Designing and Testing of a *Citrus tristeza virus* Resequencing Microarray

Z. Xiong[1,2], R. Barthelson[2], Z. Weng[1,2], and D. W. Galbraith[2]

[1]Division of Plant Pathology & Microbiology, and [2]Department of Plant Sciences, University of Arizona, AZ 85721, USA

ABSTRACT. A high density, Affymetrix resequencing microarray capable of querying 117,088 nucleotides of *Citrus tristeza virus* (CTV) genomic sequences has been designed and fabricated. This microarray comprises nearly one million tiled probes of 25-mer oligonucleotides at a feature size of 8 µm. Each CTV target nucleotide is queried by 8 probes of mismatched and perfectly matched oligonucleotides in parallel. The tiled bases on the microarray contain full-length sequences from 4 representative CTV genomes and selected sequences from six other CTV genomes. Together, the information on the microarray represents the genetic diversity equivalent to 10 full-length CTV genomes. The utility and versatility of the CTV resequencing microarray were tested using cDNA clones of T30 and T36 isolates as target DNA. The entire genomes of isolates T30 and T36 were first amplified as five PCR DNA fragments of 4 kb each. After quantification and pooling, the amplified DNA was fragmented, biotin-labeled, and hybridized to the oligonucleotide probes on the microarray. Hybridization between the target DNA and the oligonucleotide probes was visualized by multiple-stage staining with streptavidin phycoerythrin and scanned using a high density laser scanner. Signal intensities were then analyzed using the Affymetrix GDAS software to finalize base calls. Base call rates from 99.5% to 99.7% and base call accuracies between 99.9% and 100% have been achieved using the CTV resequencing microarray. In addition, several nucleotide polymorphisms were identified between the cDNA clone analyzed by the resequencing microarray and the published T30 genomic sequence.

Citrus tristeza virus (CTV) has an extremely large genome of 19,226 to 19,302 nucleotides (2, 14, 20, 33, 37, 40). Twelve open reading frames (ORFs) can be predicted from CTV genomic sequences and these encode at least 19 proteins, including proteins required for viral replication, a chaperone-like protein homologue of the HSP70 protein, the major and minor capsid proteins (CP and CPm), an RNA-binding protein, three RNA silencing suppressors (19), and several proteins of unknown functions. The biological diversity of CTV may be attributed to the complexity and variability of the viral genome. Hundreds of CTV isolates exist and these isolates can be classified into groups on the basis of distinct symptoms exhibited on a set of standardized indicator plants (27).

Many techniques have been developed to rapidly differentiate CTV isolates and to infer potential pathogenicity from either viral protein or nucleic acid sequences. The CTV CP has long been a favored target for this purpose, since various serological (17, 24, 35) and peptide mapping (1) techniques are readily available. An even more diverse collection of techniques has been developed to examine the CTV genome directly, including analysis of double-stranded RNA (dsRNA) patterns (9, 22), hybridization with cDNA and oligonucleotides probes (6, 10, 16, 21, 23, 25), and procedures based on reverse transcription-polymerase chain reaction (RT-PCR) amplification (12, 32) followed by analysis of restriction fragment length polymorphism (8) and single-strand conformation polymorphism (32, 15, 29). Recently, direct sequencing of selected regions of the CTV genome has been carried out (11, 28, 30).

Nucleotide sequencing provides perhaps the most accurate identification of CTV isolates. However, cloning and sequencing of such a large genome by current protocols is time-consuming and labor-intensive. Nev-

ertheless, efforts have been made to sequence CTV genomes. Currently there are 863 CTV sequence entries deposited in the NCBI GenBank database, which include eight full-length CTV sequences derived from seven unique isolates, a number of defective genomes, and numerous fragments from various coding and non-coding regions.

Sequence analysis of selected regions of many isolates has largely been unsuccessful in identifying pathogenic determinants in the CTV genome. Rubio et al. (30) analyzed genetic variation of four genomic regions of 30 CTV isolates from Spain and California. Their results showed that most isolates contained one predominant sequence with other minor sequence variants, but no significant correlation was observed between geographic origins and nucleotide sequences. In another study (15), genomic regions of mild and severe CTV isolates were sequenced, but no conclusion could be made about the genetic determinants of CTV pathogenicity. These data suggest that examination of a few selected regions of CTV genome does not allow definitive differentiation between CTV isolates with different biological properties. Analysis of the entire genome sequence of CTV isolates is therefore a better approach to identify potential pathogenicity determinants that can be further studied genetically. However, sequencing many CTV genomes of ~20 kb each presents a technical challenge.

Rapid, high-throughput sequencing of large genomes or genomic regions has been recently made possible with improvements in lithographic techniques used in fabrication of high density oligonucleotide resequencing microarrays. Each high density microarray comprises more than one million 25-mer oligonucleotide probes packed at a feature size of 8 µm. Resequencing microarrays are specialized types of microarrays on which all four probes (a quartet) for each individual base on both strands of the sequence of interest are sequentially tiled. The four oligonucleotide probes in a quartet differ only by substitution of A, T, C or G at the central position. In this way, each nucleotide to be determined is represented by four oligonucleotides in the sense strand and four oligonucleotides in the antisense strand. Overlapping quartets of oligonucleotides representing the entire sequence of a targeted DNA are sequentially tiled on the microarray to allow the complete interrogation of the target DNA sequence. This tiling strategy allows both complete sequencing of each strand and identification of any known or novel single nucleotide polymorphisms (SNPs) in a single experiment (5, 13).

Resequencing microarrays have been used in comparative sequencing of candidate regions identified in mapping experiments (4), identification of causative SNPs in disease populations (34), and comprehensive sequence analysis of organelle genomes such as the entire 16.6 kb human mitochondrial DNA (7), viral genomes (3, 18, 38, 39), and bacterial genomes (41). In a study of severe acute respiratory syndrome coronavirus (SARS), a microarray was designed to successfully interrogate and resequence the entire 30 kb genome of SARS isolated from 16 patients (39). The array achieved a 99.99% base call accuracy with high reproducibility and was able to identify known and novel variants of SARS. More impressive is the sequencing of the 3.1 Mb genomes of 56 strains of *Bacillus anthracis* (anthrax bacterium) with a collection of 118 microarrays, each of which is capable of interrogating 29,212 bases (41). Both of the above examples used microarrays of 24 µm feature size.

In this report, we describe the design of a higher density Affymetrix oligonucleotide resequencing microarray that allows rapid, high-throughput sequencing of the entire

20 kb CTV genome. The microarray has a feature size of 8 µm and a capability to query 117,088 bases representing four full-length CTV genomes, as well unique sequences from several other CTV isolates. The resequencing microarray yielded a call rate of above 99% and a call accuracy of above 99.9% using cloned genomic cDNA of CTV isolates T30 and T36 as target DNAs.

MATERIALS AND METHODS

CTV genomic sequences and analysis. A total of eight full-length CTV genomic sequences were retrieved from GenBank: T30, a mild isolate from Florida (2), T36, a decline-inducing isolate from Florida (T36o, an older version (14) and T36n, a revised version of the sequence (33)), SY568, a decline and sweet orange stem pitting isolate from California (40), VT, a decline and seedling yellows isolate from Israel (20), T385, an essentially symptomless isolate from Spain which is nearly identical to isolate T30 (37), NUagA, a seedling yellows isolate from Japan (GenBank Accession No. AB046398); and Qaha, an isolate from Egypt nearly identical to T36 (Accession No. AY340974). In addition, full-length sequences of CTV isolate T3 (M. E. Hilf, personal communication) and H33 (T. E. Mirkov, personal communication) and a partial sequence (13,585 nt) of the CTV T68 isolate (M. E. Hilf, personal communication) were kindly supplied.

Sequences of these CTV isolates were aligned and their phylogenetic relationship was analyzed using the ClustalX software (36). The final phylogenetic trees were calculated using the neighbor-joining method of Saitou and Nei (31) with bootstrap resampling of 1000 iterations and visualized using the TreeView program (26).

CTV resequencing microarray design and fabrication. Due to the limitation on the number of nucleotides that can be tiled on the resequencing microarray, genomic sequences of only a few, representative CTV isolates were selected for tiling. Unique sequences from other CTV isolates were identified by comparing each isolate with the most closely related isolate that was tiled on the microarray. A Java applet was written to carry out the pairwise comparison. A unique sequence was defined as a 25 nucleotide region that contains either a run of two or more different nucleotides at the central (13th nucleotide) position, or two or more non-contiguous, different nucleotides within eight nucleotides of the 13th nucleotide. These parameters were dictated by the fact that a run of mismatched nucleotides or closely-spaced mismatched nucleotides would significantly destabilize the hybridization between the labeled target DNA and the oligonucleotide probe on the microarray. Unique nucleotide sequences identified by this program and selected full genomic sequences of CTV isolates were then tiled on the GeneChip CustomSeq resequencing array using the Affymetrix photolithographic manufacturing process (Santa Clara, CA).

Amplification of cDNA clones and target DNA preparation. In order to validate the design of the microarray and its utility, cDNA clones representing the full-length genomes of isolates T30 and T36 (kindly provided by W. O. Dawson) were used as the source of target DNA. The entire genome of each CTV isolate was amplified as five PCR fragments of about 4 kb in size using the pairs of forward and reverse primers listed in Table 1. The locations of the primers are indicated on the CTV genome illustrated in Fig. 1. Long range PCR amplification was carried out with the Stratagene EXL DNA polymerase as described by the manufacturer (Stratagene, La Jolla, CA).

Microarray hybridization and data analysis. Amplified PCR fragments were first quantified with the ND-1000 spectrophotometer (Nano-

TABLE 1
OLIGONUCLEOTIDE PRIMERS USED IN AMPLIFICATION OF THE COMPLETE T30 AND T36 GENOMES

Oligos	Length	Sequence	Genome location
T30 F1*	30	CGAATTTCGATTCAAATTCACCCGTATCTC	1
T30 F2	32	TACATCAAGTCCTGTAAGTCGAGAGTGGTCAT	4140
T30 F3	26	AATTCGCAAGCGGAGATTTCTCAAAG	8313
T30 F4	34	GTACCACGTTTTCGACGGTGGCTATGGCTACATC	12028
T30 F5	31	GTCAAGGTTACGAGGAGGCAACCGAGCTTCT	15954
T30 R1*	32	ATGACCACTCTCGACTTACAGGACTTGATGTA	4140
T30 R2	26	CTTTGAGAAATCTCCGCTTGCGAATT	8313
T30 R3	34	GATGTAGCCATAGCCACCGTCGAAAACGTGGTAC	12028
T30 R4	31	AGAAGCTCGGTTGCCTCCTCGTAACCTTGAC	15954
T30_T36 R5	24	TGGACCTATGTTGGCCCCCCATAG	19226/19270
T36 F1	31	AATTTCACAAATTCAACCTGTTCGCCCAGAA	1
T36 F2	31	CGGGAAGATTATACACCGAGACGTTCCAATC	4193
T36 F3	34	GGACGATGATGTCTATGTTGTTGACTTTAATCGG	7979
T36 F4	34	GCGGTATGGTGCTTTTGGGTTTAGACTTCGGTAC	12041
T36 F5	32	GTCAAGGTTACGAGGAGGCGACCGAGCTTCTT	15997
T36 R1	31	GATTGGAACGTCTCGGTGTATAATCTTCCCG	4193
T36 R2	34	CCGATTAAAGTCAACAACATAGACATCATCGTCC	7979
T36 R3	34	GTACCGAAGTCTAAACCCAAAAGCACCATACCGC	12041
T36 R4	32	AAGAAGCTCGGTCGCCTCCTCGTAACCTTGAC	15997

*F = Forward; R = Reverse. The R5 primer for both T30 and T36 is identical.

Drop Technologies, Wilmington, DE), as recommended in the Resequencing Assay Protocol V2.1 (Affymetrix, Santa Clara, CA). Equimolar amounts of each fragment were then pooled and cleaned using the Qiagen MinElute PCR Purification Kit (Qiagen, Valencia, CA). An amount of amplified DNA equivalent to 0.055 pmoles of CTV genome was subsequently fragmented to 20 to 200 bp using the Affymetrix GeneChip Fragmentation Reagent (Affymetrix, Santa Clara, CA) and labeled with biotin-dNTP by terminal deoxynucleotidyl transferase according to the Resequencing Assay Protocol.

Hybridization of the labeled target DNA to the microarrays and subsequent washing using GeneChip Fluidics Station 450 were performed in strict accordance with the instructions provided by Affymetrix. After hybridization, the target DNA bound to the probes on the microarray was stained using a three-stage process consisting of a streptavidin phycoerythrin (SAPE) stain, an amplification with biotinylated anti-streptavidin antibodies, and a final stain again with SAPE. The stained microarray was then scanned at a resolution of 1.563 μm/pixel using a GeneChip Scanner 3000 (Affymetrix, Santa Clara, CA).

The scanned image was automatically gridded and the signal for each probe was smoothed and averaged using the Affymetrix ScanToChp V1.50.40 software. The final probe intensity data were analyzed with the Affymetrix GeneChip DNA Analysis Software (GDAS) version 3.1 to extract sequencing information. Base calls were made using the ABACUS (adaptive background genotype calling scheme) algorithm (30). The ABACUS parameters were as follows: no signal threshold = 1.2, weak signal fold threshold = 3, maximum signal to noise ratio = 20, quality score threshold = 2.0, base reliability threshold across samples = 0.5, trace threshold = 1, and sequence profile threshold = -0.175.

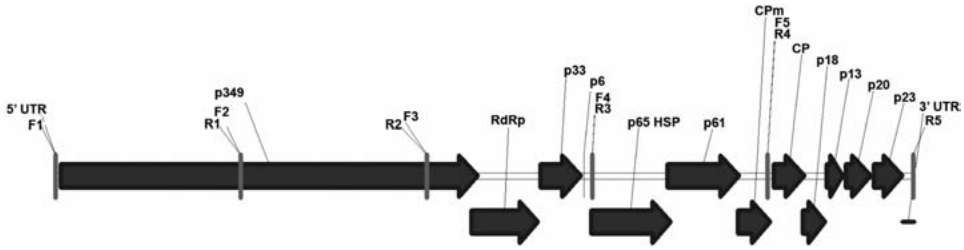

Fig. 1. Location of five pairs of reverse (R1-R5) and forward (F1-F5) PCR oligonucleotides primers on the CTV T30 and T36 genomes. Block arrows represent ORFs in the CTV genome. Protein products encoded by each ORF are listed on the top of the genome map. Locations of the primers are indicated by vertical lines.

RESULTS AND DISCUSSION

Phylogenetic analysis of CTV genomes. In order to select representative CTV genomic sequences for tiling on the microarray, the aforementioned ten full-length CTV genomes (T30, T36o, T36n, VT, SY568, T385, NUagA, Qaha, T3, and H33) were analyzed using the ClustalX program (36). The T68 genomic sequence was incomplete, containing approximately 3.7 kb near the 5' terminus and approximately 9.6 kb from the 3' terminus, and was therefore not included in the phylogenetic analysis of the full-length genomes. The resulting phylogram clearly displayed four clusters of CTV isolates (Fig. 2A). The T36 cluster contains the Egyptian Qaha isolate and the two versions of Florida isolate T36. The T30 cluster contains the T30 isolate from Florida, T385 from Spain and SY568 from California. The VT cluster contains the VT from Israel, NUagA from Japan, and H33 from Texas. The Florida T3 isolate (11) terminated on a separate branch.

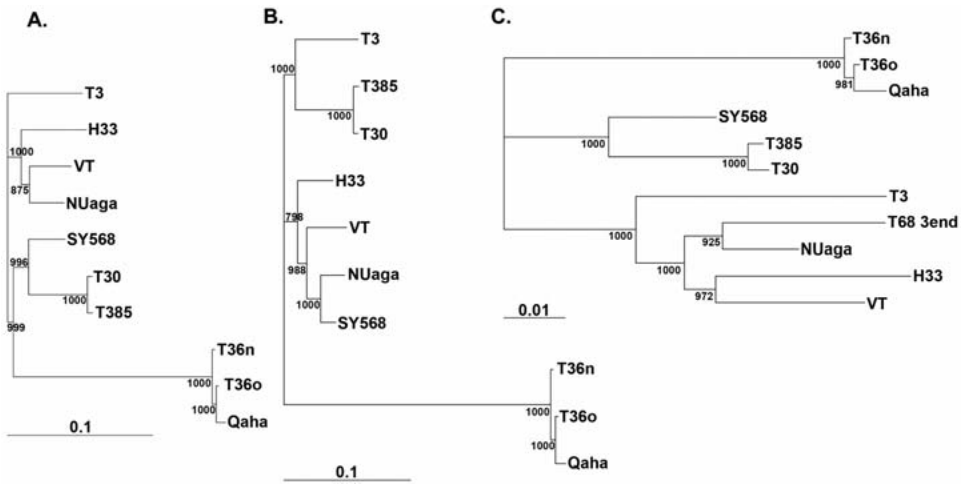

Fig. 2. Unrooted phylogenetic trees of CTV genomes. A. Ten completed CTV genomes. B. The 5' 9.7 kb of ten CTV genomes. C. The 3' 9.6 kb of 11 CTV genomes. Sequences were aligned by the ClustalX program and phylogenetic analysis was performed using the neighbor-joining algorithm (31) with 1000 bootstrap iterations of the aligned sequences. The scale at the bottom of each tree indicates the number of substitutions per site. The numbers to the left of the nodes are the bootstrap numbers that support the nodes to the right. Notice the smaller scale in C.

Phylogenetic analyses of the 5' and 3' "halves" of the available CTV genomes were performed separately (Fig. 2B, C) since the entire 5' half of the Florida T68 genome was not available. The larger amount of substitutions per site within the 5' half of the CTV genome indicated that this half is more diverse than the corresponding 3' half, as previously reported (20, 40). Interestingly, there are also some realignments of the CTV clusters. Phylogenetic analysis of the 5' half of the CTV genome (Fig. 2B) groups T3 more closely with the T30 cluster, but T3 is still distantly related to other members of that cluster. SY568, which was previously clustered with T30, now segregates with the VT cluster with a close relationship with the Japan SY isolate. On the other hand, T3 and T68 together with the previously classified VT cluster formed a cluster on the basis of the phylogenetic analysis of the 3' half of the CTV genome (Fig. 2C). The phylogenetic incongruity of SY568 is not surprising, as it has been proposed that SY568 is a recombinant between two distinct isolates (30). This possibly indicates a recombinant origin for T3 as well.

Selection and tiling of CTV genomic sequences. The maximal tiling capacity of 100,000 to 120,000 nucleotides on the microarray made it impossible to tile all the available genomic sequences of the 11 CTV isolates. Therefore, specific full-length sequences and sequence fragments were selected to ensure that the CTV genomic sequences tiled on the microarray represented maximum sequence diversity. Four individual full-length genomes were selected based on their phylogenetic clustering: T36, T30, VT, and T3, each representing a distinct cluster. Although T68 was not included in the full-length analysis and fell into the super VT cluster in the analysis of the 3' half sequence, it was shown to be uniquely different from other CTV isolates in the analysis of the 5' 3.5 kb sequence (data not shown) and other phylogenetic analysis (M. E. Hilf, personal communication). Consequently, all 13,585 available nucleotides of T68 were also completely tiled on the microarray.

For the remaining isolates, a pair-wise comparison was made between each isolate and the tiled representative isolate from each cluster to identify unique regions of sequences that are absent in the tiled representative. For example, T385 is a member of the T30 cluster so a pair-wise comparison was made between T385 and T30, the tiled representative in the cluster, to identify sequences unique to T385. This process identified 127 nucleotides unique to T385 which were then tiled on the microarray. A unique situation arose when SY568 was analyzed, as the 5' half of SY568 is more closely related to the tiled VT genome (Fig. 2B) while the 3' half of the genome is more closely related to tiled T30 genome (Fig. 2C). In order to minimize the number of nucleotides to be tiled on the microarray, the 5' and 3' halves of SY568 were compared to VT and T30, respectively, during the search for unique sequences. The tiling of 117,088 nucleotides consisting of four full-length CTV genomes, one partial genome and unique sequences from other CTV isolates is summarized in Table 2. In addition, 807 nucleotides from an artificial cDNA clone were included as internal controls. For each nucleotide tiled on the array, four 25-mer oligonucleotides corresponding to the sense strand and four oligonucleotides representing the antisense strand were tiled on the microarray. Thus, the microarray contains a total of 943,160 25-mer oligonucleotide probes.

Microarray hybridization and analysis. The complete genomes of isolates T30 and T36 were amplified as five PCR fragments in a highly specific manner (data not shown) using the sets of primers listed in Table 1. After quantification, pool-

TABLE 2
NUMBERS OF NUCLEOTIDES TILED ON THE CTV MICROARRAY FROM THE INDICATED SOURCE ISOLATES

CTV isolates	No. of bases
T30	19,259
T36	19,293
VT	19,226
T3	19,253
T68*	13,585
SY568 unique sequence (5' VT + 3' T30)**	8,090
H33 unique sequence (VT)	8,391
NUagA unique sequence (VT)	9,991
T385 unique sequence (T30)	127
Qaha unique sequence (T36)	2,298
Total nucleotides	117,088
Internal control nucleotides	807
Total Probes	943,160

*The genomic sequence of T68 is incomplete with a contig of 3.7 kb near the 5' end and a contig of 9.6 kb at the 3' end.
**Isolate within the parenthesis is the completely tiled, representative isolate of the phylogenetic group with which the pair-wide comparison was made. 5' = 5' half of the genome, 3' = 3' half of the genome.

ing, fragmentation, and end-labeling, the amplified DNA was used as target DNA to hybridize with probes on the microarray.

A scanned image of the hybridized microarray is presented in Fig. 3A. The hybridization intensities of all oligonucleotide probes in every quartet were then statistically averaged. Base calls were then made using the ABACUS algorithm with the optimized base calling parameters described in Materials and Methods. Generally, the differential hybridization intensities between two perfectly matched oligonucleotide probes and six mismatched oligonucleotide probes determine the correct base calls, as illustrated in Fig. 3B, C.

Using the CTV T36 genome of 19,293 nucleotides as the target DNA, a total of 19,269 bases were generated from the hybridization with the homologous T36 probes tiled on the microarray. The first 12 and the last 12 bases of the genome could not be queried by the microarray because of the microarray design. Among the 19,269 output bases, 19,158 bases were called, resulting in a call rate of 99.424%. Only 111 bases failed to resolve statistically and were not called ("n"). What is most impressive is that all of the called bases were identical to the published T36 sequence (33), an astounding call accuracy of 100%.

When the T36 target DNA was hybridized against heterologous probes, the call rates were lower, as expected. Sequences generated from T30 probes had a call rate of 41.235% and a call accuracy of 98.647%. Similarly, a call rate of 33.812% and a call accuracy of 98.025% were observed with the T3 probes. The lowest call rate of 29.64%, and the lowest call accuracy of 97.365% were obtained with the VT probes, as the VT isolate is more distantly related to T36 isolate than T30 isolate (Fig. 2A). Nevertheless, accurate sequences of 8058, 6633, and 5845 nucleotides were obtained from the heterologous probes of the T30, T3, and VT isolates, respectively. Although the call rate and call accuracy are constrained largely by the sequence similarity between the target DNA and the tiled probes, the ABACUS calling parameters

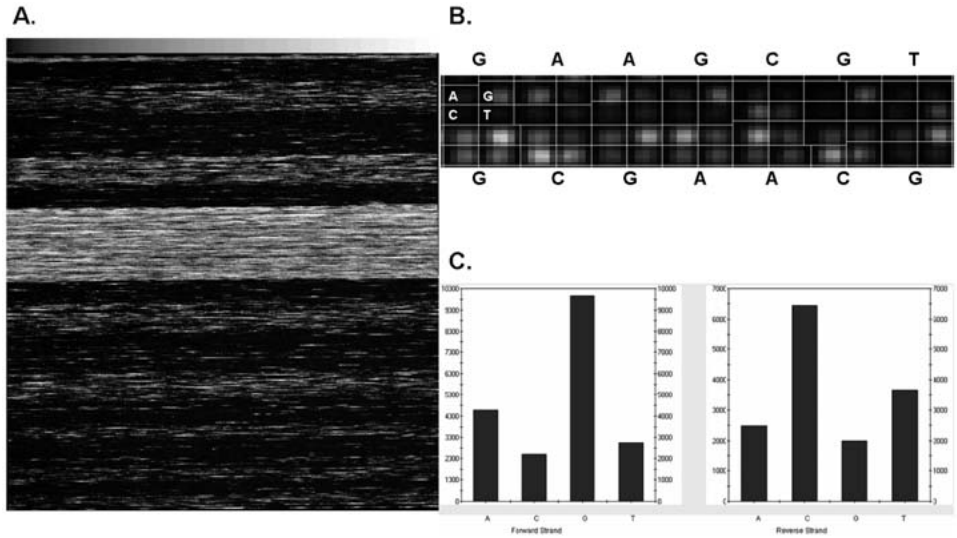

Fig. 3. A. Scanned image of the CTV resequencing microarray hybridized with the T36 target DNA. The scale bar at the top of the image indicates relative hybridization intensity. The bright section in the middle of the image contains the tiled oligonucleotide probes homologous to T36. B. An enlarged section of the microarray, showing the probe arrangement in each quartet and hybridization intensity of perfectly matched and mismatched oligonucleotide probes. Each quartet is arranged as labeled in the top left quartet: A probe at the top left, C at the bottom left, G at the top right, and T at the bottom right. Each probe has a feature size of 8 µm × 8 µm and is scanned as 5 8 5 pixels at a resolution of 1.56 µm per pixel. Bases called for each quartet are indicated on the top or the bottom of the panel. C. Hybridization intensity (y-axis) of each oligonucleotide probe (x-axis) in the sense and antisense quartets. The base queried by these eight probes is called as G. Notice the differential hybridization intensities between the perfectly matched and mismatched probes.

could be refined further to achieve better call rates and call accuracies.

When the T30 target DNA was hybridized to the resequencing microarray, similar results were obtained. Among the 19,246 bases generated from hybridization to the homologous oligonucleotide probes, 225 bases produced quality scores below the cutoff threshold of 2 and thus were not called. Twenty bases called by the GDAS program were different from the published T30 sequence (Table 3). The base call rate and accuracy using the T30 target DNA were 98.8% and 99.9%, respectively. Call rates and call accuracies from hybridization of T30 target DNA against non-homologous oligonucleotides probes were also similar to those obtained with T36 target DNA. The call rate and call accuracy were positively correlated with similarity between the target DNA and the oligonucleotide probes.

There was an unusually long stretch of 120 non-called bases between nucleotide 6,016 and 6,136 in the tiled T30 sequence that biased the base call rate obtained with the T30 target DNA. When this long stretch of non-calls is discounted, the base call rate increases to 99.5%, similar to the call rate obtained with the T36 target DNA against the tiled T36 probes. The non-calls in this region were a result of extremely low hybridization signals (about 10 fold lower than the average hybridization intensity) and, consequently, low quality scores. Either a deletion of 120 nucleotides or a run of nucleotides in the cDNA clone that are completely different from the published reference sequence could cause these aberrant non-calls. Direct

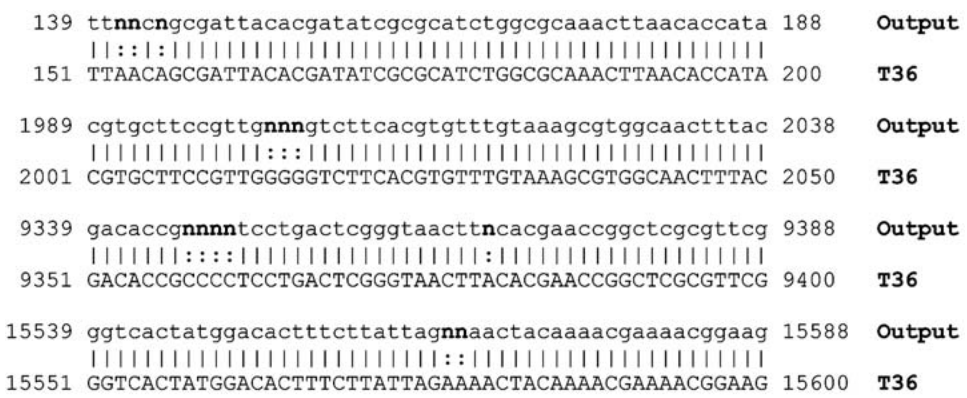

Fig. 4. Alignments of selected regions between output from the resequencing microarray and the published sequence of CTV T36 isolate, illustrating runs of non-calls (n in bold) in AT- or GC-rich contexts.

sequencing of this region of the T30 cDNA clone will be necessary to resolve this issue.

Twenty nucleotides that differed between the published T30 sequence and the sequence generated by the resequencing microarray were distributed mostly in the 5' half of the T30 genome and all differences were located within coding sequences (Table 3). About one half of the nucleotide changes resulted in amino acid changes, but none of the changes altered any reading frames significantly. The quality scores associated with most of these nucleotides are significantly higher than the cut-off threshold of 2.0 (Table 3), indicating high reliability of these base calls. Thus, the differences between

TABLE 3
NUCLEOTIDE DIFFERENCES BETWEEN PUBLISHED AND MICROARRAY GENERATED GENOMIC SEQUENCES FOR ISOLATE T30

Genome location	Published	Microarray	Quality scores	Genes	Changes in aa
1815	c	t	14.2	p349	Leu to Phe
2501	a	g	18.2	p349	
3405	a	g	13.8	p349	Ile to Ser
3860	c	t	16.9	p349	
4405	g	a	18.1	p349	Arg to Lys
4607	t	c	11.7	p349	
4655	t	c	20.7	p349	
5269	t	c	23.4	p349	Leu to Pro
6109	t	a	2.7	p349	Val to Asp
7379	t	c	2.8	p349	
7431	t	a	12.1	p349	Leu to Ile
7655	a	g	15.1	p349	Gln to Arg
7668	a	g	24.8	p349	Lys to Glu
8087	t	g	5.1	p349	
8117	t	c	22.2	p349	
9943	a	g	22.6	RdRp	Ile to Val
11181	g	a	22.8	p33	
13968	c	g	23.2	p61	Asn to Lys
14511	a	g	21.2	p61	
16828	g	a	22.0	p18	Asn to Ser

the sequence obtained by traditional sequencing and that obtained by the resequencing microarray are real. The mostly likely explanation is that the differences reflect true SNPs existing in two independent cDNA clones derived from the same region of the CTV genome.

The non-calls in both T36 and T30 sequences, with the exception of the long stretch of non-calls in the T30 sequence, were distributed throughout the genome, but were often present in short runs of two to four bases (Fig. 4). These short runs of non-calls reside in either AT-rich or GC-rich regions, features of DNA sequences that are known to be problematic in resequencing analysis (39, 41). Probes with an AT-rich region hybridize weakly with the target DNA, and probes with a GC-rich region hybridize too strongly, resulting in little differentiation between hybridization intensities of perfectly matched probes and mismatched probes, and consequently producing no calls. Improved hybridization conditions and buffer compositions might resolve some of these non-calls.

In summary, the CTV resequencing microarray has yielded call rates of above 99% and call accuracies of 99.9% to 100% for two different target DNA preparations. The performance of the CTV resequencing microarray is on a par with the published results for a resequencing microarray for SARS virus (39) and much better than that of the *B. anthracis* genome resequencing microarray (41). In addition, the CTV resequencing microarray identified a number of SNPs in the full-length cDNA clone of CTV T30 isolate, demonstrating the utility and versatility of the CTV resequencing microarray in the analysis of CTV genomes. However, how well the resequencing microarray performs against an unknown CTV isolate whose sequence differs substantially from the tiled CTV genomic sequences remain to be determined. The design and the calling algorithm of the resequencing microarray allow easy discrimination of SNPs, but do not resolve sequences containing multiple substitutions in consecutive nucleotides, deletions, or insertions. Nevertheless, sequences generated from the resequencing microarray should be sufficient to serve as digital fingerprints to differentiate various CTV isolates.

ACKNOWLEDGMENTS

The authors would like to thank Drs. M. E. Hilf and T. E. Mirkov for sharing unpublished CTV genomic sequences and Dr. W. O. Dawson for providing cDNA clones of CTV T30 and T36 isolates. This research was supported by USDA grant #2003-34-39913-764.

LITERATURE CITED

1. Albiach-Martí, M. R., J. Guerri, M. Cambra, S. M. Garnsey, and P. Moreno
2000. Differentiation of *Citrus tristeza virus* isolates by serological analysis of p25 coat protein peptide maps. J. Virol. Methods 88: 25-34.
2. Albiach-Martí, M. R., M. Mawassi, S. Gowda, T. Satyanarayana, M. E. Hilf, S. Shanker, E. C. Almira, M. C. Vives, C. Lopez, J. Guerri, R. Flores, P. Moreno, S. M. Garnsey, and W. O. Dawson
2000. Sequences of *Citrus tristeza virus* separated in time and space are essentially identical. J. Virol. 74: 6856-6865.
3. Cherkasova, E., M. Laassri, V. Chizhikov, E. Korotkova, E. Dragunsky, V. I. Agol, and K. Chumakov
2003. Microarray analysis of evolution of RNA viruses: Evidence of circulation of virulent highly divergent vaccine-derived polioviruses. Proc. Natl. Acad. Sci. USA 100: 9398-9403.
4. Chudgar, P., N. Xu, L. P. Chorich, C. Liu, J. A. Warrington, and L. C. Layman
2003. Screening candidate genes for mutations using genome custom resequencing microarrays. Amer. J. Hum. Genet. 73: 578-578.
5. Cutler, D. J., M. E. Zwick, M. M. Carrasquillo, C. T. Yohn, K. P. Tobin, C. Kashuk, D. J. Mathews, N. A. Shah, E. E. Eichler, J. A. Warrington, and A. Chakravarti
2001. High-throughput variation detection and genotyping using microarrays. Genome Res. 11: 1913-1925.

6. Derrick, K. S., M. J. Beretta, G. A. Barthe, M. Kayim, and R. Harakava
2003. Identification of strains of *Citrus tristeza virus* by subtraction hybridization. Plant Dis. 87: 1355-1359.
7. Fan, J. B., U. Surti, P. Taillon-Miller, L. Hsie, G. C. Kennedy, L. Hoffner, T. Ryder, D. G. Mutch, and P. Y. Kwok
2002. Paternal origins of complete hydatidiform moles proven by whole genome single-nucleotide polymorphism haplotyping. Genomics 79: 58-62.
8. Gillings, M., P. Broadbent, J. Indsto, and R. Lee
1993. Characterization of isolates and strains of citrus tristeza closterovirus using restriction analysis of the coat protein gene amplified by the polymerase chain reaction. J. Virol. Methods 44: 305-317.
9. Habili, N.
1993. Detection of Australian field isolates of citrus tristeza virus by double-stranded RNA analysis. J. Phytopathol. 138: 308-316.
10. Halbert, S. E., H. Genc, B. Cevik, L. G. Brown, I. M. Rosales, K. L. Manjunath, M. Pomerinke, D. A. Davison, R. F. Lee, and C. L. Niblett
2004. Distribution and characterization of *Citrus tristeza virus* in South Florida following establishment of *Toxoptera citricida*. Plant Dis. 88: 935-941.
11. Hilf, M. E., A. V. Karasev, M. R. Albiach-Martí, W. O. Dawson, and S. M. Garnsey
1999. Two paths of sequence divergence in the citrus tristeza virus complex. Phytopathology 89: 336-342.
12. Huang, Z. P., P. A. Rundell, and C. A. Powell
2004. Detection and isolate differentiation of *Citrus tristeza virus* in infected field trees based on reverse transcription-polymerase chain reaction. Plant Dis. 88: 625-629.
13. Jaakson, K., J. Zernant, M. Kulm, A. Hutchinson, N. Tonisson, D. Glavac, M. Ravnik-Glavac, M. Hawlina, M. R. Meltzer, R. C. Caruso, F. Testa, A. Maugeri, C. B. Hoyng, P. Gouras, F. Simonelli, R. A. Lewis, J. R. Lupski, F. P. M. Cremers, and R. Allikmets
2003. Genotyping microarray (gene chip) for the ABCR (ABCA4) gene. Hum. Mut. 22: 395-403.
14. Karasev, A. V., V. P. Boyko, S. Gowda, O. V. Nikolaeva, M. E. Hilf, E. V. Koonin, C. L. Niblett, K. Cline, D. J. Gumpf, R. F. Lee, S. M. Garnsey, D. J. Lewandowski, and W. O. Dawson
1995. Complete sequence of the citrus tristeza virus RNA genome. Virology 208: 511-520.
15. Kong, P., L. Rubio, M. Polek, and B. W. Falk
2000. Population structure and genetic diversity within California *Citrus tristeza virus* (CTV) isolates. Virus Genes 21: 139-145.
16. Lair, S. V., T. E. Mirkov, J. A. Dodds, and M. F. Murphy
1994. A single temperature amplification technique applied to the detection of citrus tristeza viral RNA in plant nucleic acid extracts. J. Virol. Methods 47: 141-151.
17. Lin, Y. J., P. A. Rundell, L. H. Xie, and C. A. Powell
2000. *In situ* immunoassay for detection of *Citrus tristeza virus*. Plant Dis. 84: 937-940.
18. Long, W. H., H. S. Xiao, X. M. Gu, Q. H. Zhang, H. J. Yang, G. P. Zhao, and J. H. Liu
2004. A universal microarray for detection of SARS coronavirus. J. Virol. Methods 121: 57-63.
19. Lu, R., A. Folimonov, M. Shintaku, W. X. Li, B. W. Falk, W. O. Dawson, and S. W. Ding
2004. Three distinct suppressors of RNA silencing encoded by a 20-kb viral RNA genome. Proc. Natl. Acad. Sci. USA 101: 15742-15747.
20. Mawassi, M., E. Mietkiewska, R. Gofman, G. Yang, and M. Bar-Joseph
1996. Unusual sequence relationships between two isolates of citrus tristeza virus. J. Gen. Virol. 77: 2359-2364.
21. Mehta, P., R. H. Brlansky, S. Gowda, and R. K. Yokomi
1997. Reverse-transcription polymerase chain reaction detection of citrus tristeza virus in aphids. Plant Dis. 81: 1066-1069.
22. Moreno, P., J. Guerri, J. F. Ballester-Olmos, R. Albiach, and M. E. Martínez
1993. Separation and interference of strains from a citrus tristeza virus isolate evidenced by biological-activity and double-stranded-RNA (dsRNA) analysis. Plant Pathol. 42: 35-41.
23. Narváez, G., B. S. Skander, M. A. Ayllón, L. Rubio, J. Guerri, and P. Moreno
2000. A new procedure to differentiate citrus tristeza virus isolates by hybridisation with digoxigenin-labelled cDNA probes. J. Virol. Methods 85: 83-92.
24. Nikolaeva, O. V., A. V. Karasev, S. M. Garnsey, and R. F. Lee
1998. Serological differentiation of the citrus tristeza virus isolates causing stem fitting in sweet orange. Plant Dis. 82: 1276-1280.
25. Olmos, A., M. Cambra, O. Esteban, M. T. Gorris, and E. Terrada
1999. New device and method for capture, reverse transcription and nested PCR in a single closed-tube. Nucleic Acids Res. 27: 1564-1565.
26. Page, R. D. M.
1996. TreeView: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12: 357-358.

27. Rocha-Peña, M. A., R. F. Lee, R. Lastra, C. L. Niblett, F. M. Ochoa-Corona, S. M. Garnsey, and R. K. Yokomi
1995. Citrus tristeza virus and its aphid vector *Toxoptera citricida*—Threats to citrus production in the Caribbean and Central and North America. Plant Dis. 79: 437-445.
28. Roy, A. and R. H. Brlansky
2004. Genotype classification and molecular evidence for the presence of mixed infections in Indian *Citrus tristeza virus* isolates. Arch. Virol. 149: 1911-1929.
29. Rubio, L., M. A. Ayllón, J. Guerri, H. Pappu, C. L. Niblett, and P. Moreno
1996. Differentiation of citrus tristeza closterovirus (CTV) isolates by single-strand conformation polymorphism analysis of the coat protein gene. Ann. Appl. Biol. 129: 479-489.
30. Rubio, L., M. A. Ayllón, P. Kong, A. Fernandez, M. Polek, J. Guerri, P. Moreno, and B. W. Falk
2001. Genetic variation of *Citrus tristeza virus* isolates from California and Spain: Evidence for mixed infections and recombination. J. Virol. 75: 8054-8062.
31. Saitou, N. and M. Nei
1987. The neighbor-joining method—a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.
32. Sambade, A., C. López, L. Rubio, R. Flores, J. Guerri, and P. Moreno
2003. Polymorphism of a specific region in gene p23 of *Citrus tristeza virus* allows discrimination between mild and severe isolates. Arch. Virol. 148: 2325-2340.
33. Satyanarayana, T., S. Gowda, V. P. Boyko, M. R. Albiach-Martí, M. Mawassi, J. Navas-Castillo, A. V. Karasev, V. Dolja, M. E. Hilf, D. J. Lewandowski, P. Moreno, M. Bar-Joseph, S. M. Garnsey, and W. O. Dawson
1999. An engineered closterovirus RNA replicon and analysis of heterologous terminal sequences for replication. Proc. Natl. Acad. Sci. USA 96: 7433-7438.
34. Tebbutt, S. J., J. Q. He, K. M. Burkett, J. Ruan, I. Opushnyev, B. W. Tripp, J. A. Zeznik, C. O. Abara, C. C. Nelson, and K. R. Walley
2004. Microarray genotyping resource to determine population stratification in genetic association studies of complex disease. Biotechniques 37: 977-985.
35. Terrada, E., R. J. Kerschbaumer, G. Giunta, P. Galeffi, G. Himmler, and M. Cambra
2000. Fully "recombinant enzyme-linked immunosorbent assays" using genetically engineered single-chain antibody fusion proteins for detection of *Citrus tristeza virus*. Phytopathology 90: 1337-1344.
36. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins
1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876-4882.
37. Vives, M. C., L. Rubio, C. Lopez, J. Navas-Castillo, M. R. Albiach-Marti, W. O. Dawson, J. Guerri, R. Flores, and P. Moreno
1999. The complete genome sequence of the major component of a mild *Citrus tristeza virus* isolate. J. Gen. Virol. 80: 811-816.
38. Wang, D., A. Urisman, Y. T. Liu, M. Springer, T. G. Ksiazek, D. D. Erdman, E. R. Mardis, M. Hickenbotham, V. Magrini, J. Eldred, J. P. Latreille, R. K. Wilson, D. Ganem, and J. L. DeRisi
2003. Viral discovery and sequence recovery using DNA microarrays. Plos Biol. 1: 257-260.
39. Wong, C. W., T. J. Albert, V. B. Vega, J. E. Norton, D. J. Cutler, T. A. Richmond, L. W. Stanton, E. T. Liu, and L. D. Miller
2004. Tracking the evolution of the SARS coronavirus using high-throughput, high-density resequencing arrays. Genome Res. 14: 398-405.
40. Yang, Z. N., D. M. Mathews, J. A. Dodds, and T. E. Mirkov
1999. Molecular characterization of an isolate of *Citrus tristeza virus* that causes severe symptoms in sweet orange. Virus Genes 19: 131-142.
41. Zwick, M. E., F. McAfee, D. J. Cutler, T. D. Read, J. Ravel, G. R. Bowman, D. R. Galloway, and A. Mateczun
2004. Microarray-based resequencing of multiple *Bacillus anthracis* isolates. Genome Biol. 6: R10.