1133	944	693	68X
		SPLCSPV	2	133	1016	133	11X
F11	222	SPVG	4	3941	6291	2094	36X
		SPVC	14	705	3202	210	7X
		SPFMV	3	1727	5649	965	18X
KF1	95.4	SPCSV-RNA1	5	66	696	39	6X
		SPCSV-RNA2	9	287	902	98	12X
		SPFMV	2	4968	10,463	4885	52X
		SPLCSPV	3	108	849	81	10X
		SPMaV	1	27	393	27	7X
L18	115.3	SPCSV-RNA1	2	1302	7984	1281	25X
		SPCSV-RNA2	2	2953	7507	2279	46X
		SPFMV	1	14,834	10,540	14,834	222X
		SPLCSPV	4	766	562	61	14X
KT6	99.4	SPFMV	7	955	2423	208	9X
		SPLCSPV	4	127	1089	87	8X
P2	82.1	SPFMV	3	1270	5646	660	13X
		SPVC	6	1176	3095	306	11X
L9	67	SPVC	11	573	2199	154	8X
		SPFMV	8	52	423	10	3X
F4	60.1	SPVC	11	138	955	33	3X
P14	99.6	SPCV	1	8	252	8	5X
L11	120.7	SPLCSPV	1	9	356	9	3X
Z24	126.6	SPVC	6	2832	7101	1335	22X
		SPFMV	19	564	1214	97	9X
KZ17	149.9	SPVC	1	36,214	10,442	36,214	477X
M19	75.1	–	–	–	–	–	–
FH14	162.6	–	–	–	–	–	–
M11	159.1	–	–	–	–	–	–
K10	135.9	–	–	–	–	–	–
F22	89.5	–	–	–	–	–	–

Near complete sequences of sweet potato feathery mottle virus (SPFMV), sweet potato virus G (SPVG), sweet potato virus C (SPVC), Sweet potato chlorotic stunt virus (SPCSV) and short transcripts of DNA viruses sweet potato leaf curl Sao Paulo virus (SPLCSPV), sweet potato caulimo-like virus (SPCV) and sweet potato mosaic virus (SPMaV) were generated.
[a] Contigs identified as viruses during BLASTn and BLASTx searches.
[b] Overall number of virus-specific sequence reads.
[c] Total number of reads used to assemble maximum contig.

Sanger sequencing. Sequences were analysed using the Sequence Scanner Software v2.0 (Applied Biosystems). The edited sequences were then subjected to BLASTn and BLASTx searches to determine their identities.

3. Results

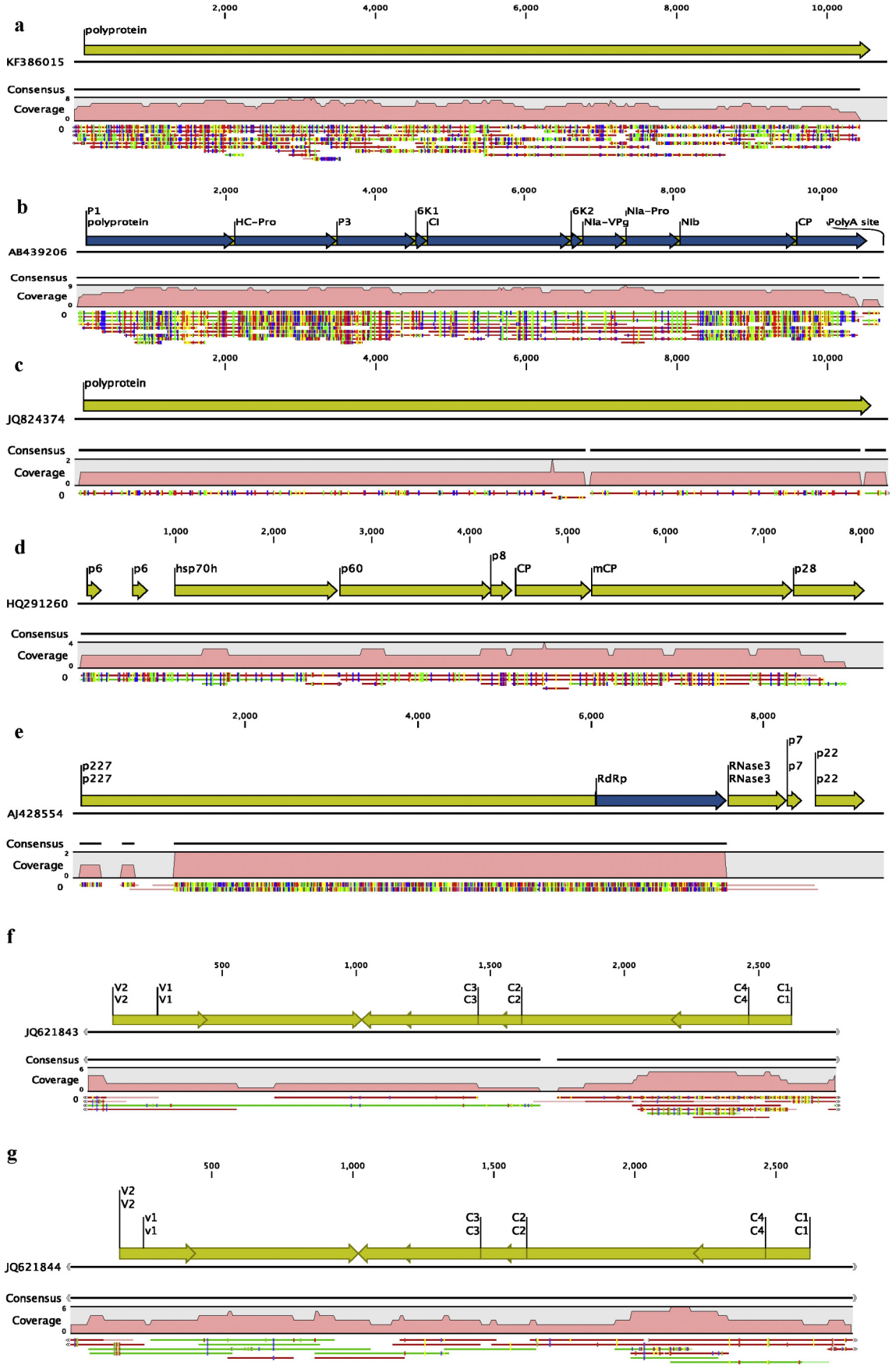
3.1. Field symptoms associated with viral infection

In the field a variety of symptoms were observed on sweet potato plants. Symptoms ranged from upward curling of leaves, purple-edged vein feathering and purple ring spots in samples KT10, F11, KF1 and L18; to chlorotic spots in KT6, P2, Z24, and KZ17; and vein clearing in Z24, while other samples were asymptomatic (Table 2 and Fig. 1). Symptom severity scoring of field samples can be viewed in Table 2. The most severe symptoms were observed on sample KF1, which was collected from the WC province. After one week of harvesting the leaves from sample KF1, the whole plant died. Samples F11 and KT10 displayed moderate symptoms characterised by purple ring spots and leaf curling respectively. These plants were also collected from the WC province. The phenotypic data suggests that sweet potato viruses were more prevalent in the WC, since most of the samples collected from this province were symptomatic. Interestingly, during sample collection in the WC province, unidentified whitefly species were observed on the sweet potato leaves. These could be possible vectors of several viruses that are associated with the symptoms recorded (*e.g.* leaf curling is associated with begomoviruses). Only two of the six samples from the EC province (KZ17 and Z24) displayed noticeable virus symptoms (Table 2). Furthermore during field collection sweet potato leaves were either symptomless or displayed mild symptoms in each of the four sampling locations of the EC province.

3.2. Sequence data and de novo assembly

The 17 individually labelled RNA libraries were sequenced to generate approximately 7 gigabases (Gb) of data. The primary data consisted of over 19 million sequences. After quality control, 56% of the data was retained for further analysis. Sixty eight percent of the retained data aligned to sweet potato chloroplast and EST sequences. For each sample,

Fig. 2. Genome coverage achieved by *de novo* assembly. Total contigs assembled for SPVC aligned along the full genome (a); contigs aligned to SPFMV covering the partial polyprotein sequence (b), the complete sequences of the HC-Pro, P3, 6 K1, CI, 6 K2, Nia-VPg, Nia-Pro, and NIb and partial CP genes were obtained; four contigs aligned to SPVG from sample F11 (c); the RNA2 segment of SPCSV had contigs aligned to the p6, hsp70h, p60, p8, CP, mCP, and partial p28 proteins (d); 4 contigs spanning over the p227 and RdRp proteins for RNA1 segment of SPCSV (e). The total number of contigs generated for DNA viruses SPMaV (f) and SPLCSV (g) achieved 80–100% genome coverage.


a
polyprotein
KF386015
Consensus
Coverage
b
P1
polyprotein
HC-Pro
P3
6K1
CI
6K2
NIa-VPg
NIa-Pro
NIb
CP
PolyA site
AB439206
Consensus
Coverage
c
polyprotein
JQ824374
Consensus
Coverage
d
p6
p6
hsp70h
p60
p8
CP
mCP
p28
HQ291260
Consensus
Coverage
e
p227
p227
RdRp
RNase3
RNase3
p7
p7
p22
p22
AJ428554
Consensus
Coverage
f
V2
V2
V1
V1
C3
C3
C2
C2
C4
C4
C1
C1
JQ621843
Consensus
Coverage
g
V2
V2
v1
v1
C3
C3
C2
C2
C4
C4
C1
C1
JQ621844
Consensus
Coverage

sequence reads that did not map to host sequences were assembled to generate contigs varying in number and size (Table 3). The *de novo* assemblies generated large numbers of contigs, with the largest contigs being in the range of 5–10 kilobases (kb), while the N50 values were in the range 300–486 bp (Table 3).

3.3. Detection of RNA viruses

The analysis of the assembled contigs was undertaken by matching all contigs to the viral sequences in the NCBI database using BLASTn and BLASTx. Overall, the majority of the assembled contigs showed no similarity to viral sequences, but significant matches to known sweet potato viruses were found in a large number of samples isolated from symptomatic plants (*e.g.* KT10, KF1, KT6, F11, L18, L9, P2, F4, Z24 and KZ17). Notably, five samples (M19, FH14, M11, K10 and F22) from asymptomatic plants had no detectable viral sequences in this analysis. Low counts of virus sequences were detected in two asymptomatic samples (L11 and P14) collected from the WC province. The viral sequences identified in each sample are shown in Table 4a. Viral sequences accounted for 1–2% of the total sequence data that was generated from the RNA libraries. In many cases the RNA virus genomes were assembled into a single contig (Fig. 2a and Table 4a) with a large number of reads supporting the contig assembly. In other instances there were lower levels of infection and only fragmentary assemblies of the genomes were achieved, however the total contig assemblies represented almost complete genome sequences when aligned with the reference genomes (Fig. 2 and Table 4a).

The BLAST search results revealed the presence of SPCSV RNA1 segment and SPCSV RNA2 segment in 3 samples. The total assembled contigs for SPCSV RNA1 represented a maximum of 72% of the East African strain reference sequence (accession number AJ428554), and for RNA2 the contigs represented a maximum of 92% of the m2–74 RNA2 sequence (accession number HQ291260). The two largest contigs generated for RNA1 were from samples KT10 and L18, at lengths of 7965 bp and 7984 bp respectively. Alignment and pairwise comparisons of these two contigs showed that were identical, except for the 5′ and 3′ ends, due to incomplete sequences. The RNA1 sequences shared 76% nucleotide (nt) similarity to the Ugandan reference sequence, while the RNA2 segment shared 97% nt identity to the Peruvian m2–74 isolate. This suggests that the two genome segments are of distinct origin, based on the widely divergent sequences observed. The two contigs matching to RNA1 of SPCSV were merged to generate a reference contig for extension with the PRICE genome assembler software using default parameters (Ruby et al., 2013). The final contig generated using PRICE was 8572 bp.

SPFMV was detected in 8 samples. The largest contig for SPFMV, 10540 bp, was assembled from sample L18. This contig covered 97% of the reference genome and shared 94% sequence identity with the SPFMV 10-O strain (accession number AB439206). Over 80% genome coverage was achieved for the *de novo* assembly of SPFMV in 7 samples (Table 4a) and over 50,000 sequence reads were assembled into 45 contigs. According to the *de novo* assembly, SPVC was the second most prevalent virus in the samples analysed. It was detected in 6 samples and the longest contig of 10442bp (Table 4a), which was assembled from sample KZ17, had a sequence depth of 477-fold. The contig shared 95% nt identity with the SPVC isolate from Argentina (accession number KF386015) and maximum genome coverage of 96% was achieved. BLAST search results also revealed the presence of SPVG in samples KT10 and F11. The largest contig of 6291 bp was assembled from sample F11. A total of 4 contigs were generated for SPVG, and when these were aligned to the reference genome a consensus of 10,577 bp, representing 97% of the genome, was generated (Fig. 2c). The consensus sequence shared 98% nt identity with an isolate from Argentina (accession number JQ824374). PCR amplification of the coat protein genes for the 4 RNA viruses was successful (Fig. 5). The PCR amplicons were Sanger sequenced to confirm their identity. The optimised PCR assay has the potential to be used in future studies for easy detection of SA virus isolates, especially SPCSV, which could not be detected previously using primers available in the literature.

Table 4b
Sweet potato viruses identified using reference-guided assembly. New full-length and partial virus genomes were assembled from less than 500 Mb of data.

Sample ID	Virus reference genome	Number of sequence reads mapped to reference genome	Size of newly generated sequence (bp)	Genome coverage (%); sequence depth
KT10	SPFMV	22,650	10803	(99); 186X
	SPCSV-RNA1	164	1419	(16); 1X
	SPCSV-RNA2	9101	8210	(99); 104X
	SPMaV	785	2544	(91); 26X
	SPLCSPV	510	2690	(97); 16X
	SPVC	171	785	(7); 0.9X
	SPVG	107	439	(4); 0.6X
F11	SPVG	3862	10739	(99); 39X
	SPVC	786	10161	(94); 7X
	SPFMV	1632	10192	(94); 16X
KF1	SPCSV-RNA1	2	230	(2); 0.03X
	SPCSV-RNA2	293	5948	(72); 3X
	SPFMV	4266	10392	(96); 44X
	SPLCSPV	111	1995	(72); 4X
	SPMaV	74	1421	(51); 2X
L18	SPCSV-RNA1	18	842	(9); 0.2X
	SPCSV-RNA2	3301	8193	(99); 62X
	SPFMV	12,103	10585	(97); 175X
	SPLCSPV	127	2186	(78); 5X
KT6	SPFMV	839	9859	(91); 8X
	SPLCSPV	53	1612	(57); 2X
	SPMaV	84	1639	(58); 3X
P2	SPFMV	1103	10317	(95); 11X
	SPVC	1150	10534	(97); 12X
L9	SPVC	524	9676	(89); 5X
	SPFMV	99	5411	(50); 1X
F4	SPVC	239	8455	(78); 2X
P14	SPCV	9	344	(4); 0.1
Z24	SPVC	2756	10486	(97); 30X
	SPFMV	532	9845	(90); 5X
KZ17	SPVC	35,939	10794	(100); 457X
L11	–	–	–	–
M19	–	–	–	–
FH14	–	–	–	–
K10	–	–	–	–
M11	–	–	–	–
F22	–	–	–	–

Reference genomes used in reference-guided assembly; SPFMV (Accession number: AB439206); SPCSV-RNA1 (Accession number: HQ291259); SPCSV-RNA2 (Accession number: HQ291260); SPVC (Accession number: KF386015); SPVG (Accession number: JQ824374); SPMaV (Accession number: JQ621843): SPLCSPV (Accession number: JQ621844): SPCV (Accession number: HQ694978).

3.4. Detection of DNA viruses

Interestingly gene transcripts of ssDNA begomoviruses (SPLCSPV and SPMaV) and sweet potato caulimo-like virus (SPCV) were detected in the RNA dataset (Table 4a). This could have occurred possibly as a result of purification of DNA viral transcripts. High genome coverage was achieved for SPLCSPV detected in sample KT10. While fragmentary assemblies were generated for SPMaV which was detected in samples KT10 and KF1. We conducted RCA and deep sequencing to generate full-length genomes for the DNA viruses. The individually labelled RCA libraries generated over 1,1 Gb of sequence data. Fifty nine percent of the paired-end data (approximately 2,3 million reads) was retained after adaptor and quality trimming and used for down stream analysis. Over 800 thousand reads mapped to the host sequences and the remaining 65% of the

clean reads were used for *de novo* assembly. The two DNA viruses (SPLCSPV and SPMaV), detected in the RNA datasets were also detected in the RCA libraries (Table 5). The *de novo* assembly generated a total of 11 contigs that were identified as SPMaV and 21 contigs were matching SPLCSPV (Table 5 and Fig. 2f & g). The consensus sequence for SPMaV shared 98% nt similarity to the SA SPMaV-isolate PstI-01 (accession number JQ621843), while the SPLCSPV consensus sequence shared 99% nt sequence similarity to the SA SPLCSPV-isolate PstI-012 (accession number JQ621844).

DNA viruses were detected in only 4 of the 17 plants that underwent RCA and deep sequencing (samples KT10, KF1, KT6 and L18) (Table 5). Begomoviruses were not detected in the plants collected from the EC. Both SPMaV and SPLCSPV were detected previously in SA from the Limpopo province in 2012, and now detected in this study for the first time in the WC province. A PCR assay for the two begomoviruses was performed and the expected band sizes of approximately 322 bp and 314 bp for SPLCSPV and SPMaV respectively were obtained (Fig. 5). Sanger sequencing confirmed the identity of the begomoviruses.

3.4.1. De novo assembly efficiency

When contigs were aligned to full-length viral genomes, gaps were observed in the consensus sequences. The only virus that was assembled with no gaps was SPVC (Fig. 2a). From this study we noticed that near full-length virus genome could be *de novo* assembled using datasets of 75 to 300 Mb (*e.g.* samples L18, F11, KT10, KZ17, Z24 and KF1) (Table 4a). This resulted in overall high genome coverage and sequence depth. In cases where large amounts of data were generated and there was low viral sequence count (*e.g.* P14) or no virus detection (M19, FH14, M11, K10 and F22), it was concluded that viruses were either absent or present at very low concentrations. The sequence data is supported by the phenotypic data. Samples M19, FH14, M11, K10, F22 and P14 showed no visible symptoms prior to sequencing (Table 2 and Fig. 1). The *de novo* assembly approach is effective for virus discovery and for the assembly of near complete viral sequences. This strategy is also efficient in the assembly of distinct viral sequences, where reference-guided assembly could pose a limitation.

3.5. Reference sequence-guided assembly

The reference-guided assembly showed that a total of 43,224 sequence reads originated from SPFMV (Table 4b). SPVC specific sequence reads were 41,265, while the *Crinivirus* (SPCSV) had a total of 12,879 sequence reads, and only 3862 sequence reads mapped to SPVG. Four new South African RNA virus genomes were generated from the reference-guided assembly. The SPCSV RNA2 segment sequence was 8210 bp long with a sequence depth of 104-fold (accession number KT069221); the SPFMV genome sequence was 10803 bp long and had a sequence depth186-fold and the new SPVG isolate was 10739 bp long, with a sequence depth of 39-fold (accession number KT069224). The new South African SPVC genome was 1 nucleotide (nt) longer than the reference sequence (10794 bp), and was assembled with a sequence depth of 457-fold (accession number KT069223). Few sequence reads aligned to the Peruvian and Ugandan SPCSV RNA1, possibly as a result of high variability within the South African genome sequence. The longest assembled contig for SPCSV RNA1 (accession number KX932096) was generated from the *de novo* assembly. From our RCA data we show that more reads corresponded to SPLCSPV (114211) than SPMaV (69950) (Table 5). Reference-guided assembly was able to generate full-length genome sequences for both begomoviruses. The new sequence for SPLCSPV was 2769 bp long (accession number KX859238) and SPMaV was 2781 bp long (accession number KX859239). Sequence depth for SPMaV ranged from 75 to 826-fold while it was 140–1932-fold for SPLCSPV. Sweet potato caulimo-like virus (SPCV) was also detected in the RCA data set, however a short sequence with gaps was assembled (Table 5).

3.6. Co-infections and mixed virus infections in sweet potato

The sequence data revealed co-infections of potyviruses (SPVC and SPFMV) in samples P2 and Z24 collected from the WC and EC provinces, respectively. Sequence reads matching SPFMV, SPVC and SPVG were detected from F11, which was collected from the WC. Three plants samples from the WC showed evidence of a virus complex including SPLCSPV, SPMaV, SPFMV and SPCSV (KF1, KT6 and L18) (Tables 4a & 4b; 5). A mixed infection of 6 viruses was detected in sample KT10

Table 5
Summary statistics of the DNA viruses detected by rolling circle amplification (RCA) and deep sequencing.

De novo assembly							Reference-guided assembly		
Sample ID	Virus detected	Length of reference genome (bp)	Maximum contig length (bp)	Total number of sequence reads	Total number of contigs	Total genome coverage (%)	Sequence reads mapped to reference	Length of new sequence (bp); coverage (%); depth (X)	Data after QC (Mb)
KT10	SPMaV	2783	757	25,649	6	80	28,993	2781; 99; 826	22,4
	SPLCSPV	2769	1349	28,695	3	90	27,083	2769; 100; 794	
F11	–	–	–	–	–	–	–	–	14
KF1	SPMaV	2783	1689	9194	3	80	22,748	2780; 99; 585	33,6
	SPLCSPV	2769	561	53,085	9	85	68,800	2769; 100; 1932	
L18	SPMaV	2783	1295	7097	1	46	18,209	2781; 99; 533	30,3
	SPLCSPV	2769	1251	15,875	5	45	18,328	2769; 100; 553	
KT6	SPMaV	2783	2781	1900	1	99	2631	2781; 99; 75	24,8
	SPLCSPV	2769	1499	5991	4	80	4808	2769; 100; 140	
P2	–	–	–	–	–	–	–	–	28,2
L9	–	–	–	–	–	–	–	–	21
F4	–	–	–	–	–	–	–	–	17,1
P14	SPCV	7723	–	–	–	–	39	608; 0.07; 0.07	271,1
L11	–	–	–	–	–	–	–	–	43,9
Z24	–	–	–	–	–	–	–	–	27,3
KZ17	–	–	–	–	–	–	–	–	23,1
M19	–	–	–	–	–	–	–	–	20,3
FH14	–	–	–	–	–	–	–	–	30,1
M11	–	–	–	–	–	–	–	–	16,6
K10	–	–	–	–	–	–	–	–	21,6
F22	–	–	–	–	–	–	–	–	22,6

from the WC (Table 4a). The sequence data shows that sweet potato viruses were found mostly as co-infections and mixed infections in plant samples from both provinces.

3.7. Phylogenetic analysis of RNA viruses

Near complete sequences of SPVC, SPVG, SPFMV and SPCSV RNA2 generated from the reference-guided assembly were used for phylogenetic analysis. Phylogenetic trees assigned RNA2 to the East African (EA) strain group (Fig. 3a). The SPFMV sequence grouped with the ordinary strains (Fig. 3b). The SPVC from the EC clustered with the SPVC group, which shares 95% nucleotide identity with the isolate from Argentina. The SPVG sequence from the WC clustered with isolates from Taiwan, USA and South Korea (Fig. 3b). A phylogenetic tree of begomoviruses (Fig. 4) placed the two DNA viruses from the coastal WC province of SA with previous SPLCSPV and SPMaV isolates detected from Waterpoort, Limpopo province, South Africa in 2012 (Esterhuizen et al., 2012).

4. Discussion

In this study we detected 6 different sweet potato viruses in various combinations in the EC and WC provinces of SA. A variety of known symptoms were observed in the infected field samples, which depended on the sweet potato cultivar and virus combination. This has been shown in several other studies (Gibson et al., 2004), where for example a combination of SPCSV and SPFMV caused severe symptom development on susceptible cultivars (Gibson et al., 1998). This study reports for the first time, the detection of two begomoviruses and four RNA viruses in a single plant in SA. Multiple infections of SPFMV, SPCSV, SPLCSPV and SPMaV (found in samples KT10, KF1, KT6 and L18 from the WC province), resulted in severe symptoms including upward curling of leaves, chlorotic spots, mottling and necrosis. The occurrence of multiple viruses in single plants, and correlation between mixed infections and symptom severity, has been reported in sweet potato (Mukasa et al., 2006; Tugume et al., 2016). However the combination of six viruses has not been reported. It is therefore necessary to further investigate how

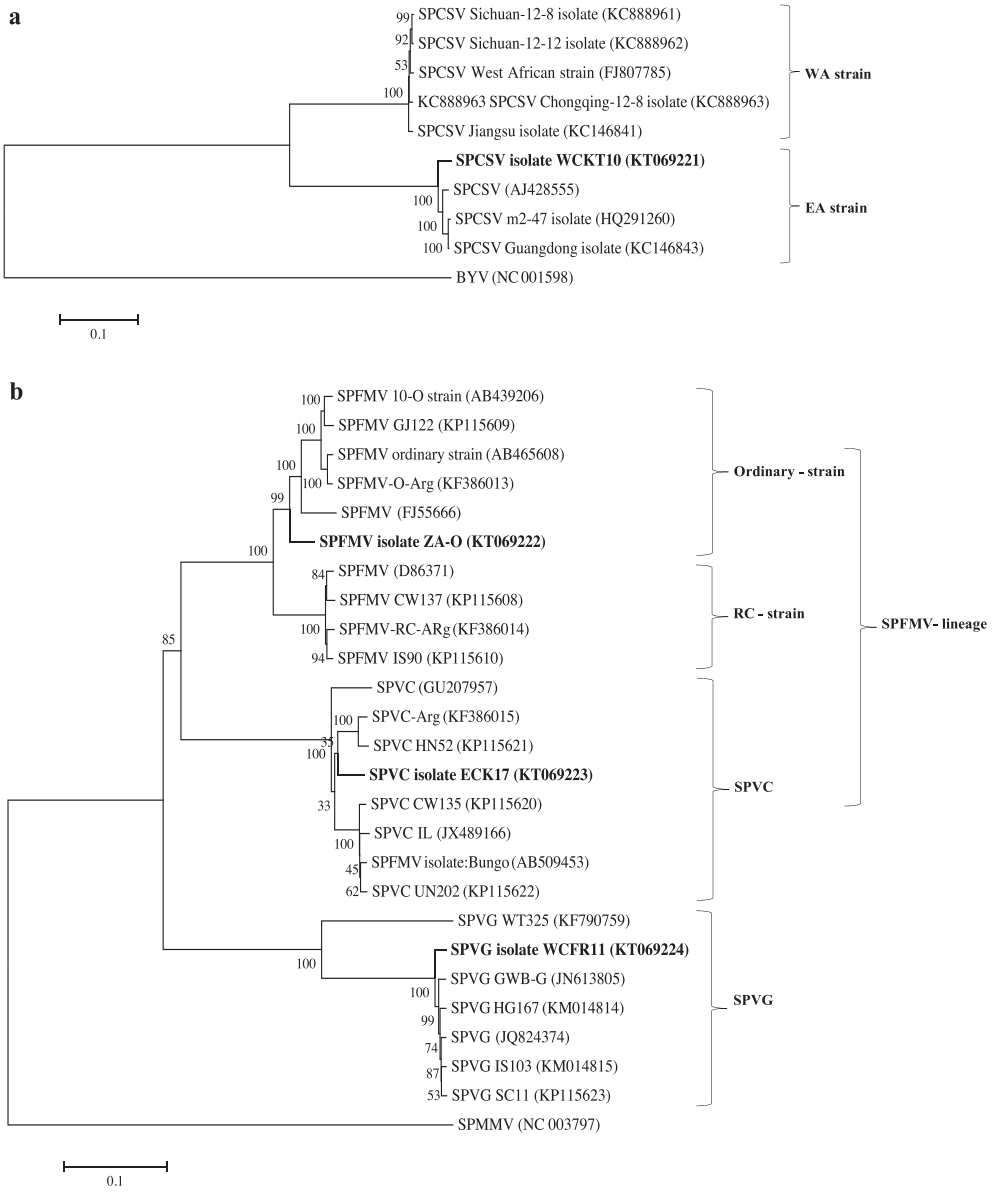


Fig. 3. Neighbour-joining tree of SPCSV RNA2 segments assigned the SA isolate to the East African (EA) group (a), beet yellow virus (BYV) was used as an outgroup. Phylogenetic analysis of potyviruses (SPVC, SPFMV, and SPVG) based on complete and near-complete sequences (b). Sweet potato mild mottle virus (SPMMV) was included as an outgroup. The trees were generated using the neighbour-joining algorithm and the bootstrap values (1000 replicates) are indicated on the branches.

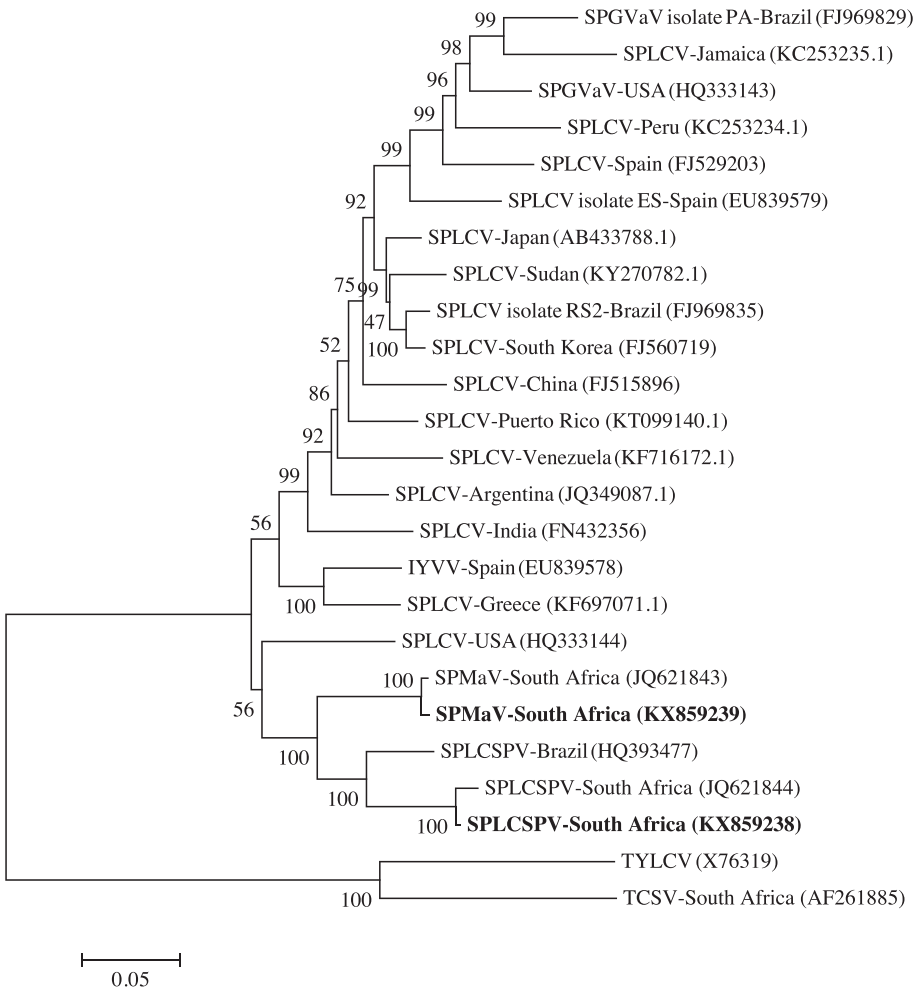


Fig. 4. Neighbour-joining tree showing the phylogenetic relationship of begomovirus isolates from different geographic locations. Bootstrap values (of 1000 replicates) are indicated on the branches. The acronyms denote the following viruses: SPLCV – sweet potato leaf curl virus; SPLCSPV – sweet potato leaf curl Sao Paulo virus; SPMaV - sweet potato mosaic virus; IYVV – Ipomoea yellow vein virus; TYLCV – tomato yellow leaf curl virus and tomato curly stunt virus were included as outgroups.

sweet potato cultivars will respond to infection by six viruses because other viruses (such as SPMaV, SPLCSPV, SPVC and SPVG) could be playing a role in causing severe disease symptoms.

Consistent with previous reports, it is evident from the NGS data that SPFMV remains the most common sweet potato virus in SA, occurring wherever sweet potato is grown. Other studies have also reported SPFMV to be the most widely occurring virus in sweet potato to date (Valverde et al., 2007; Rännäli et al., 2009; Clark et al., 2012). Infection with SPFMV often causes no symptoms (Kreuze and Fuentes, 2008), and SPCSV causes mild symptoms in single infection but when the two viruses are in co-infection, they often cause severe symptoms

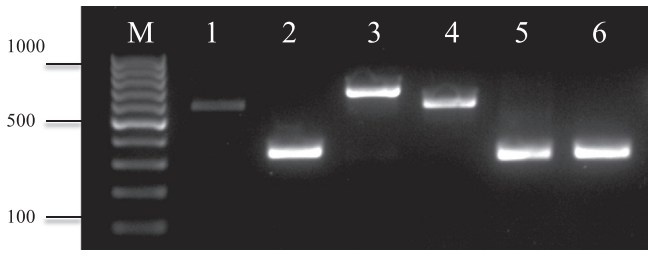


Fig. 5. Agarose gel electrophoresis confirming the presence of the RNA and DNA viruses detected in this study by PCR and RT-PCR. Lane M, 100 bp molecular marker; Lane 1, SPFMV; Lane 2, SPVC; Lane 3, SPVG; Lane 4, SPCSV; Lane 5, SPLCSPV; Lane 6, SPMaV.

(Kreuze et al., 2009). The co-infection results in a synergistic interaction ultimately causing SPVD, the most devastating disease affecting sweet potato (Ateka et al., 2004; Kreuze et al., 2009; Cuellar et al., 2011). There are distinct strain groups within SPFMV; these are the O, EA, RC and C strains (Untiveros et al., 2010). In this study, phylogenetic analysis (Fig. 3) clustered SPFMV isolates into two distinct groups; the O strain and the C strain, now classified as a new *Potyvirus* species (SPVC) (Untiveros et al., 2010). As reported previously (Rännäli et al., 2009; Sivparsad and Gubba, 2012), mixed infections of O and C strains have been detected in sweet potato from other regions of South Africa. The SPFMV-O strain detected in the WC province shared 94% nt identity with the isolate from Japan and SPVC shared 95% nt sequence identity with the isolate from Argentina. The full genome sequence of the South African SPVG isolate shared 98% nt identity with the isolate from Argentina. Limited genetic variability observed between SPVG isolates worldwide suggests geographic distribution by infected vegetative material, and that previous diagnostic tests were not sensitive or available at time of material importation/exportation to detect this virus. Co-infection of potyviruses SPVG and SPFMV (both common (C) and ordinary (O) strains) has been previously reported in French Polynesia, New Zealand, Zimbabwe and South Africa (Rännäli et al., 2009). RNA viruses are prone to variation therefore studies investigating the evolution and adaptability of these viruses are necessary in order to develop effective diagnostic assays and disease control strategies (Rubio et al., 2013). Since this is the first study to generate near

complete reference sequences of South African isolates, we were also able to successfully design oligonucleotide primers to develop diagnostic assays for all the viruses detected.

The SPCSV RNA2 segment assembled in this study is highly conserved and shares 97% nt identity with the Peruvian isolate. The RNA1 segment shares 76% nt identity with the Ugandan isolate. This result suggested a possible reassortment (Hou and Gilbertson, 1996; Savory et al., 2014) between RNA1 from an "unknown" variant of an East African isolate or RNA2 from a West African isolate. It is also possible that reassortment could have occurred between SPCSV RNA1 and another sweet potato virus species. Since fewer full-length sequences of RNA1 and RNA2 segments of the SPCSV are not available in the GenBank, screening for reassortment becomes a challenge. Reassortment events between RNA segments of two different or closely related viruses have been documented in viruses infecting other crops including tomato (Chen et al., 2009) and banana (Hu et al., 2007). Reassortment between virus species, especially of closteroviruses in the family *Closteroviridae*, increases genetic variability and accelerates evolution (Rubio et al., 2013). Genetic diversity observed in SPCSV which belongs to the genus *Crinivirus* within the family *Closteroviridae*, may have arisen from the interaction of mixed viral infections and migration (exchange of sweet potato cuttings) along distant geographical areas (Rubio et al., 2013). Alternatively, reassortment may have occurred in RNA1 with viruses from natural wild hosts. SPFMV, which co-exists often with SPCSV, was detected in 22 *Ipomoea* spp., *Hewittia sublobata*, and *Lepistemon owariensis* in Uganda (Tugume et al., 2008). SPCSV has been found in complexes with viruses such as SPVG, SPV2, SPLCV, SPMMV and cucumber mosaic virus (CMV), where it enhances replication and ultimately increases virus titers by approximately 1000-fold (Valverde et al., 2007; Kreuze and Fuentes, 2008). The interaction of SPCSV with other viruses also exacerbates viral symptoms (Mukasa et al., 2006; Kreuze and Fuentes, 2008), and it has been documented in many cases that SPCSV plays a major role in the enhancement of disease severity (Mukasa et al., 2006; Valverde et al., 2007). The *Hsp70* gene sequence on RNA2 of SPCSV from KwaZulu-Natal (KZN) province in SA is from the West African (WA) strain group (Sivparsad and Gubba, 2012), while the phylogenetic analysis of the full sequenced segments from this study assigned the SPCSV RNA2 from the WC province to the East African (EA) group. This finding suggests that high genetic diversity of SA SPCSV isolates may exist in different sweet potato growing regions. Our study also demonstrates the need to sequence full-length segments of SPCSV in southern and northern Africa in order to further examine the genetic diversity of SPCSV and to identify potential geographic regions where reassortment could occur, as this could lead to the emergence of new strains and increased disease severity.

SPLCSPV and SPMaV have been reported to co-infect sweet potato plants in SA (Esterhuizen et al., 2012). Based on our phylogenetic analysis, begomoviruses may be more widespread in the country, necessitating screening for these viruses in all 9 provinces of SA. Plants infected with *Begomovirus* are often symptomless, however symptoms such as yellow vein, leaf distortion, chlorosis (Kreuze and Fuentes, 2008) and upward curling of young leaves can be associated with infection (Valverde et al., 2007). The recorded field symptoms in our study showed that sweet potato plants infected with begomoviruses displayed upward curling of old and young leaves. The sequence data generated in this study confirmed that plants displaying leaf curl symptoms were indeed infected with begomoviruses. Furthermore, sequence data revealed begomoviruses co-existing with a *Crinivirus* (SPCSV) and a *Potyvirus* (SPFMV). Synergistic interactions of begomoviruses and SPCSV were reported recently (Cuellar et al., 2015). Studies have also shown that single infection with SPCSV or SPFMV causes mild symptoms or no symptoms at all, while con-infection of both viruses causes severe symptoms resulting in sweet potato virus disease (SPVD) (Cuellar et al., 2008). Results from our study infer that SPLCSPV and SPMaV isolates could be contributing significantly to sweet potato disease severity; further experimentation is required to confirm our results. Sequence reads identified as sweet potato caulimo-like virus (SPCV) were detected in the RNA and DNA datasets. Future studies must explore screening more material from the Western Cape and other provinces in SA for the presence of this virus. Disease control strategies, including the use of virus-free cuttings and vector control, should be implemented especially in the WC province commercial farms, to prevent further spread of the viruses and crop decline. It is necessary to conduct a nationwide survey employing NGS in order to a) screen for viruses, b) assemble full-length genomes and c) gain better understanding on virus diversity and virus complexes.

In conclusion, this study describes a metagenomic approach employing the use of high throughput deep sequencing for the detection of RNA and DNA viruses in sweet potato without *a priori* knowledge. This approach clearly reveals the comprehensive profile of the entire viral community in a sample. We established that a survey of two provinces detected 6 viruses in South Africa, including a distinct SPCSV RNA1 sequence. We can also infer that SPCSV, together with SPFMV and begomoviruses, is still a major role player in SPVD.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

This work is based on the research support in part by the National Research Foundation of South Africa (Grant reference number UID 79983). The authors would like to acknowledge our colleagues Sidwell Tjale, Thakhani Ramathavhatha, Sunette Laurie, for assistance with sample collection; Molemi Rauwane and Mr. Thobejane for assistance with glasshouse trials; Jonathan Featherston, Marija Kvas, and Kerry-Anne Pillay for assistance with library preparation.

References

Albuquerque, L.C., Inoue-Nagata, A.K., Pinheiro, B., Ribeiro, S.D.G., Resende, R.O., Moriones, E., Navas-Castillo, J., 2011. A novel monopartite *Begomovirus* infecting sweet potato in Brazil. Archives of Virology 156, 1291–1294.

Albuquerque, L.C., Inoue-Nagata, A.K., Pinheiro, B., Resende, R.O., Moriones, E., Navas-Castillo, J., 2012. Genetic diversity and recombination analysis of sweepoviruses from Brazil. Virology Journal 9, 241.

Aronesty, E., 2013. Comparison of sequencing utility programs. The Open Bioinformatics Journal 7.

Ateka, E., Njeru, R., Kibaru, A., Kimenju, J., Barg, E., Gibson, R., Vetten, H., 2004. Identification and distribution of viruses infecting sweet potato in Kenya. Annals of Applied Biology 144, 371–379.

Boonham, N., Kreuze, J., Winter, S., Van Der Vlugt, R., Bergervoet, J., Tomlinson, J., Mumford, R., 2014. Methods in virus diagnostics: from ELISA to next generation sequencing. Virus Research 186, 20–31.

Brown, J.K., Zerbini, F.M., Navas-Castillo, J., Moriones, E., Ramos-Sobrinho, R., Silva, J.C., Fiallo-Olivé, E., Briddon, R.W., Hernández-Zepeda, C., Idris, A., 2015. Revision of *Begomovirus* taxonomy based on pairwise sequence comparisons. Archives of Virology 160, 1593–1619.

Chen, L., Rojas, M., Kon, T., Gamby, K., Xoconostle-Cazares, B., Gilbertson, R., 2009. A severe symptom phenotype in tomato in Mali is caused by a reassortant between a novel recombinant *Begomovirus* (tomato yellow leaf curl Mali virus) and a betasatellite. Molecular Plant Pathology 10, 415–430.

Clark, C.A., Davis, J.A., Abad, J.A., Cuellar, W.J., Fuentes, S., Kreuze, J.F., Gibson, R.W., Mukasa, S.B., Tugume, A.K., Tairo, F.D., 2012. Sweetpotato viruses: 15 years of progress on understanding and managing complex diseases. Plant Disease 96, 168–185.

Cuellar, W.J., Tairo, F., Kreuze, J.F., Valkonen, J.P., 2008. Analysis of gene content in sweet potato chlorotic stunt virus RNA1 reveals the presence of the p22 RNA silencing suppressor in only a few isolates: implications for viral evolution and synergism. Journal of General Virology 89, 573–582.

Cuellar, W.J., Cruzado, R.K., Fuentes, S., Untiveros, M., Soto, M., Kreuze, J.F., 2011. Sequence characterization of a Peruvian isolate of sweet potato chlorotic stunt virus: further variability and a model for p22 acquisition. Virus Research 157, 111–115.

Cuellar, W.J., Galvez, M., Fuentes, S., Tugume, J., Kreuze, J., 2015. Synergistic interactions of begomoviruses with sweet potato chlorotic stunt virus (genus *Crinivirus*) in sweet potato (*Ipomoea batatas* L.). Molecular Plant Pathology 16, 459–471.

Domola, M., Thompson, G., Aveling, T., Laurie, S., Strydom, H., Van Den Berg, A., 2008. Sweet potato viruses in South Africa and the effect of viral infection on storage root yield. African Plant Protection 14, 15–23.

Esterhuizen, L., Van Heerden, S., Rey, M., Van Heerden, H., 2012. Genetic identification of two sweet-potato-infecting begomoviruses in South Africa. Archives of Virology 157, 2241–2245.

Fauquet, C.M., Bisaro, D., Briddon, R., Brown, J., Harrison, B., Rybicki, E., Stenger, D., Stanley, J., 2003. Virology division news: revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of *Begomovirus* species. Archives of Virology 148, 405–421.

Gibson, R., Mpembe, I., Alicai, T., Carey, E., Mwanga, R., Seal, S., Vetten, H., 1998. Symptoms, aetiology and serological analysis of sweet potato virus disease in Uganda. Plant Pathology 47, 95–102.

Gibson, R.W., Aritua, V., Byamukama, E., Mpembe, I., Kayongo, J., 2004. Control strategies for sweet potato virus disease in Africa. Virus Research 100, 115–122.

Hou, Y.-M., Gilbertson, R.L., 1996. Increased pathogenicity in a pseudorecombinant bipartite geminivirus correlates with intermolecular recombination. Journal of Virology 70, 5430–5436.

Hu, J.-M., Fu, H.-C., Lin, C.-H., Su, H.-J., Yeh, H.-H., 2007. Reassortment and concerted evolution in banana bunchy top virus genomes. Journal of Virology 81, 1746–1761.

Idris, A., Al-Saleh, M., Piatek, M.J., Al-Shahwan, I., Ali, S., Brown, J.K., 2014. Viral metagenomics: analysis of begomoviruses by illumina high-throughput sequencing. Viruses 6, 1219–1236.

Kashif, M., Pietilä, S., Artola, K., Jones, R., Tugume, A., Mäkinen, V., Valkonen, J., 2012. Detection of viruses in sweetpotato from Honduras and Guatemala augmented by deep-sequencing of small-RNAs. Plant Disease 96, 1430–1437.

Kays, S.J., 2004. Sweetpotato production worldwide: assessment, trends and the future. I International Symposium on Root and Tuber Crops: Food Down Under. 670, pp. 19–25.

Kreuze, J., Fuentes, S., 2008. Sweetpotato viruses A2 - mahy. In: Brian, W.J., Regenmortel, M.H.V.V. (Eds.), Encyclopedia of Virology, Third edition Academic Press, Oxford.

Kreuze, J., Savenkov, E., Valkonen, J., 2002. Complete genome sequence and analyses of the subgenomic RNAs of sweet potato chlorotic stunt virus reveal several new features for the genus *Crinivirus*. Journal of Virology 76, 9260–9270.

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R., 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.

Massart, S., Olmos, A., Jijakli, H., Candresse, T., 2014. Current impact and future directions of high throughput sequencing in plant virus diagnostics. Virus Research 188, 90–96.

Mukasa, S.B., Rubaihayo, P.R., Valkonen, J., 2006. Interactions between a *Crinivirus*, an *Ipomovirus* and a *Potyvirus* in coinfected sweet potato plants. Plant Pathology 55, 458–467.

Mwanga, R., Moyer, J., Zhang, D., Carey, E., Yencho, G., 2001. Nature of resistance of sweetpotato to sweetpotato virus disease. I International Conference on Sweetpotato. Food and Health for the Future. 583, pp. 113–119.

Paprotka, T., Boiteux, L., Fonesca, M., Resende, R., Jeske, H., Faria, J., Ribeiro, S., 2010. Genomic diversity of sweet potato geminiviruses in a Brazilian germplasm bank. Virus Research 149, 224–233.

Rännäli, M., Czekaj, V., Jones, R., Fletcher, J., Davis, R., Mu, L., Valkonen, J., 2009. Molecular characterization of sweet potato feathery mottle virus (SPFMV) isolates from Easter Island, French Polynesia, New Zealand, and southern Africa. Plant Disease 93, 933–939.

Rubio, L., Guerri, J., Moreno, P., 2013. Genetic variability and evolutionary dynamics of viruses of the family Closteroviridae. Frontiers in Microbiology 4, 151.

Ruby, J.G., Bellare, P., Derisi, J.L., 2013. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. Genes, Genomes, Genetics 3, 865–880.

Savory, F.R., Varma, V., Ramakrishnan, U., 2014. Identifying geographic hot spots of reassortment in a multipartite plant virus. Evolutionary Applications 7, 569–579.

Sivparsad, B., Gubba, A., 2012. Molecular resolution of the genetic variability of major viruses infecting sweet potato (*Ipomoea batatas* L.) in the province of KwaZulu-Natal in the Republic of South Africa. Crop Protection 41, 49–56.

Stobbe, A.H., Roossinck, M.J., 2014. Plant virus metagenomics: what we know and why we need to know more. Frontiers in Plant Science 5, 150.

Studholme, D.J., Glover, R.H., Boonham, N., 2011. Application of high-throughput DNA sequencing in phytopathology. Annual Review of Phytopathology 49, 87–105.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30, 2725–2729.

Tesfaye, T., Feyissa, T., Abraham, A., 2011. Survey and serological detection of sweet potato (*Ipomoea batatas* (L.) Lam) viruses in Ethiopia. Journal of Applied Biosciences 41, 2746–2756.

Tugume, A., Mukasa, S., Valkonen, J., 2008. Natural wild hosts of sweet potato feathery mottle virus show spatial differences in virus incidence and virus-like diseases in Uganda. Phytopathology 98, 640–652.

Tugume, A.K., Mukasa, S.B., Valkonen, J.P., 2016. Mixed infections of four viruses, the incidence and phylogenetic relationships of sweet potato chlorotic fleck virus (Betaflexiviridae) isolates in wild species and sweetpotatoes in Uganda and evidence of distinct isolates in East Africa. PLoS One 11, e0167769.

Untiveros, M., Quispe, D., Kreuze, J., 2010. Analysis of complete genomic sequences of isolates of the sweet potato feathery mottle virus strains C and EA: molecular evidence for two distinct *Potyvirus* species and two P1 protein domains. Archives of Virology 155, 2059–2063.

Valverde, R.A., Clark, C.A., Valkonen, J.P., 2007. Viruses and virus disease complexes of sweetpotato. Plant Viruses 1, 116–126.

Van Jaarsveld, P.J., Faber, M., Tanumihardjo, S.A., Nestel, P., Lombard, C.J., Benadé, A.J.S., 2005. β-Carotene–rich orange-fleshed sweet potato improves the vitamin a status of primary school children assessed with the modified-relative-dose-response test. The American Journal of Clinical Nutrition 81, 1080–1087.