Diversity of *Citrus tristeza virus* populations in commercial Star Ruby orchards in Southern Africa, using Illumina MiSeq technology.

D.A Read[1*] and G. Pietersen[1, 2]

[1]Department of Microbiology and Plant Pathology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0028, South Africa

[2]Agricultural Research Council—Plant Protection Research Institute, Pretoria 0186, South Africa

*Current address: Department of Microbiology and Plant Pathology, Centre for Microbial Ecology and Genomics (CMEG), University of Pretoria, Pretoria 0028, South Africa

ABSTRACT

Grapefruit cultivars are highly sensitive to CTV infections and in order to increase their productive lifespan, the Southern African citrus industry makes use cross-protection. However, CTV symptoms in the form of stem grooving is commonly observed in commercial grapefruit plantings pre-immunised with CTV GFMS 12, the initial CTV source used. Samples (n = 192) of *Citrus x paradisi* (Macfad.) cv. Star Ruby, pre-immunised with GFMS 12 were collected from six grapefruit production areas of South Africa and Swaziland. Six samples from non-pre-immunised plants were also collected. The p33 gene was amplified and direct Sanger sequencing performed on the resulting amplicons. A sample subset was randomly selected, including three non-pre-immunised samples and subjected to Illumina MiSeq amplicon sequencing. High levels of CTV diversity were observed within trees, orchards and between orchards. Most populations were made up of a dominant component with several minor sequence types. Resistance Breaking (RB) sequences were most numerous, especially in recently planted orchards and were present within all populations. The Kpg3/SP/T3 group was the second most prevalent, with increased prevalence in older

orchards. Sequences mapping to the HA 16-5, VT, AT-1, T36, Taiwan-Pum/M/T5 and T30 references, were represented sporadically within numerous collection sites.

INTRODUCTION

CTV, the largest known plant virus (Karasev et al. 1995) causes one of the most economically devastating disease of citrus (Bar-Joseph et al. 1989) and has become endemic to almost all of the major citrus production areas worldwide (Moreno et al. 2008). The severity of the disease varies from being asymptomatic to displaying severe stem-pitting, reduced fruit size and decline, and depends on the specific infecting CTV strains (Niblett et al. 2000). Grapefruit cultivars are among the most sensitive to CTV infections (van Vuuren and Manicom 2005) and strains known to cause severe stem-pitting are endemic to Southern Africa (van Vuuren et al. 2000).

CTV strains are commonly introduced into an area through the propagation of infected budwood material and the subsequent spread of the virus through aphid vector populations. *Toxoptera citricida* (Kirkaldy) is the most efficient vector (Moreno et al. 2008) and is the also the most important vector species in Southern Africa. It has been shown that areas with endemic populations of *T. citricida* can experience rapid spread of the virus with an up to 95% rate of infection over 2-4 years (Gottwald et al. 1996).

In 1972 the South African Citrus Improvement Program (SACIP) was established and one of its functions was to implement a mild-strain cross-protection scheme for South African citrus (von Broembsen and Lee 1988). Cross-protection in plant viruses is the partial or complete resistance of a plant to infection of a severe strain of a virus after the intentional inoculation of a mild strain of the same virus (Gal-On and Shiboleth

2005; Yoon et al. 2006). Candidate mild-strain sources were obtained from trees that were still highly productive after 15 years (van Vuuren and Collins 1993). After performance trials, the GFMS 12 mild source was used to pre-immunise all grapefruit cultivars. However in 1993, severe stem-pitting symptoms were observed on a number of mother block budwood trees (van Vuuren and Manicom 2005). This prompted the approval of GFMS 35 as an alternate cross-protection source in 1998 (Luttig et al. 2002). Many nurseries however continued to use budwood pre-immunised with GFMS 12 until 2007 (van Vuuren, personal communication).

Six distinct lineages of CTV have been proposed and are now referred to as strains rather than genotypes (Harper 2013). The virus often exists as mixtures of these within a single infection (Scott et al 2012). It has recently been shown that cross-protection (described as super-infection exclusion) is strain-specific (Folimonova et al 2010), which has important implications for commercial cross-protection schemes, since pre-immunising populations may only be able to protect against the same strains that they are themselves made up of. It was shown that a CTV specific gene, p33 and its expressed functional protein are fundamental for the functioning of strain-specific superinfection exclusion (Folimonova et al, 2012; Folimonova 2013).

A number of studies have been done to characterise the pre-immunising populations of GFMS 12 and 35, as well as a number of their sub-isolates (Luttig et al. 2002; Read and Pietersen 2015; Scott et al. 2012; van Vuuren et al. 2000; Zablocki and Pietersen 2014), however, relatively little data is available regarding the CTV population compositions of pre-immunised trees in the field. In this study a total of 192 samples were collected from pre-immunised Star Ruby trees in the production areas of Hoedspruit, Malelane, Swaziland, Northern Cape, Sundays River Valley and Nkwalini Valley, which were all pre-immunised with the GFMS 12 cross-protecting

source. In addition to these, six samples were collected from non-pre-immunised Star Ruby trees in the Letsitele grapefruit production area, only three of which tested positive for the presence of CTV by PCR.

Due to the association of the p33 gene and the mechanism of cross-protection, this gene region was chosen for the analysis of variation in population composition of the collected samples. All of the resulting PCR amplicons from the non-pre-immunised trees were subjected direct Sanger sequencing. A subset of 92 of these was selected for Illumina MiSeq sequencing. This data will assist with the targeted selection of mild-strain sources that are able to protect against strains that are circulating in the field and provide a basis for further research regarding the interaction between CTV strains and their Star Ruby host.

MATERIALS AND METHODS

Collection of material

Leaf material was collected, usually from one side of each tree, along the rows of the orchards. *Table 1* indicates the name of each collection site, the GPS coordinates within the orchard, the number of samples collected and the date of collection. The exact position of each sampled tree was mapped, using a differential GPS (Trimble Geo, Trimble Navigation, Sunnyvale, CA, USA).

Six samples were collected, toward the end of 2014, from biological replicates of Star Ruby trees that had been planted CTV-free (not pre-immunised) in 2007, in the production area of Letsitele.

Table 1: Details of the sites from which samples were collected for characterisation of the CTV populations. Each site was named according to its geographical location. The GPS coordinates are the points close to the centre of the orchard that was sampled.

Collection site	GPS coordinates	Number of samples collected	Year that orchard was planted	Date of collection
Hoedspruit 1	24°24'20.15''S 30°49'47.5''E	16	1996	Feb 2013
Hoedspruit 2	24°24'20.15''S 30°49'47.5''E	16	2007	Feb 2013
Malelane 1	25°32'18.84''S 31°26'26.92''E	16	1990	Feb 2013
Malelane 2	25°25'26.69''S 31°40'32.71''E	16	1999	Feb 2013
Swaziland (Mananga)	25°56'28.02''S 31°42'58.24''E	32	1995	Feb 2013
Northern Cape (Kakamas)	28°47'45.18''S 20°40'06.24''E	32	2002	Mar 2013
Sundays River Valley 1	33°28'29.55''S 25°35'21.28''E	16	1990	Jun 2013
Sundays River Valley 2	33°28'40.13''S 25°35'17.93''E	16	2004	Jun 2013
Nkwalini Valley	28°42'17.75''S 31°40'26.19''E	32	1991	Jul 2013

RNA isolation, reverse transcription and PCR amplification

RNA extractions were carried out with a GeneJET plant RNA isolation kit (Thermo, Vilnius, Lithuania) using 100mg of plant material that had been macerated in liquid nitrogen. The amplification of the p33 gene of each sample was carried out using a two-step RT-PCR as follows. The reverse transcription reaction was primed using 50pmol of the Univ-p33-R reverse primer (5' CCCGTTTAAACAGAGTCAAACGG 3'), 12µl of total RNA extracts, 50pmol of reverse primer, 10U avian myeloblastosis virus (AMV) reverse transcriptase (Roche, Mannheim, Germany), 5U RNase inhibitor (Roche, Mannheim, Germany), 1x RT buffer (Roche, Mannheim, Germany) and 0.1 mM dNTP mix (Thermo, Vilnius, Lithuania). Reactions were held at 42°C for 1 hour.

The PCR reactions for the amplification of the p33 gene consisted of 10µl Promega GoTaq Hot Start Green Master Mix (Promega, Madison, WI, USA), 0.375µl Univ-p33-F forward primer (10µM) (5' GATGTTTGCCTTCGCGAGC 3'), 0.375µl Univ-p33-R reverse primer (10µM), 2µl of cDNA product and molecular grade water to a total volume of 20µl. PCR reactions conditions were; 1 cycle of 92°C for 2min, 35 cycles of 92°C for 30s, 65°C for 45s, 72°C for 1min and 1 cycle of 72°C for 10min. The final PCR products were visualised on an agarose gel.

Direct sequencing of p33 gene amplicons

Amplicons were purified for sequencing by adding 2µl FastAP (1U/ µl) and 0.5µl (20U/ µl) ExoI enzymes (Thermo, Vilnius, Lithuania) to 19µl of the PCR product and incubated at 37°C for 15 min. This was followed by an 85°C for 15 min incubation. Sanger sequencing reactions on 2µl of the purified PCR was by adding 1µl BigDye® Terminator mix v3.1 (Applied Biosystems, Foster City, CA, USA), 2.25µl 5x BigDye® v3.1 sequencing buffer, 0.75µl Univ-p33-F primer (2µM) and molecular grade water to a total volume of 10µl and using 1 cycle of 94°C for 1 min, 30 cycles of 94°C for 10

seconds, 50°C for 5 seconds and 60°C for 4 minutes. Sequencing products were purified using ethanol precipitation, according to Sambrook (2001). The purified sequencing products were submitted to the African Centre for Gene Technologies (ACGT), Automated Sequencing Facility, Department of Genetics, University of Pretoria, South Africa and sequenced using an ABI Prism® 3500xl Genetic Analyser (Applied Biosystems, Foster City, CA, USA). Sequences not conforming to a quality criterion of a minimum PHRED score of 30 were discarded from further analysis.

Phylogenetic analysis of the direct sequence data

Chromas Lite 2.1 (Technelysium, Brisbane, Australia) was used to edit and correct errors in chromatograms. Alignments of sequences were carried using the CLUSTAL W alignment software (EBI, Cambridgeshire, England) within the BioEdit Sequence alignment editor 7.1.3 (Hall 1999). The cognate p33 gene region from a total of 45 full-genome reference sequences were accessed from GenBank (www.ncbi.nlm.nih.gov/genbank) and include accession numbers (strain names in brackets): NC_001661 (T36); AY 340974 (Qaha); U16304 (T36); DQ272579 (Mexico); AY170468 (T36); EU937521 (T36); KC517485 (FS674-T36); KC517486 (FS701-T36); KC517487 (FS703-T36); KC517488 (FS577); JX266713 (Taiwan-Pum/M/T5); AF001623 (SY568); AF260651 (T30); Y18420 (T385); KC517489 (FS701-T30); KC517490 (FL278-T30); KC517491 (FS703-T30); JF957196 (B301); FJ525432 (NZRB-G90); GQ454869 (HA 18-9); FJ525435 (NZRB-M17); JX266712 (Taiwan-Pum/SP/T1); FJ525431 (NZRB-M12); FJ525433 (NZRB-TH28); FJ525434 (NZRB-TH30); JQ798289 (A18); KC525952 (T3); HM573451 (Kpg3); EU857538 (SP); GQ454870 (HA 16-5); DQ151548 (T318A); AB0463981 (NUagA); JQ911664 (CT11A); KC517493 (FL202-VT); U56902 (VT); KC517492 (FS703-VT); EU937519

(VT); KC517494 (FS701-VT); KC262793 (L192GR); JQ911663 (CT14A); FJ525436 (NZ-B18); JQ965169 (T68); EU076703 (B165); JQ061137 (AT-1) and KC333869 (CT-ZA3). Neighbour-joining phylogenetic trees were constructed for each alignment, using MEGA 6 (Tamura et al. 2013) and the Maximum Composite Likelihood substitution model with a 1000 bootstrap replicates.

Illumina MiSeq sequencing

The p33 amplicon from a subset of 92 samples were subjected to Illumina MiSeq sequencing (Illumina, San Diego, CA, United States). Paired-end DNA libraries were prepared using the Nextera V2 sample kit (Epicentre, Madison, WI, United States). The samples were sequenced at the Agricultural Research Council (ARC), Biotechnology Platform, Pretoria, South Africa.

MiSeq data analysis

All trimming and analyses of the Illumina MiSeq datasets was carried out using CLC Genomics workbench 5.5.1 (CLC, Aarhus, Denmark). Data was imported as paired-end reads with a distance range of 180-300. Adapter and quality trimming was performed using the default program settings with Nextera V2 transposase adapter sequences (Transposase1: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG; Transposase2: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG). Quality control was carried out using the Fast QC function. Datasets were mapped to the cognate p33 region of a set of reference sequences. Nine p33 clades/lineages, referred to here as “p33 sequence types” were defined and each of these consisted of the following reference sequences. **RB**: F957196 (B301); FJ525432 (NZRB-G90); GQ454869 (HA 18-9); FJ525435 (NZRB-M17); JX266712 (Taiwan-Pum/SP/T1); FJ525431 (NZRB-M12); FJ525433 (NZRB-TH28); FJ525434 (NZRB-TH30). **Kpg3/SP/T3**: HM573451

(Kpg3); EU857538 (SP); KC525952 (T3). **HA 16-5**: GQ454870 (HA 16-5). **VT**: JQ911664 (CT11A); KC517493 (FL202-VT); U56902 (VT); KC517492 (FS703-VT); EU937519 (VT); KC517494 (FS701-VT); KC262793 (L192GR); DQ151548 (T318A); AB0463981 (NUagA); KC333869 (CTZA3). **AT-1**: JQ061137 (AT-1). **T36**: NC_001661 (T36); AY 340974 (Qaha); U16304 (T36); DQ272579 (Mexico); AY170468 (T36); EU937521 (T36); KC517485 (FS674-T36); KC517486 (FS701-T36); KC517487 (FS703-T36); KC517488 (FS577). **Taiwan-Pum/M/T5**: JX266713 (Taiwan-Pum/M/T5). **T30**: AF001623 (SY568); AF260651 (T30); Y18420 (T385); KC517489 (FS701-T30); KC517490 (FL278-T30); KC517491 (FS703-T30). **B165**: JQ911663 (CT14A); JQ965169 (T68); EU076703 (B165); FJ525436 (NZ-B18).

Threshold to account for index leaching

The 16S RNA sequence of "Candidatus" *Liberibacter africanus* was amplified using the OA1 forward (5' GCGCGTATTTTATACGAGCGGCA 3') and the OI2c reverse primer (5' ATGGGTTGCGAAGTCGCGAGGC 3') (Teixeira et al. 2005), with 1 cycle of 92°C for 2min, 40 cycles of 92°C for 30s, 65°C for 45s, 72°C for 1.5min and 1 cycle of 72°C for 10min. The amplicon was gel purified, quantified and prepared for Illumina sequencing, as described previously and sequenced on the same flow cell in parallel with the CTV field survey samples.

RESULTS

Direct sequencing of p33 gene amplicons

A total of 130 samples yielded direct sequences of an acceptable quality, and were subjected to phylogenetic analysis (dendrograms not shown). The strain identities of

Table 2: Dominant CTV genotype present in samples obtained through direct sequencing of the p33 gene PCR amplicons. Numbers represent the number of sequenced samples with the dominant sequence grouping with the respective reference sequences within a Neighbour-joining phylogeny.

		Number of samples with sequences obtained through direct sequencing grouping with respective reference sequences								
Collection site	Age of orchard at time of collection (years)	RB	Kpg3/ SP/T3	AT-1	No ref (T36)	VT	HA 16-5	T36	Taiwan-Pum/ M/T5	T30
Hoedspruit 1	18	10	-	2	-	-	-	-	-	-
Hoedspruit 2	8	14	-	-	-	-	-	-	-	-
Malelane 1	24	2	9	-	-	-	-	-	-	-
Malelane 2	14	4	6	3	-	-	-	-	-	-
Swaziland (Mananga)	18	4	15	-	1	-	-	-	-	-
Northern Cape (Kakamas)	12	21	-	-	-	-	-	-	-	-
Sundays River Valley	9	4	-	-	-	-	-	-	-	-
	23	12	4	-	-	-	-	-	-	-
Nkwalini Valley	23	3	15	-	1	-	-	-	-	-

the predominant sequences obtained by direct sequencing (Table 2) were determined through their clustering with the reference sequences. Kpg3 (HM573451), SP (EU857538) or T3 (KC525952) cannot be resolved with the p33 gene and any direct sequencing data falling within this branch is referred to as Kpg3/SP/T3-like. The Nkwalini Valley sequences showed one of the samples (13-3747) grouping away from any references. This unique branch represents that of 13-3427, suggesting that these samples may represent a previously undescribed strain.

Illumina MiSeq sequencing data

The percentage of total Illumina MiSeq reads mapping to the members of various reference p33 clades is presented in *Table 3*. The RB sequence type was the most prevalent and occurred in all datasets, while the Kpg3/SP/T3-type sequence was the second most numerous and was found to be present in all sample datasets except for eight from Hoedspruit 2, eight from Northern Cape and five from Sundays River Valley. VT was detected in samples from all collection sites except Malelane 1, generally at low percentages of the total mapped reads. A few sample datasets contained sequences represented by some of the other p33 clades at a number of collection sites such as AT-1, T36, Taiwan-Pum/M/T5 and T30.

Only three of the six samples from non-pre-immunised trees, collected in Letsitele, yielded amplicons from the p33 gene PCR. The results of the read mappings for these populations are shown in *Table 4*. Seven of the p33 sequence types were present across all three samples. The dominant p33 sequence type in sample 14-6003 was VT-like, with AT-1, CT14A, RB and A18 representing minor components. In sample 14-6004 RB-like p33 sequences were dominant and AT-1 a minor component. In

Table 3: Percentage of total reads of p33 gene sequences from Illumina MiSeq of citrus samples mapping to reference sequences representing nine p33 sequence clades/lineages. The percentages of mapped reads do not add up to exactly 100% in all cases due to the discarding of a small percentage of reads in each dataset, due to their potential origin from index leaching.

Collection site	Age of orchard at time of collection (years)	Sample number	Total number of reads mapping to refs	RB	Kpg3 / SP/ T3	HA 16-5	VT	AT-1	T36	Taiwan-Pum/ M/T5	T30	B165
Hoedspruit 1	18	13-3023	92768	99.6	0.15	-	-	-	-	-	-	-
		13-3067	174009	98	0.1	-	0.57	0.2	-	0.86	-	-
		13-3040	12421	95.5	0.1	-	2.9	1.1	-	-	-	-
		13-3042	73495	92.4	0.4	-	3.9	0.4	-	2.3	-	-
		13-3015	181749	91.2	7.5	0.4	-	-	-	0.1	-	-
		13-3059	133533	73.3	0.1	-	25.7	-	-	0.2	-	-
		13-3013	57627	10.9	78.8	6.5	0.2		3.1	0.3		-
		13-3062	327014	2.7	-	-	14.5	82.2	-	-	-	-
Hoedspruit 2	8	13-3113	144570	100	-	-	-	-	-	-	-	-
		13-3160	57211	99.9	-	-	-	-	-	-	-	-
		13-3152	109041	99.9	-	-	-	-	-	-	-	-
		13-3161	58759	99.8	-	-	-	-	-	-	-	-
		13-3105	190405	99.7	-	-	-	-	-	-	-	-
		13-3155	53385	99.6	-	-	-	-	-	-	-	-
		13-3125	25583	97.9	-	-	2	-	-	-	-	-
		13-3130	143871	34	62.1	-	3.8	-	-	-	-	-
Malelane 1	24	13-3251	69030	90.5	9	0.4	-	-	-	-	-	-
		13-3242	104646	90.2	8.5	1.2	-	-	-	-	-	-
		13-3271	74353	48.3	23.6	27.8	-	-	-	-	-	-
		13-3257	19511	35.8	58.7	5.3	-	-	-	-	-	-
		13-3260	152311	11.4	85.1	3.4	-	-	-	-	-	-
		13-3267	149687	9.4	86.2	4.3	-	-	-	-	-	-
		13-3205	80987	8.9	87.2	3.8	-	-	-	-	-	-
		13-3255	318511	6.4	76.4	17	-	-	-	-	-	-
Malelane 2	14	13-3321	129010	99.7	0.1	-	-	-	-	-	-	-
		13-3312	158354	99.1	0.1	-	-	-	-	-	-	-
		13-3305	39568	86.2	12	0.7	0.3	0.6	-	-	-	-
		13-3362	78835	22.5	4.8	0.7	23.2	48.4	-	-	0.1	-
		13-3324	87079	22.8	23.8	1.9	20.8	30.5	-	-	-	-
		13-3309	70784	12.6	71.3	4.1	4.9	7.1	-	-	-	-
		13-3343	125140	2.7	92.6	4.7	-	-	-	-	-	-
		13-3347	284891	1.6	92.7	5.4	-	-	-	-	-	-
Swaziland (Mananga)	18	13-3410	51624	95.1	2.2	0.2	-	-	2.3	-	-	-
		13-3445	62214	88.7	8.6	0.3	2.2	-	-	-	-	-
		13-3436	210778	81.2	16.9	1	0.5	0.1	-	-	-	-
		13-3416	105078	80.8	12.6	1.4	4.9	-	-	-	-	-
		13-3421	50038	73.2	15.7	0.8	0.5	-	9.5	-	-	-
		13-3427	57164	66.9	2.3	-	-	-	30.1	0.1	-	-
		13-3439	258754	61.5	35.4	2.8	-	-	-	-	-	-
		13-3447	95957	32.1	29.8	37.9	-	-	-	-	-	-
		13-3422	258480	24.2	62.6	12.8	-	-	-	0.1	-	-
		13-3406	103216	10.7	74.7	9.9	0.2	-	4.2	-	-	-
		13-3418	541918	8.3	89.6	2	-	-	-	-	-	-
		13-3425	86853	8	81.2	10.1	-	-	-	-	-	-
		13-3454	35808	3.5	71.8	24.7	-	-	-	-	-	-
		13-3417	231254	2.2	87.7	9.9	-	-	-	-	-	-

		13-3449	92851	1.8	93.1	5	-	-	-	-	-	-
		13-3405	107866	1.5	44.8	5.7	47.7	-	-	-	-	-
Collection site	Age of orchard at time of collection (years)	Sample number	Total number of reads mapping to refs	RB	Kpg3 / SP/ T3	HA 16-5	VT	AT-1	T36	Taiwan-Pum/ M/T5	T30	B 165
Northern Cape (Kakamas)	12	13-3553	28218	99.9	-	-	-	-	-	-	-	-
		13-3545	352081	99.9	-	-	-	-	-	-	-	-
		13-3512	212758	99.7	-	-	-	-	-	-	-	-
		13-3556	123561	99.7								-
		13-3520	105466	99.6	-	-	-	-	-	0.2	-	-
		13-3570	81865	98.7	0.4	0.2	0.2	-	0.2	-	-	-
		13-3536	103234	97.9	-	-	1.6	-	-	0.2	-	-
		13-3552	48994	97.8	-	-	-	-	-	-	-	-
		13-3508	109759	94.5	-	-	4.6	-	-	0.5	-	-
		13-3526	45601	88	8.8	1.6	1.4	-	-	-	-	-
		13-3539	55325	79.9	18.6	0.9	0.4	-	-	-	-	-
		13-3523	61645	75.3	20.4	3.5	0.7	-	-	-	-	-
		13-3534	32842	29.2	66.3	3.7	0.7	-	-	-	-	-
Sundays River Valley	9	13-3651	120438	99.5	0.1	-	-	-	-	0.2	-	-
		13-3674	117329	98.5	-	-	-	-	-	1.3	-	-
		13-3655	106127	94.7	-	-	0.9	3.9	-	0.2	-	-
		13-3658	149419	57.9	35.3	6.4	-	-	-	0.2	-	-
		13-3661	329326	2.3	-	-	19.1	78.3	-	-	-	-
	23	13-3621	127293	99.5	-	-	-	-	-	0.2	-	-
		13-3637	196489	98.9	-	-	0.1	-	-	0.3	0.4	-
		13-3643	26209	97.8	1.3	0.5	0.3	-	-	-	-	-
		13-3642	34903	97.4		-	1.8	-	-	-	-	-
		13-3608	67631	97.1	0.3	-	2	-	-	-	-	-
		13-3612	52122	89.1	5.6	3.2	1.2	0.8	-	-	-	-
		13-3647	57568	78.4	12.2	8.1	1.1	-	-	-	-	-
		13-3627	180682	63.3	28.5	5.9	1.9	-	-	-	-	-
		13-3619	110849	27.8	68.8	3.1	-	-	-	-	-	-
		13-3626	87694	8.3	91.1	0.4	-	-	-	-	-	-
		13-3604	110345	4.1	79.4	16.5	-	-	-	-	-	-
Nkwalini Valley	23	13-3768	66898	93.3	6.1	0.5	-	-	-	-	-	-
		13-3776	56981	86.6	8.8	4.2	-	-	-	-	-	-
		13-3770	98996	83.2	13.4	1.4	1.6	-	-	-	-	-
		13-3747	87318	76.8	3.2	0.1	-	-	19.3	-	-	-
		13-3719	37751	51.4	2.7	0.2	-	-	45.6	-	-	-
		13-3743	80819	26.8	32.3	40.7	-	-	-	-	-	-
		13-3705	109482	24.5	29.4	45.7	0.3	-	-	-	-	-
		13-3714	136211	20.7	79.2	-	-	-	-	-	-	-
		13-3745	93095	20	74.5	5.5	-	-	-	-	-	-
		13-3732	348076	19.2	77.3	3.3	-	-	-	-	-	-
		13-3713	271667	18	50.7	31.1	-	-	-	-	-	-
		13-3748	290491	16.9	79.1	3.9	-	-	-	-	-	-
		13-3729	345812	9.2	65.7	24.9	-	-	-	-	-	-
		13-3708	213313	7.9	87.4	4.2	-	-	0.4	-	-	-
		13-3727	292114	4.3	93	2.5	-	-	-	-	-	-

Table 4: Population compositions of the samples that were collected from a trial that was planted virus-free in the Letsitele citrus production area.

Collection site	Age of orchard at time of collection (years)	Sample number	Total number of reads mapping to refs	RB	Kpg3 / SP/ T3	HA 16-5	VT	AT-1	T36	Taiwan-Pum/ M/T5	T30	A 18	B1 65
Letsitele	7	14-6003	278055	0.3	-	-	81	16.3	-	-	-	0.2	1.7
		14-6004	378037	99.6	-	-	-	0.4	-	-	-	-	-
		14-6005	451572	81.5	3	0.1	0.4	11.1	-	-	-	0.4	0.9

sample 14-6005 RB-like components were dominant while AT-1 and the remaining p33 sequence types being minor components.

Threshold to account for index leaching

The total number of reads mapping back to the 16S rRNA Liberibacter africanus reference sequence (EU921620) not amplified by the CTV p33 gene PCR, varied between 0% and 0.1% among all of the samples and represents the levels of indexing leaching hence all reads at levels of 0.1% and lower were considered to possibly be the result of index leaching amongst samples and were therefore eliminated from the dataset.

DISCUSSION

In spite of all trees sampled being Star Ruby grapefruit and all being pre-immunised with a CTV GFMS 12 population (with the exception of six) considerable variation in CTV genotypes were observed between orchards as well as between different trees within the same orchard. Of the 192 samples analysed using direct sequencing of amplicons of the p33 gene, evidence of populations with a clearly prevalent (Pawlotsky et al. 1998) p33 gene genotype was obtained in 130 samples. Sixty-two samples had low PHRED scores and multiple nucleotide traces indicative of multiple sequencing templates (Fontana et al. 2014), and hence at least two p33 gene genotypes in the population at similar levels. The following genotypes were found to be clearly dominant amongst the 130 CTV populations: RB in 72 samples (56.7%), a member of the Kpg3/SP/T3 genotypes in 48 samples (37.8%), AT-1 in 5 samples (3.9%), and T36 in 2 samples (1.6%).

Due to the large number of samples with evidence of multiple p33 gene genotypes in the population, amplicons of 92 samples were randomly selected from the pool of 192 and analysed using the Illumina MiSeq platform. High levels of CTV p33 gene genotype diversity were observed. From the 92 analysed, 76 (82.6%) had sequence reads mapping to specific p33 gene genotypes that comprised at least 70% or more of the total CTV sequences generated for that sample. Generally older trees contained more heterogeneous CTV p33 gene populations

Illumina sequence reads mapping to the p33 gene of reference strains within the RB clade were present in all of the populations analysed, being the dominant strain in 54 of these. They often represented more than 80% of total reads obtained. This supports the results obtained with direct sequencing and confirms the fact that it is a component of the original GFMS 12 pre-immunising population (Scott et al. 2012). The dominance of RB components in two of the three samples that were planted virus free reflects the high inoculum pressure of RB genotypes within Southern Africa.

Reads mapping to reference sequences within the Kpg3/SP/T3 p33 gene clade were also present in most (76%) of the CTV p33 gene populations from pre-immunised trees and were the dominant p33 gene variant in 30% of the CTV population analysed. The apparent absence of these components in almost a quarter of the pre-immunised populations, its absence in the majority of the youngest trees, especially at the Hoedspruit 2 collection site (8 years old) and the trend that it generally becomes more prevalent and also attains dominance only in older trees suggests that it was introduced into these populations through infection post- pre-immunisation by aphids in the field. More than half of the CTV populations from trees in 23 and 24 year old orchards in Nkwalini valley and Malelane 1 were dominated by Kpg3/SP/T3 p33 gene reads forming up to 93% of total CTV p33 gene reads obtained in the population.

The p33 gene region does not resolve between the Kpg3 (strain within VT genotype), SP (strain within T3 genotype) or T3 genotype and as incongruences exist between different genomic region of CTV strains (Harper 2013) it is not possible to relate these sequences back to the presence of the “VT-like” component obtained in the GFMS 12 pre-immunizing source which had been characterised using the A fragment of ORF1a gene and the p23 gene (Scott et al. 2012).

Reads mapping to the p33 gene of the HA 16-5 reference sequence were next most prevalent within the CTV populations analysed, with 58 out of the 92 pre-immunised trees being positive for this genotype. However, the relatively high number of populations lacking HA 16-5 p33 gene reads, its tendency to accumulate to higher levels in older trees and the regional distribution (29 of the 32 populations from the geographically close Malelane and Swaziland being infected by this strain, yet only 2 of the 16 populations from Hoedspruit being infected with this strain), suggests that it has also been introduced into pre-immunised Star Ruby trees through aphid transmission in the field.

Members from the VT clade (according to p33 gene delineation) were represented in just over a third of the populations that were analysed and occurred sporadically in all but one of the sample sites. Generally, VT associated p33 gene reads were represented at low levels in each dataset (<5%) however in 5 populations were between 19.1 and 47.7%. VT reads constituted 81% of the total reads in one out of the three non-pre-immunised populations. While VT-like sequences generally occurred more in older CTV populations it was completely absent from trees sampled at one the oldest orchard sites (Malelane 1).

AT-1, T36, Taiwan-Pum/M/T5 and T30 were detected at varying levels within a small number of populations, usually within specific sample sites and probably do not

play a major role in the overall CTV pathogenicity within Southern African Star Ruby plantings. AT-1 was detected in twelve samples from Hoedspruit 1, Malelane 2 and Sundays River Valley. In most of these populations AT-1 p33 gene associated reads were present at low percentages of the total reads, however four populations showed levels of between 30.5 and 82.2%. Reads of T36 were detected in only 9 CTV populations and in only three exceeding levels of 10% of the total, with one originating from Swaziland (30.1%) and two from Nkwalini Valley (19.3% and 45.6%). Based on the sequence for T36 sources obtained following direct sequencing, these form a unique branch in dendrograms and warrant further characterisation to determine whether they represent a previously undescribed genotype. The total absence of B165-like sequences in GFMS 12 pre-immunised trees was unexpected, since Zablocki and Pietersen (2014) found that sub-isolates of GFMS 12 comprised were almost homogenous for the CTZA-3 strain, which when considering the p33 gene sequence only, is indistinguishable from B165 and T68. The absence of CTZA-3, B165 or T68 in the field could be the result of host or environmental selection pressures or vector selection during the production of the sub-isolates.

Three samples collected from seven year old Star Ruby trees in the Letsitele production area, had not been subjected to pre-immunisation and were CTV free at the time of planting. CTV genotypes detected in these samples would have been introduced solely by aphid transmission of CTV in the field. The CTV genotype composition of these populations vary considerably from one another although they all have a dominant sequence component, representing over 80% of the total CTV sequence reads. Two of the populations were dominant for RB-like CTV sequences, which, is expected since these components were probably present at high levels within the surrounding pre-immunised populations.

While there was a general hierarchy in terms of the overall representation of strains, there was significant variation in terms of strain representation between specific collection sites. As little as three of the clades were represented at Hoedspruit 2 and Malelane 1, which were the youngest and oldest tested sites respectively, while overall diversity increased to eight of the clades being represented at Sundays River Valley. Generally collection sites showed overall diversity at levels between these extremes, with seven groups in Hoedspruit 1 and Swaziland, six in Malelane 1 and Northern Cape and five in Nkwalini Valley. In view of an increased period of orchard exposure to field conditions, it would have been expected that Malelane 1 would have the greatest diversity of clades and not just the three observed. It is possible that the populations at the Malelane 1 site were under very little re-infection pressure could have reached equilibrium. The possibility of CTV population equilibrium was shown by Harper et al. (2015), where components within populations of CTV reach equilibrium over time, regardless of the relative titres of the initial inoculum or whether certain components are introduced into the population subsequently. Our study however suggests that population equilibrium was not reached even after periods of more than two decades. The lack of equilibrium within Southern African Star Ruby populations could be due to the large numbers of genotypes present, which may delay equilibrium with the host potentially interacting differently with the respective genotype combinations. Also, repeated aphid inoculations over many years, from a diverse pool of inoculum, could result in the populations being in a state of flux, never attaining equilibrium.

Future research should include efforts to isolate RB, Kpg3/SP/T3, HA 16-5 and VT components from the above sources in order to determine various biological properties

including their importance in eliciting disease and to sequence their entire genome. This will assist in prioritising targets for cross-protection.

Direct Sanger sequencing provides the identity of the dominant component only (Pawlotsky et al. 1998) and where populations are composed of various disparate strains, will yield ambiguous base calls (Fontana et al. 2014). In this study, direct sequence data not adhering to stringent quality criteria were assumed to be a result of populations consisting of more than one detectible sequence type and only the identities of those appearing to contain a seemingly single, clearly dominant sequence are reported. In addition to sometimes missing critical components of populations, direct sequencing may also misidentify the dominant sequence type in populations that are heterogeneous. While Mullan et al. (2001) advises against using direct sequencing for the characterisation of viral populations, direct sequencing of CTV amplicons may provide a useful pre-screening of samples prior to more in depth analyses.

A total of fifty-two samples were identified as having both useful direct sequencing data (supported by PHRED scores) and Illumina sequencing data. Amongst these forty-six had Illumina data that supported those of direct sequencing data with regards the dominant genotype. In only six populations were differences between the direct sequencing and Illumina data regarding the dominant p33 gene observed. In the most extreme case, direct sequencing suggested the prevalence of a sequence type that was represented by only 23.6% of CTV Illumina MiSeq reads. Hence, while direct sequencing appears to generally give a good indication of dominant components, all minor components will be overlooked and therefore sequencing of CTV populations by Illumina is preferable.

In establishing the Illumina-based p33 population assay, a concurrent study was carried out to determine the primer-associated PCR amplification bias associated with a number of previously published primer sequences, including those targeting the p33 gene used in this survey (Read and Pietersen, *unpublished*). While a significant degree of bias was found to be associated with the primers targeting the p33 gene, particularly against members of the RB clade, this clade was the most prevalent one obtained during the survey. It can be assumed that the bias did not materially affect the results of the survey, possibly only underestimating the level of dominance of the RB clade. Therefore the data represented are most likely a true reflection of the CTV populations analysed in this survey. However to mitigate potential inaccuracies obtained Read and Pietersen, (*unpublished*) have provided guidelines for improving the p33 gene primer pair for future use, which should further reduce the potential for PCR amplification bias.

ACKNOWLEDGMENTS

We gratefully acknowledge funding from Citrus Research International (CRI), Agricultural Research Council - Plant Protection Research Institute (ARC-PPRI) and the National Research Foundation (NRF) - Technology and Human Resources for Industry Programme (THRIP) program. We also acknowledge the numerous producers that allowed samples to be collected from their orchards, and we wish to thank Dr. Fanie van Vuuren, CRI, for his continued support and advise during the course of this project.

LITERATURE CITED

Bar-Joseph, M., Marcus, R., and Lee, R.F. 1989. The continuous challenge of Citrus Tristeza Virus Control. Annu. Rev. Phytopathol. 27:291-316

Folimonova, S.Y., Robertson, C.J., Shilts, T., Folimonov, A.S., Hilf, M.E., Garnsey, S.M. and Dawson, W.O. 2010. Infection with Strains of *Citrus Tristeza Virus* Does Not Exclude Superinfection by Other Strains of the Virus. J. Virol. 84:1314–1325

Folimonova, S.Y. 2012. Superinfection Exclusion Is an Active Virus-Controlled Function That Requires a Specific Viral Protein. J. Virol. 86:5554-5561

Folimonova, S.Y. 2013. Developing an understanding of cross-protection by *Citrus tristeza virus*. Front. Microbiol. 4:76

Fontana, A., Debreczeni, D.E., Albanese, G., Davino, S., Flores, R., and Rubio, L. 2014. Evolutionary analysis of *Citrus tristeza virus* outbreaks in Calabria, Italy: two rapidly spreading and independent introductions of mild and severe isolates. Eur. J. Plant. Pathol. 140:607-613

Gal-On, A., and Shiboleth, Y.M., 2005. Cross-protection. Pages 261-288 in: Natural resistance mechanisms of plants to viruses. G. Loebenstein, and J.P. Carr, eds. Springer, Dordrecht, The Netherlands.

Gottwald, T.R., Garnsey, S.M., Cambra, M., Moreno, P., Irey, M., and Borbón, J. 1996. Differential Effects of *Toxoptera citricida* vs. *Aphis gossypii* on Temporal

Increase and Spatial Patterns of Spread of Citrus Tristeza. . In: Proc. 13th Conf. IOCV. IOCV, Riverside, California. 120-129

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic. Acids. Symp. Ser. 41:95-98

Harper, S.J. 2013. Citrus tristeza virus: evolution of complex and varied genotypic groups. Front. Microbiol 4:93

Harper, S.J., Cowell, S.J., and Dawson, W.O. 2015. Finding balance: Virus populations reach equilibrium during the infection process. Virology. 485:205-212

Karasev, A.V., Boyko, V.P., Gowda, S., Nikolaeva, O.V., Hilf, M.E., Koonin, E.V., Niblett, C.L., Cline, K., Gumpf, D.J., Lee, R.F., Garnsey, S.M., Lewandowski, D.J., and Dawson, W.O. 1995. Complete Sequence of the Citrus Tristeza Virus RNA Genome. Virology. 208:511-520

Luttig, M., van Vuuren, S.P., and van der Vyver, J.B. 2002. Differentiation of single aphid cultured sub-isolates of two South African *Citrus tristeza virus* isolates from grapefruit by single-strand conformation polymorphism. In: Proc 15th Conf. IOCV, IOCV, Riverside, California. 186-196

Moreno, P., Ambrós, S., Albiach-Marti, M.R., Guerri, J., and Peña, L. 2008. *Citrus tristeza virus*: A pathogen that changed the course of the citrus industry. Mol. Plant. Pathol. 9:251-268

Mullan, B., Kenny-Walsh, E., Collins, J.K., Shanahan, F., and Fanning, L.J. 2001. Inferred Hepatitis C Virus Quasispecies Diversity Is Influenced by Choice of DNA Polymerase in Reverse Transcriptase-Polymerase Chain Reactions. Anal. Biochem. 289:137-146

Niblett, C.L., Genc, H., Cevik, B., Halbert, S., Brown, L., Nolasco, G., Bonacalza, B., Manjunath, K.L., Febres, V.J., Pappu, H.R., and Lee, R.F. (2000). Progress on strain differentiation of *Citrus tristeza virus* and its application to the epidemiology of citrus tristeza disease. Virus Research vol. 71 pp97-106

Pawlotsky, J.M., Germanidis, G., Neumann, A.U., Pellerin, M., Frainais, P.O., and Dhumeaux, D. 1998. Interferon Resistance of Hepatitis C Virus Genotype 1b: Relationship to Nonstructural 5A Gene Quasispecies Mutations. J. Virol. 72:2795-2805

Read D., and Pietersen, G. 2015. Genotypic diversity of Citrus tristeza virus within red grapefruit, in a field trial site in South Africa. Eur. J. Plant. Pathol. 142:531-545

Sambrook J. (2001). Molecular Cloning: 'A laboratory manual'. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, New York

Scott, K.A., Hlela, Q., Zablocki, O., Read, D., van Vuuren, S., and Pietersen, G. 2012. Genotype composition of populations of grapefruit-cross-protecting citrus tristeza virus strain GFMS12 in different host plants and aphid-transmitted sub-isolates. Arch. Virol. 158:27-37

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol.Evol. 30:2725-2729

Teixeira, D.C., Lopes, S.A., Yamamoto, P.T., Eveillard, S., Martins, E.C., de Jesus Junior, W.C., Bassanezi, R.B., Ayres, A.J., Danet, J.L., Saillard, C., and Bové, J.M. 2005. PCR Detection of the Two Liberibacter Species Associated with Citrus Huanglongbing in São Paulo State, Brazil. In: Proc. 16th Conf. IOCV. IOCV, Riverside, CA. 432-438

van Vuuren, S.P., and Collins, R.P. 1993. Evaluation of Citrus Tristeza Virus Isolates for Cross Protection of Grapefruit in South Africa. Plant. Dis. 77:24-28

van Vuuren, S.P., van der Vyver, J.B,. and Luttig, M. 2000. Diversity Among Sub-Isolates of Cross-Protecting Citrus Tristeza Virus Isolates in South Africa. In: Proc. 14th Conf. IOCV. IOCV, Riverside, California. 103-110

van Vuuren S.P. and Manicom B.Q. (2005). The Response of Star Ruby Grapefruit to Different *Citrus tristeza virus* isolates. In: Proc 16th Conf. IOCV. IOCV, Riverside, California. 112-116

von Broembsen, L. A., and Lee, A.T.C. 1988. South Africa's Citrus Improvement Program. In: Proc. 10th Conf. IOCV. IOCV, Riverside, California. 407-416

Yoon, J.Y., Ahn, H.I., Kim, M., Tsuda, S., and Ryu, K.I. 2006. *Pepper mild mottle virus* pathogenicity determinants and cross protection effect of attenuated mutants in pepper. Virus. Res. 118:23-30

Zablocki, O., and Pietersen, G. 2014. Characterization of a novel citrus tristeza virus genotypewithin three cross-protecting source GFMS12 sub-isolates in South Africa by means of Illumina sequencing. Arch. Virol. 158:2133-2139