

## ABSTRACT

KAYE, AMANDA CLAIRE. Population Genetic Analysis of Tomato Spotted Wilt Virus (TSWV) on Peanut in North Carolina and Virginia. (Under the direction of Marc A. Cubeta and George G. Kennedy.)

Exploring the genetic diversity and evolutionary history of plant viruses is critical to understanding their ecology and epidemiology. Tomato spotted wilt virus (TSWV) is an ambisense RNA virus that infects over 1000 species of plants and has been shown to reassort its genome. To further investigate this virus, maximum-likelihood and population genetics-based methods were used to investigate the population structure, genetic diversity, and sources of genetic variation in field isolates of TSWV from peanut in North Carolina and Virginia. Selected regions of the nucleocapsid, movement, and RNA-dependent RNA polymerase genes were amplified and sequenced to identify haplotypes and infer genetic relationships between isolates of TSWV with heuristic methods. The haplotype structure of each locus consisted of one or two predominant haplotypes and more than 100 haplotypes represented by a single isolate. No specific haplotypes or clades were associated with geographic area, peanut cultivar or year. The population was panmictic at the regional level and high levels of genetic diversity were observed among isolates. There was evidence for negative selection acting upon each locus and maximum-likelihood analyses, which indicated that exponential growth was occurring in the population. The results of compatibility analyses and the persistence of specific gene sequences in isolates collected over three field seasons suggest that recombination was occurring in the population. Phylogenetic analysis supports that each locus has an independent evolutionary history.

Also, high diversity in viruses has been attributed to the occurrence of recombination and mutation. To investigate recombination in TSWV, matrix compatibility and ancestral

recombination methods were used to detect, quantify, and reconstruct the history of recombination in three genes of TSWV in isolates collected from three cultivars of peanut (Gregory, NC12C, and Perry) sampled in multiple years. Site commonalities among the isolates collected from different cultivars were found in each gene where the majority of the recombination appeared to be occurring. These corresponded to sequence near the 3' terminus in the N and RdRP genes and to sequence near the 5' terminus in the NSm gene. Between isolates collected from different peanut cultivars, the Gregory isolates proportionally had the most recombination occurring in the N gene and the RdRP genes, while the Perry isolates had the most recombination occurring in the NSm. Recombination rates per site estimated for the N, NSm, and RdRP genes in the NC12C isolates were 0.00934, 0.01559, and 0.00495, respectively.

Population Genetic Analysis of Tomato Spotted Wilt Virus (TSWV) on Peanut in North  
Carolina and Virginia

by  
Amanda Claire Kaye

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Plant Pathology

Raleigh, North Carolina

2009

APPROVED BY:

---

Marc A. Cubeta  
Committee Co-Chair

---

George G. Kennedy  
Committee Co-Chair

---

Barbara B. Shew

---

James W. Moyer

## **DEDICATION**

I would like to dedicate this work to the public school system of the Commonwealth of Virginia, which literally taught me everything I know, and to the memory of former President Calvin Coolidge, because indeed, nothing can take the place of Persistence.

## **BIOGRAPHY**

Amanda C. Kaye was born and raised in Virginia Beach, Virginia. And she lived happily ever after, to the end of her days...

## **ACKNOWLEDGMENTS**

I would like to thank all the people who supported me along my way. You already know who you are.

## TABLE OF CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES.....	vii
<b>Chapter 1: Population Genetic Analysis of Tomato Spotted Wilt Virus on Peanut in North Carolina and Virginia.....</b>	<b>1</b>
Abstract.....	2
Introduction.....	2
Materials and Methods.....	4
Results.....	9
Discussion.....	13
Literature Cited.....	21
<b>Chapter 2: Examining Recombination in the Nucleocapsid, Movement, and RNA-dependent RNA Polymerase Genes of Tomato Spotted Wilt Virus .....</b>	<b>43</b>
Abstract.....	44
Introduction.....	45
Materials and Methods.....	46
Results.....	50
Discussion.....	58
References.....	65

## LIST OF TABLES

### Chapter 1: Population Genetic Analysis of Tomato Spotted Wilt Virus on Peanut in North Carolina and Virginia

	Page
Table 1. Year of collection, geographic area, peanut cultivar and number of amplified nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) fragments of 196 isolates of <i>Tomato spotted wilt virus</i> .....	31
Table 2. Summary of the number of isolates, length of gene fragment, haplotype number, haplotype diversity, and haplotype frequencies of the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of <i>Tomato spotted wilt virus</i> on peanut.....	32
Table 3. Frequency of haplotypes in the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of <i>Tomato spotted wilt virus</i> that were distributed across different peanut cultivars in different geographical areas from 2005 to 2007.....	33
Table 4. Summary of the number of segregating sites, nucleotide diversity, and estimates of the population mutation rate of the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of <i>Tomato spotted wilt virus</i> on peanut.....	35
Table 5. Summary of neutrality test statistic values for the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of <i>Tomato spotted wilt virus</i> on peanut.....	36

### Chapter 2: Examining Recombination in the Nucleocapsid, Movement, and RNA-dependent RNA Polymerase Genes of Tomato Spotted Wilt Virus

Table 1. Year of collection, cultivar, isolate number and accession numbers of amplified nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) fragments of 97 isolates of <i>Tomato spotted wilt virus</i> .....	69
Table 2. The ratios of the number of inferred recombination events to the number of possible recombination sites for the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) fragments of NC12C, Perry, and Gregory isolates of <i>Tomato spotted wilt virus</i> .....	72

## LIST OF FIGURES

### **Chapter 1: Population Genetic Analysis of Tomato Spotted Wilt Virus on Peanut in North Carolina and Virginia**

	Page
Figure 1. Unrooted majority rules consensus tree of the N gene for isolates of <i>Tomato spotted wilt virus</i> obtained using nucleotide sequences without conflict.....	37
Figure 2. Unrooted majority rules consensus tree of the NSm gene for isolates of <i>Tomato spotted wilt virus</i> obtained using nucleotide sequences without conflict.....	39
Figure 3. Unrooted majority rules consensus tree of the RdRP gene for isolates of <i>Tomato spotted wilt virus</i> obtained using nucleotide sequences without conflict.....	41

### **Chapter 2: Examining Recombination in the Nucleocapsid, Movement, and RNA-dependent RNA Polymerase Genes of Tomato Spotted Wilt Virus**

Figure 1. Ancestral recombination graph (ARG) for the N gene region of the NC12C isolates.....	73
Figure 2. Ancestral recombination graph (ARG) for the NSm gene region of the NC12C isolates.....	74
Figure 3. Ancestral recombination graph (ARG) for the RdRP gene region of the NC12C isolates.....	75

## CHAPTER 1

### **Population Genetic Analysis of Tomato Spotted Wilt Virus on Peanut in North Carolina and Virginia<sup>1</sup>**

Amanda C. Kaye

James W. Moyer

Elizabeth J. Parks

Ignazio Carbone

Marc A. Cubeta

*Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina, 27695.*

---

<sup>1</sup> This chapter was formatted for submission to the journal *Phytopathology*

## **ABSTRACT**

Exploring the genetic diversity and evolutionary history of plant viruses is critical to understanding their ecology and epidemiology. In this study, maximum-likelihood and population genetics-based methods were used to investigate the population structure, genetic diversity, and sources of genetic variation in field isolates of Tomato spotted wilt virus (TSWV) from peanut in North Carolina and Virginia. Selected regions of the nucleocapsid, movement, and RNA-dependent RNA polymerase genes were amplified and sequenced to identify haplotypes and infer genetic relationships between isolates of TSWV with heuristic methods. The haplotype structure of each locus consisted of one or two predominant haplotypes and more than 100 haplotypes represented by a single isolate. No specific haplotypes or clades were associated with geographic area, peanut cultivar or year of isolation. The population was panmictic at the regional level and high levels of genetic diversity were observed among isolates. There was evidence for negative selection acting upon each locus and maximum-likelihood analyses indicated that exponential growth was occurring in the population. The results of compatibility analyses and the persistence of specific gene sequences in isolates collected over three field seasons suggest that recombination was occurring in the population. Phylogenetic analysis supports that each locus has an independent evolutionary history.

## **INTRODUCTION**

Plant viruses are a significant barrier to optimal worldwide crop production and can cause severe economic losses to growers (84). This is due to viral disease symptoms that

lead to host plant death, yield loss and post harvest damage (3). *Tomato spotted wilt virus* (TSWV) can infect over 1000 plant species (49) and has been estimated to cause annual losses of one billion US dollars(1). In addition, high infection rates, which can be in excess of 90% of a field (15), have lead to TSWV long being ranked as one of the ten most economically destructive plant viruses (80). Control of this virus is effected indirectly, through a multifaceted approach combining resistant varieties, crop cultural practices, and integrated pest management (11, 15, 60), but more effective management strategies could be facilitated by a better understanding of the population dynamics involved in the pathogenicity of this virus.

The TSWV genome has a single-stranded, tripartite RNA genome (6, 85). The virion is spherical, 80-120 nm in diameter, with a lipid bilayer envelope containing viral-coded surface glycoproteins (31). These characteristics place TSWV in the virus family Bunyaviridae, but the genome organization places TSWV in its own genus, where it is the type member (43). The smallest segment, S (2.9kb), is ambisense and codes for the nonstructural protein, NSs, which is identified as a silencing suppressor (76), in the viral sense and the viral nucleocapsid protein, N, in the viral complementary sense (18). The medium segment, M (4.8kb), is also ambisense and codes for the nonstructural protein, NSm, which is identified as the movement protein (40, 68, 69) in the viral sense and the precursor to the G1/G2 external glycoproteins, in the viral complementary sense (39). The large segment, L (8.9kb), codes for the viral RNA-dependent RNA polymerase in the viral complementary sense (2, 10, 17).

A previous study examined the population genetics of TSWV and found that populations of the virus were differentiated geographically on a global scale and presented evidence that positive selection favors divergence in different Tospovirus species (82). However, this study did not address whether similar structural characteristics are present in localized TSWV subpopulations. The objective of this research was to characterize the genetic diversity and infer the phylogenetic history of TSWV in field isolates collected from symptomatic peanut (*Arachis hypogaea*) tissue in a localized geographical area. To better elucidate the evolutionary processes underlying populations of TSWV in a single plant host in a localized geographic region, we examined three regions of the TSWV genome in a multi-locus approach. We used maximum-likelihood methods with estimations of population parameters to examine the similarity of evolutionary trends in the local and global TSWV populations including population subdivision, population growth, and evolutionary forces such as selection. This study represents the most comprehensive examination of localized field populations in a single host collected over time on a member of the family Bunyaviridae.

## **MATERIALS AND METHODS**

**Isolates.** One hundred ninety-six isolates of Tomato spotted wilt virus (TSWV) from peanut (*Arachis hypogaea* L.) leaves exhibiting typical symptoms (e.g., ring spots, "rusting", stunting, shortened internodes, and bud necrosis) were used in this study (Table 1). A hierarchical sampling strategy was deployed to collect leaf samples from one field of peanut in Bertie County, NC (77.1755° W, 36.1324° N) in 2005; one field each in Bertie County,

NC and Suffolk County, VA (76.6097° W, 36.7411° N) in 2006; and one field each in Bertie and Sampson Counties, NC (78.3318° W, 34.8462°N) and Suffolk County, VA in 2007.

Geographically, these areas are nearly contiguous. Quadrifoliate leaves were collected every 4 weeks for 16 weeks from arbitrarily-chosen symptomatic plants during each year. Five peanut cultivars were represented in the sample: NC\_12C (32), Gregory (35), Brantley (34), Perry (33) and NC-V11 (90). Each geographic area represented a different population in this study. All leaf samples were transported from the field to the laboratory on ice in a cooler and were either processed for testing the same day or stored at 4°C. Each leaf sample was tested with either a double antibody sandwich enzyme-linked immunosorbent assay (DAS-ELISA) or ImmunoStrip (Agdia, Elkhart, IL) to determine the presence of TSWV.

**RNA Extraction.** Total plant RNA was extracted from each TSWV-positive leaf sample with either TRIzol Reagent (Invitrogen Corporation, Carlsbad, CA), RNeasy Plant Mini Kit (Qiagen, Valencia, CA) with 1% polyvinyl pyrrolidone in RLT buffer (64), or directly from leaf sap-saturated ImmunoStrips using the RNeasy Plant Mini Kit with 1% polyvinyl pyrrolidone according to the manufacturer's protocol.

**RT-PCR and Sequencing.** Extracted RNA from each sample was used as a template for reverse transcription-polymerase chain reaction (RT-PCR) to generate complementary DNA (cDNA) for selected portions of the nucleocapsid (N), movement (NsM), and RNA dependent RNA polymerase (RdRP) genes of TSWV (81). A 720 bp fragment of the N gene was amplified with the S1976 forward primer (5'-ACAGCTG(ACT)TTTTAAGCAAGTTCTG-3') and the S2847 reverse primer (5'-AAGAAACGACTGCGG(AG)ATACA-3'). A 699 bp fragment of the NSm gene was

amplified with the M2 forward primer (5'- GAGCAATCAGTGC(AG)TCAGA-3') and the M824 reverse primer (5'- CTTGCA(AG)GC(CT)TCAATGAATG-3'). A 999 bp fragment of the RdRp gene was obtained in two independent amplifications with the L7541 forward primer (5'-CTCTGATAAATGATGCAT-3') and L8250 reverse primer (5'- (AG)GG(AG)CAATTGACATTCTT-3'); and the L7109 forward primer (5'- T(AG)AGGCT(CT)CCTGAAGTAGG-3') and L7750Y reverse primer (5'- ATGTGGCATATGT(CT)CTGC-3'). All oligonucleotide primers used in this study were synthesized by Integrated DNA Technologies, Inc.. (Coralville, IA). cDNA was synthesized with AMV Reverse Transcriptase (Promega) using the reverse primer by the method of Law and Moyer (1990) (41). PCR amplification was performed in a total volume of 25 µl with 1X PCR buffer (Promega), 25 µM MgCl<sub>2</sub>, 2.5 µM dNTPs (Promega), GoTaq<sup>®</sup> polymerase (Promega) and 10 µM of each forward and reverse primer for each locus with 2.5 µl of cDNA. The cycles were: 94°C for 10 min, followed by 40 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, and an extension at 72°C for 10 min. Amplified DNA was subjected to electrophoresis in 1% agarose and visualized with UV transillumination. A 1 kb+ ladder (Promega) was used to determine size of the amplified product. Amplified products were cleaned with either a QIAquick PCR Purification (Qiagen) or a PerfectPrep PCR Cleanup 96 Kit (Eppendorf, Germany) according to the manufacturer's protocol. Sequencing reactions were performed with 3730 xl DNA Analyzers (Applied Biosystems, Foster City, CA) using Big Dye chemistry (version 3) at the Duke University Institute for Genomic Sciences and Policy DNA Sequencing Facility (Durham, NC). Forward and reverse chromatograms were aligned using Sequencher<sup>™</sup> version 4.6 (Gene Codes Corporation, Ann Arbor, MI) and with

visual examination. All sequences for each locus were submitted to GenBank (N accession numbers FJ234451-FJ234639, NSm accession numbers FJ234640-FJ234829, and RdRP accession numbers FJ261988-FJ262170).

**Haplotype, site compatibility and phylogenetic analysis of cDNA sequences.** A suite of nucleotide analyses programs, SNAP Workbench, was used in this study for phylogenetic and population genetics analyses of the sequence data (57). Briefly, the sequences from each separate locus were aligned with ClustalW version 1.7 (79) and collapsed into haplotypes after removing insertions, deletions or infinite sites violations with SNAP Map (5).

Site compatibility matrices for each locus or population were generated using SNAP Clade (7). Compatibility matrices were graphically illustrated using SNAP Matrix (7) and Inkscape version 0.45.1. Matrices were examined for incompatible sites (8, 36) that were manually removed from the dataset using SNAP Combine (5). The removal of incompatible sites is a prerequisite step needed for other analyses used in this study that assume no conflict in the data (4).

Maximum parsimony (MP) analysis was used to examine grouping of isolates by minimizing the number of mutations required to reconstruct the phylogeny. MP analysis was performed for each locus with unweighted parsimony using PAUP\* 4.0 (72). The heuristic search was set for randomized branch swapping and 500 bootstrap replicates. The maximum number of generated parsimony trees was set at 500 with no increases. Bootstrap confidence limits were obtained with 500 replicates. If more than one most parsimonious tree was generated, a strict consensus tree was inferred. Maximum-likelihood (ML) analysis was performed by using the program ModelTest to select a best-fit nucleotide substitution model

based on hierarchical likelihood ratio tests (HLRT) and the Akaike Information Criterion (AIC) (56) in conjunction with PAUP\* 4.0 (72). ML calculates the likelihood of a predictive model producing the observed data (62), but relies on the investigator to provide a model of sequence evolution to improve the approximation of the evolutionary history of the population (52, 55). When the choices of models differed, the AIC best-fit nucleotide substitution model was chosen (52) to provide the prior assumptions to infer each phylogenetic tree. To reconstruct the phylogenetic tree, a quartet puzzling algorithm (70, 71) was applied under the maximum-likelihood criterion using PAUP\* 4.0 (72). The algorithm was modified according to parameters specified in the ModelTest AIC outfile for each locus. The number of puzzling steps was set to 1000 and quartets were evaluated using exact likelihood calculations. All phylogenetic trees were visualized with Mega 4.0 (77).

**Neutrality and population subdivision.** The population mutation rate, theta ( $\theta$ W), was estimated from the number of segregating sites,  $S$  (86), and  $k$ , the average pairwise distance between sequences (75). DnaSP version 4.50 (63) was used to estimate haplotype diversity ( $H_d$ ) (45, 46) and calculate test statistics and p-values for Tajima's  $D$  (74), Fu and Li's  $F^*$  and  $D^*$  without an outgroup (23), Fu's  $F_s$  (22) and Wright's  $F_{ST}$ , a measure of the proportion of the total genetic variance contained in a subpopulation relative to the total genetic variance (89). The average number of pairwise differences per nonsynonymous site,  $dN$ , the average number of pairwise differences per synonymous site,  $dS$ , and their ratio,  $dN/dS$  ( $\omega$ ), were calculated for each locus using the counting method of Yang and Nielsen (93) as implemented in the YN00 program in PAML version 4 (91).

To test for differentiation in the TSWV population, SNAP Map was used to generate files for Seqtomatrix which converts sequence data into a distance matrix. The test PermtestK uses the molecular variation to test if two sampled localities are genetically different (30). Hudson's estimates of  $K_S$ ,  $K_T$ ,  $K_{ST}$  and  $S_{nn}$  were used to determine the presence of subdivision in populations (29, 30).

## **RESULTS**

**Haplotype analysis and site compatibility of cDNA sequences.** A total of 2418 nucleotides from the N, NSm and RdRP genomic regions of TSWV were amplified for each isolate and collapsed into haplotypes (Table 2). The haplotype structure in each of the three loci consisted of one major haplotype, which contained the majority of isolates, and more than 100 minor haplotypes frequently represented by a single isolate (singleton haplotype). One hundred fifty haplotypes from 189 isolates were identified in the N gene, which had a haplotype diversity of 99.3%. One hundred fifty-one haplotypes from 190 isolates were identified in the NSm gene, which had a haplotype diversity of 99.7%. One hundred sixty-seven haplotypes from 183 isolates were identified in the RdRP gene, which had a haplotype diversity of 99.9%. The largest major haplotype for each locus contained 8.5%, 6.8%, 2.2% and of all isolates of the N, NSm, and RdRP gene sequences, respectively. The NSm gene had the highest percentage, 29.4%, of haplotypes that contained more than a single isolate and the RdRP gene had the lowest percentage of 14.7%. The RdRP gene had 156 singleton haplotypes while the N gene and NSm genes both had 135 singleton haplotypes each.

The specific isolate composition of each haplotype differed for each locus with one exception: the only haplotype that was maintained across all three loci consisted of F22N7 and F40N7, two isolates from NC\_12C in Suffolk County, VA in 2007. Haplotypes that contained more than one isolate and a mixture of isolate categories, such as collection year and geographical origin, were examined for patterns of association and frequency (Table 3). The major haplotype for the N (N1) and NSm (M144) loci contained isolates from each year, geographic area, and several cultivars. In contrast, the major haplotype for the RdRP locus (R166) contained only isolates that were sampled from peanuts in Bertie County in 2007. Further examination of the relevant minor haplotypes for each locus showed similar breadth of diversity across isolate characteristics. Two of the minor haplotypes of the N locus and 1 of the NSm locus were found each collection year of the study. One haplotype of the N locus and 2 of the NSm locus were found in each geographic area of the study. Three haplotypes of both the N and NSm loci were found in at least three different cultivars during the study. In contrast, only 1 haplotype besides the major haplotype contained more than 1 cultivar or geographic area in the RdRP locus. All other haplotypes in all three loci that were not singletons were not distinguished by differences in year, geographic area or cultivar.

The population statistics were determined for each locus and compared (Table 4). Nucleotide diversity was variable among the loci. The genetic distances for the N and NSm were similar at 6.907 and 6.806, respectively, while the RdRP locus had the highest average pairwise distance of 13.035; this also serves as an estimate of the population mutation rate,  $\theta$  (75). Watterson's estimates of average  $\theta$  were approximately 5 times higher than Tajima's estimate for each locus, which is indicative of over-inflation. The genetic distances at

nonsynonymous sites between isolates were highest in the NSm gene (0.003), but similar for the N and RdRP locus at 0.0015 and 0.0019, respectively. The genetic distances at synonymous sites between isolates were highest in the RdRP gene (0.0471), but similar for the N and NSm at 0.0335 and 0.0303, respectively. The values of the dN/dS ratio,  $\omega$ , were less than 1.0 for each locus (0.045, 0.099 and 0.039 for the N, NSm, and RdRP, respectively), which indicates negative selection on each locus (93).

When examining the site compatibility matrices for each locus, boundaries for partitions of distinct evolutionary history could not be determined because the sites of incompatibility did not appear to form any distinct patterns. These sites represented areas of conflict in the data, which could be interpreted as possible different evolutionary branch histories. They were removed from each locus alignment when performing phylogenetic analysis and neutrality testing, as required by model assumptions (4). The N gene had 47 incompatible sites removed, the NSm had 64 incompatible sites removed, and the RdRP had 112 incompatible sites removed.

**Phylogenetic analysis of cDNA sequences.** Maximum parsimony (MP) analysis of the N gene showed unresolved comb structures where numerous branches radiated from the main branch with no definition of descent (Data not shown). Bootstrap analysis indicated low support (< 70%) for the branching topology of the N and RdRP MP trees, but four branches of the NSm MP tree were supported (>94%) for groups containing isolates from the same geographic area. No discernible pattern of association with geographic area, year or peanut cultivar was found for each locus. Each MP tree showed some grouping characterized by geographic area and year, but most branches were examples of singleton haplotypes.

The best-fit nucleotide models differed for the three loci. The N gene was best modeled by the General Time Reversible model (GTR), which is characterized by all variable base, transition and transversion frequencies (61), with a gamma distributed rate variation of 1.4805 across sites. The NSm gene was best modeled by the Tamura-Nei model (TrN), which is characterized by variable base frequencies, variable transition frequencies and equal transversion frequencies (78). The gamma distributed rate variation was 0.9352 across sites. The RdRP is best modeled by the GTR with a 0.2936 proportion of invariable sites and a gamma distributed rate variation of 1.1497 across sites.

The ML quartet puzzling analysis revealed exponential population growth patterns for each locus (Figs. 1, 2, 3). In all three loci, the branches were shortest when closest to the main trunk and subsequent branching of secondary nodes was lengthened when further away from the main trunk, which indicates exponential growth of the population (25). The spatial structure of the isolate groupings in each locus was also weakly defined by geographic area, year, or peanut cultivar (25) and showed no characteristic grouping or clade formation pattern.

**Neutrality and population subdivision.** The values for the Tajima's D, Fu and Li's D\* and F\*, and Fu's Fs test statistics were significant and negative, which indicates significant deviation from neutrality (Table 5). Hudson's test statistics and Wright's  $F_{ST}$  indicated that the population was panmictic (Table 6). For all loci, the calculated value of Hudson's  $S_{nn}$  was close to 0.5 and not significant, which indicates no population differentiation.

Calculations of Wright's  $F_{ST}$  for the N, NSm and RdRP gene (the maximum value of 0.011 for both the N and NSm genes, and 0.047 for the RdRP gene) indicated that the percentage of

genetic differentiation accounting for the total amount of genetic variation did not exceed 11% in any comparison.

## **DISCUSSION**

Our sampling method used in this study was sufficient to obtain the maximum amount of genetic variation data with a minimum of sampling effort (21, 51). The amount of genetic diversity found in our population from peanut in NC and VA is similar to that previously reported for global populations of TSWV (82) and other plant RNA viruses (24). In all three loci examined, there was one major haplotype and more than 100 minor haplotypes that were sampled from the population. Similar population structures have been shown for other ssRNA plant virus populations such as *Turnip yellow mosaic virus* (24).

We also provide evidence that specific genes of TSWV can be sampled repeatedly from season to season. In general, isolate composition for each haplotype was not restricted by geographic area, year or cultivar. The major haplotypes for the N and NSm loci were found in each geographic area each sampled field season and in Sampson County after sampling for only one season. This pattern suggests that there are segments of the genome which are persistent in the genetic population from season to season. These segments are recombined with other segments during replication, such as in reassortment.

Our haplotype analysis measured the number of haplotypes, their frequency and diversity, and genetic distances within and between sequences (24). According to Moury et al. (2006), these three parameters are required to accurately measure the genetic diversity of a viral population (42). To calculate genetic distance, we used Tajima's calculations of S, the

number of segregating sites, and  $\pi$ , the average number of nucleotide differences (73), because of *a priori* knowledge of the non-equilibrium state of the population (i.e. the occurrence of bottlenecks) (37, 38). The most appropriate estimate of the population mutation rate,  $\theta$ , for our data is Tajima's estimate, which is based on the average number of pairwise distances (75). Watterson's estimate of  $\theta$  (86) is based upon the number of segregating sites, which is influenced by the size of the current population, and can be inflated by the presence of numerous unique sequences (73). Our data showed that Watterson's estimates of  $\theta$  were approximately five times greater than Tajima's estimate. This is supported by numerous recent mutations indicated in our data by significant values of the neutrality tests and frequent occurrence of singleton haplotypes.

The measures of haplotype diversity,  $H_d$ , can range from zero, meaning no diversity, to 1.000, which indicates high levels of haplotype diversity (46). For our data,  $H_d$  ranged from 0.993 to 0.999, indicating very high levels of diversity for each locus. The RdRP locus showed the most diversity, based on the number of haplotypes (e.g., 167 haplotypes from 183 isolates), their frequency in the population and the estimates of genetic distance. Other studies of Tospoviruses have focused on the diversity of the N gene for strain identification (14, 16, 47, 48, 66, 94) or the NSs (13), but none have examined the RdRP region to measure its diversity. The RdRP functions as a polymerase in the infected plant cell (10) and, in our data, appears to be under the same amount of purifying selection as the N and NSm loci, based on data from the neutrality tests and values of  $\omega$ , the dN/dS ratio. However, it could be that our data showed that RdRP has the most diversity because the higher numbers of segregating sites in the longer sequence lengths were creating more haplotypes when the data

was collapsed. An interesting follow up study would be to test haplotype diversity of these loci is influenced by the number of nucleotides examined. This should also include non-coding regions of the TSWV genome to provide enhanced statistical benefits to the data analysis (62).

We have provided evidence that the population of TSWV is panmictic across the coastal plain region of North Carolina and Virginia. We examined the data for the presence of subdivision in our samples through the use of Hudson's test statistics,  $PermtestK$  and  $Snn$ , and the index for Wright's  $F_{ST}$  (1969, 1978).  $PermtestK$  is powerful when sample sizes are small, but it can be compromised when there are unequal sample sizes (30) whereas Hudson's  $Snn$  is powerful at all sample sizes and diversities (29). A value of the  $Snn$  statistic close to 1 indicates high subpopulation differentiation and a value less than or equal to 0.5 indicates panmixia (29). In our data, the value of the  $Snn$  statistic depended on the presence of conflict from incompatible sites or recombination in the data set. In data without conflict, the  $Snn$  value was closer to 0.5. In contrast, when conflict was present, the value ranged from 0.6 to 0.9, which suggested that the conflict influenced the subdivision results. As a further check, we estimated  $F_{ST}$  as a measure of the genetic structure. Its values can range from 0 to 1.000 and values of  $F_{ST}$  greater than 0.05 (5%) suggest a degree of differentiation among populations. Our data showed that in data both with and without incompatible sites, the percentage of genetic differentiation accounting for the total amount of genetic variation was less than 11%. We therefore concluded that the subdivision seen in our data was influenced by the recombination occurring in the population and was not a function of subspecies divergence. The presence of subdivision in our sequence data with conflict

suggests that this subdivision has arisen more recently within each geographical population as recombination.

Estimates of the population summary statistics were similar in value across all loci, which could be possibly explained by the ecology of the virus population and sampling from a limited geographic area and a single host. The levels of nucleotide diversity of the N gene sampled from our population is similar to other N gene studies with *Cucurbit yellow stunting disorder virus* (.002) and *Sweet potato chlorotic stunt virus* (.006) (24). Nucleotide diversity calculations reported by Tsompana et al. (2005) were higher for their sampled N (0.024) and NSm (0.030) genes of TSWV. This study was conducted on isolates from 10 different host plants, such as chrysanthemum and pepper, from 11 different geographic areas. This global collection would be expected to show higher diversity due to the wide range of sampling. A similar recent study from South Africa on TSWV N gene did not provide an estimate of haplotype or nucleotide diversity on a small sample of six isolates (67). Therefore, it cannot be compared to our data; however, the authors state that their isolates exhibited low diversity based on sequence similarities and distance phylogeny methods. While this is not sufficient for determining evolutionary relationships, their conclusions support the findings of localized evolution presented in this study.

We showed that the population is experiencing strong negative selection for each of the three genes, as has been shown in other TSWV studies (82). To further examine evolutionary forces acting on the populations of TSWV, four neutrality test statistics (Tajima's D, Fu and Li's D\* and F\*, and Fu's Fs) were used in this study to examine the sequence data for departure from neutrality. All of the Tajima's D test statistics were

significant and negative, which is a strong indication that each of the three loci are experiencing population bottlenecks, where the population is severely limited by some event and only a few sequences start the new population (74). Almost all  $D^*$  and  $F^*$  test statistics were significant and negative, which implies that background selection (22), in which strong negative selection suppresses variation in linked regions, and rapid population growth have both occurred. Also, the large significant negative values of  $F_u$ 's  $F_s$  suggest that both genetic hitchhiking, which can happen when linked alleles become more common through positive selection, and rapid population growth have occurred in the population.  $F_u$ 's  $F_s$  is also an indicator of the presence of recombination (22), which is consistent with the results of the compatibility matrices and our conclusions from the subdivision testing. These results also strongly support our conclusions of population growth from our examination of haplotype diversity, population summary statistics and phylogenetic analysis.

As a further measure of selection, we calculated values for  $\omega$ , a comparison of synonymous to non-synonymous nucleotide substitution rates. This comparison detects and measures the selective pressure on encoded proteins. We used the counting method of Yang and Nielsen (2000) that is implemented in PAML, which takes into account the transition/transversions rate bias and the codon frequency bias while counting sites and differences (93). This method of estimating the  $\omega$  ratio has been shown to be the most reliable compared to other computations (92). For our data, the calculation of the  $\omega$  ratio for each gene was less than 1, which indicates strong negative selection acting on each locus (93), especially for the NSm gene. These results also support the conclusions from the neutrality tests of strong negative selection acting on the loci. Although our calculated

values of  $\omega$  ratio for the N gene were lower than previous estimates of 0.221 published for TSWV (82), this could indicate that there is stronger selection in a more localized population than might be seen on a global scale. There are a number of elements of the biology and ecology of TSWV that could possibly account for strong negative selection, such as population bottlenecks through vector transmission and host compatibility.

Parsimony and neighbor-joining (NJ) methods are traditionally used to reconstruct and infer phylogenetic relationships. However, both of these methods have limitations which can be avoided by using maximum-likelihood (ML) methods. Parsimony methods assume that shared characteristics are a result of common descent (26) and group resultant clades according to minimum numbers of mutations. This method can produce numerous trees of equally parsimonious probability and cannot score different mutation rates on different branches (28). For our data, we used parsimony trees as a preliminary investigation of clade formation and population structure, but not as a method of inferring evolutionary relationships. Although our sampling should have been sufficiently dense, even when we removed conflict from our data unresolved polytomies occurred at the main branch, which prevented resolution of relationships between isolates.

Because the MP trees were not sufficient to infer phylogenies for each locus, ML analysis with a quartet puzzling algorithm was performed to estimate a majority rule consensus tree after determining the best-fit nucleotide model. For our data, we found the best-fit nucleotide substitution rates models were different for each of our genes, which may indicate that each gene is experiencing different evolutionary forces. It is possible that specific regions of each locus can have different evolutionary histories due to recombination

(36). This seems likely because each gene encodes a protein that performs a different function in the ecology of the virus. Using these best-fit models to construct a ML phylogenetic tree provides a more accurate way to model the population structure. In general, there were several examples of minor incongruence between the MP and ML tree of each locus, but the general order of the tree branching was similar. The topology of the generated ML trees can provide information for building a phylodynamic framework to characterize TSWV as a pathogen (25). For our data, each locus ML tree shows evidence of exponential growth in a strongly selective environment. These conclusions are supported by our results from the neutrality tests and measurements of genetic diversity. It is possible that the environment provides strong selection from the many different host plants and vectors that are known to harbor TSWV. This selection and the high adaptability of the virus genome may be combining to drive TSWV genetic variability, as is seen in HIV and influenza populations (25). As a result, TSWV demonstrates increased evasion of plant resistance in the field.

Because the topology of phylogenetic trees can also be affected by recombination, leading to errors in phylogeny inferences, we investigated if recombination was present in the populations. If recombination is not addressed in phylogenetic analysis, it can negatively impact parameter estimation by overestimation and lead to incorrect inferences (4, 53). Furthermore, an assumption of neutrality testing is that the data contains no recombination (22, 23, 74). Our method of detecting recombination with compatibility matrices is most effective in detecting recombination in sequences with high mutation rates (54) and we were able to detect recombination in all three of our genes, but not to quantify it. While it is

currently unknown when in the TSWV infection cycle that recombination may take place, most recombination in viruses is assumed to occur during replication in the host (24). The virus is transmitted by at least 10 species of thrips (50, 65, 87) in a persistent and circulative manner and it also replicates in its vector (83, 88) and secondary weed hosts (9, 12). It has also been shown that TSWV can occupy plants at the same time as a closely related virus, *Impatiens necrotic spot virus* (27) and that reassortment does occur in the TSWV genome (58, 59). Because of its ecology, it appears that TSWV has numerous opportunities to exchange genetic information with other strains or other viruses. Therefore further investigation of the occurrence and rate of recombination for this virus is critical to correctly infer phylogenetic relationships and estimate population summary statistics.

Future evolutionary studies should consider including other loci, such as the NSs, which has been shown to be a gene silencing protein (76) and G1/G2, which have both been shown to interact with thrips vector gut cells (44). Examining these genes under a similar sampling strategy could further elucidate important epidemiological patterns of the disease. It will also be critical to maintain a frequent collection of TSWV in the same localized areas to examine the range of evolutionary rates of heterochronous data, as has been done for other RNA plant viruses such as *Rice yellow mottle virus* (19, 20).

## LITERATURE CITED

1. Adkins S. 2000. Tomato spotted wilt virus—positive steps towards negative success. *Mol. Plant Pathol.* 1 3:151-157.
2. Adkins S., Quadt R., Choi T. -, Ahlquist P., and German T. L. 1995. An RNA-Dependent RNA Polymerase Activity Associated with Virions of Tomato Spotted Wilt Virus, a Plant- and Insect-Infecting Bunyavirus. *Virology* 207 1:308-311.
3. Agrios G. N. 2005. *Plant Pathology*. 5th ed. Agrios GN, editor. Elsevier Academic Press, New York.
4. Awadalla P. 2003. The evolutionary genomics of pathogen recombination. *Nat. Rev. Genet.* 4 1:50-60.
5. Aylor D. L., Price E. W., and Carbone I. 2006. SNAP: Combine and Map modules for multilocus population genetic analysis. *Bioinformatics* 22:1399-1401.
6. Best R. J. 1968. Tomato Spotted Wilt Virus. *Adv. Virus Res.* 13:65-145.
7. Bowden L. C., Price E. W., and Carbone I. 2008. SNAP Clade and Matrix. Department of Plant Pathology, North Carolina State University 2.0.  
<http://www.cals.ncsu.edu/plantpath/faculty/carbone/home.html>.
8. Carbone I., Yir-Chung L., Hillman B. I., and Milgroom M. G. 2004. Recombination and Migration of *Cryphonectria hypovirus 1* as Inferred From Gene Genealogies and the Coalescent. *Genetics* 166 4:1611-1629.
9. Chamberlin J. R., Culbreath A. K., Todd J. W., and Demski J. W. 1993. Detection of Tomato spotted wilt virus in Tobacco thrips (Thysanoptera: Thripidae) overwintering in harvested peanut fields. *J. Econ. Entomol.* 86 1:40-45.

10. Chapman E. J., Hilson P., and German T. L. 2003. Association of L protein and in vitro Tomato spotted wilt virus RNA-dependent RNA polymerase activity. *Intervirology* 46:177-181.
11. Chatzivassiliou E. K. 2008. Management of the Spread of Tomato spotted wilt virus in Tobacco Crops with Insecticides Based on Estimates of Thrips Infestation and Virus Incidence. *Plant Dis.* 92 7:1012-1020.
12. Chatzivassiliou E. K., Peters D., and Katis N. I. 2007. The Role of Weeds in the Spread of Tomato Spotted Wilt Virus by Thrips *tabaci* (Thysanoptera: Thripidae) in Tobacco Crops. *J. Phytopathol.* 155 11:699-705.
13. Chen T., Huang C., Kuo Y., Liu F., Yuan C. H., Hsu H., and Yeh S.. 2006. Identification of Common Epitopes on a Conserved Region of NSs Proteins Among Tospoviruses of Watermelon silver mottle virus Serogroup. *Phytopathology* 96 12:1296-1304.
14. Chiemsombat P., Gajanandana O., Warin N., Hongprayoon R., Bhunchoth A., and Pongsapich P. 2008. Biological and molecular characterization of tospoviruses in Thailand. *Arch. Virol.* 153 3:571-577.
15. Culbreath A. K., Todd J. W., and Brown S. L. 2003. Epidemiology and management of Tomato spotted wilt in peanut. *Annu. Rev. Phytopathol.* 41 1:53-75.
16. de Avila A. C., de Haan P., Kormelink R., Resende R. d. O., Goldbach R. W., and Peters D. 1993. Classification of tospoviruses based on phylogeny of nucleoprotein gene sequences. *J. Gen. Virol.* 74:153-159.
17. de Haan P., Kormelink R., Resende R. d. O., van Poelwijk F., Peters D., and Goldbach R. W. 1991. Tomato spotted wilt virus L RNA encodes a putative RNA polymerase. *J. Gen. Virol.* 72 9:2207-2216.
18. de Haan P., Wagemakers L., Peters D., and Goldbach R. W. 1990. The S RNA segment of tomato spotted wilt virus has an ambisense character. *J. Gen. Virol.* 71:1001-1007.

19. Fargette D., Pinel A., Abubakar Z., Traoré O., Brugidou C., Fatogoma S., Hébrard E., Choisy M., Séré Y., Fauquet C. M., and Konaté G. 2004. Inferring the Evolutionary History of Rice Yellow Mottle Virus from Genomic, Phylogenetic, and Phylogeographic Studies. *J. Virol.* 78 7:3252-3261.
20. Fargette D., Pinel A., Rakotomalala M., Sangu E., Traoré O., Sérémé D., Sorho F., Issaka S., Hébrard E., Séré Y., Kanyeka Z., and Konaté G. 2008. Rice Yellow Mottle Virus, an RNA Plant Virus, Evolves as Rapidly as Most RNA Animal Viruses. *J. Virol.* 82 7:19-19.
21. Felsenstein J. 2006. Accuracy of Coalescent Likelihood Estimates: Do We Need More Sites, More Sequences, or More Loci? *Mol. Biol. Evol.* 23 3:691-700.
22. Fu Y. 1997. Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* 147 2:915-925.
23. Fu Y., and Li W. H. 1993. Statistical Tests of Neutrality of Mutations. *Genetics* 133 3:693-709.
24. García-Arenal F., Fraile A., and Malpica J. M. 2001. Variability and Genetic Structure of Plant Virus Populations. *Annu. Rev. Phytopathol.* 39 1:157.
25. Grenfell B. T., Pybus O. G., Gog J. R., Wood J. L. N., Daly J. M., Mumford J. A., and Holmes E. C. 2004. Unifying the Epidemiological and Evolutionary Dynamics of Pathogens. *Science* 303 5656:327-332.
26. Harrison C. J., and Langdale J. A. 2006. A step by step guide to phylogeny reconstruction. *Plant J.* 45 4:561-572.
27. Hoffmann K., Geske S. M., and Moyer J. W. 1998. Pathogenesis of tomato spotted wilt virus in peanut plants dually infected with peanut mottle virus. *Plant Dis.* 82 6:610-614.
28. Holder M., and Lewis P. O. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nat. Rev. Genet.* 4 4:275-284.

29. Hudson R. R. 2000. A new statistic for detecting genetic differentiation. *Genetics* 155:2011-2014.
30. Hudson R. R., Boos D. D., and Kaplan N. L. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol.* 9 1:138-151.
31. Tomato spotted wilt virus (CMI/AAB Descriptions of Plant Viruses, no. 39). 1970. Available from: <http://www.dpvweb.net/dpv/showdpv.php?dpvno=39>.
32. Isleib T. G., Rice P. W., Bailey J. E., Mozingo R. W., and Pattee H. E. 1997. Registration of 'NC\_12C' Peanut. *Crop Sci.* 37 6:1976.
33. Isleib T. G., Rice P. W., Mozingo, R. W., II, Bailey J. E., Mozingo R. W., and Pattee H. E. 2003. Registration of 'Perry' Peanut. *Crop Sci.* 43 2:739-740.
34. Isleib T. G., Rice P. W., Mozingo, R. W., II, Copeland S. C., Graeber J. B., Novitzky W. P., Pattee H. E., Sanders T. H., Mozingo R. W., and Coker D. L. 2006. Registration of 'Brantley' Peanut. *Crop Sci.* 46 5:2309-2311.
35. Isleib T. G., Rice R. W., Mozingo R. W., Mozingo R. W. I., and Pattee H. E. 1999. Registration of 'Gregory' peanut. *Crop Sci.* 39:1526.
36. Jakobsen I. B., Wilson S. R., and Easteal S. 1997. The partition matrix: exploring variable phylogenetic signals along nucleotide sequence alignments. *Mol. Biol. Evol.* 14 5:474-484.
37. Kaye A. C., Abad J., Shew B. B., Cubeta M. A., and Moyer J. W. 2006. Exploring the role of *Tomato spotted wilt virus* and soilborne fungi in yellowing syndrome of peanut. *Phytopathology* 96 S59.
38. Kaye A. C., Kennedy G. G., Shew B. B., Cubeta M. A., and Moyer J. W. 2008. Population genetic data analysis of *Tomato spotted wilt virus* on peanut in North Carolina and Virginia. *Phytopathology* 98 S79.

39. Kormelink R., de Haan P., Meurs C., Peters D., and Goldbach R. W. 1992. The nucleotide sequence of the M RNA segment of tomato spotted wilt virus, a bunyavirus with two ambisense RNA segments. *J. Gen. Virol.* 73 11:2795-2804.
40. Kormelink R., Storms M. M. H., van Lent J., Peters D., and Goldbach R. W. 1994. Expression and Subcellular Location of the NSm Protein of Tomato Spotted Wilt Virus (TSWV), a Putative Viral Movement Protein. *Virology* 200 1:56-65.
41. Law M. D., and Moyer J. W. 1990. A tomato spotted wilt-like virus with a serologically distinct N protein. *J. Gen. Virol.* 71:933-938.
42. Moury B., Desbiez C., Jacquemond M., and Lecoq H. 2006. Genetic diversity of plant virus populations: Towards hypothesis testing in molecular epidemiology. *Adv. Virus Res.* 67:49-87.
43. Moyer J. W. 2000. Tospoviruses. Pages 592-597 in: *Encyclopedia of microbiology*. in: Lederberg J., ed. Academic Press, San Diego.
44. Nagata T., Inoue-Nagata A. K., Prins M., Goldbach R. W., and Peters D. 2000. Impeded Thrips Transmission of Defective Tomato spotted wilt virus Isolates. *Phytopathology* 90 5:454-459.
45. Nei M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York.
46. Nei M., and Tajima F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97 1:145-163.
47. Nischwitz C., Pappu H. R., Mullis S. W., Sparks A. N., Langston D. R., Csinos A. S., and Gitaitis R. D. 2007. Phylogenetic Analysis of Iris yellow spot virus Isolates from Onion (*Allium cepa*) in Georgia (USA) and Peru. *J. Phytopathol.* 155 9:531-535.

48. Pappu H. R., du Toit L. J., Schwartz H. F., and Mohan S. K. 2006. Sequence diversity of the nucleoprotein gene of iris yellow spot virus (genus Tospovirus, family Bunyaviridae) isolates from the western region of the United States. *Arch. Virol.* 151 5:1015-1025.
49. Parrella G., Gognalons P., Gebre-Selassie K., Vovlas C., and Marchoux G. 2003. An update of the host range of *Tomato spotted wilt virus*. *J. Plant Pathol.* 85 4:227-264.
50. Pittman H. A. 1927. Spotted Wilt of Tomatoes. *Journ. Counc. Sci. Ind. Res.* 1:74-77.
51. Pluzhnikov A., and Donnelly P. 1996. Optimal Sequencing Strategies for Surveying Molecular Genetic Diversity. *Genetics* 144 3:1247-1262.
52. Posada D., and Buckley T. 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Syst. Biol.* 53 5:793-808.
53. Posada D., and Crandall K. A. 2002. The Effect of Recombination on the Accuracy of Phylogeny Estimation. *J. Mol. Evol.* 54 3:396-402.
54. Posada D., and Crandall K. A. 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. U. S. A.* 98 24:13757-13762.
55. Posada D., and Crandall K. A. 2001. Selecting the Best-Fit Model of Nucleotide Substitution. *Syst. Biol.* 50 4:580-601.
56. Posada D., and Crandall K. A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14 9:817-818.
57. Price E. W., and Carbone I. SNAP: workbench management tool for evolutionary population genetic analysis. *Bioinformatics.* 2005 21 3:402-404.

58. Qiu W. P., Geske S. M., Hickey C. M., and Moyer J. W. 1998. Tomato spotted wilt Tospovirus genome reassortment and genome segment-specific adaptation. *Virology* 244 1:186-194.
59. Qiu W. P., and Moyer J. W. 1999. Tomato Spotted Wilt Tospovirus adapts to the TSWV N gene-derived resistance by genome reassortment. *Phytopathology* 89 7:575-582.
60. Riley D. G., and Pappu H. R. 2004. Tactics for Management of Thrips (Thysanoptera: Thripidae) and Tomato Spotted Wilt Virus in Tomato. *J. Econ. Entomol.* 97 5:1648-1658.
61. Rodríguez F., Oliver J. L., Marín A., and Medina J. R. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* 142 4:485-501.
62. Rosenberg N. A., and Nordborg M. 2002. Genealogical Trees, Coalescent Theory and the Analysis of Genetic Polymorphisms. *Nat. Rev. Genet.* 3 5:380-390.
63. Rozas J., Sanchez-Delbarrio J. C., Messeguer X., and Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496-2497.
64. Salzman R. A., Hasegawa P. M., Bressan R. A., Fujita T., and Zhu-Salzman K. 1999. An Improved RNA Isolation Method For Plant Tissues Containing High Levels of Phenolic Compunds or Carbohydrates. *Plant Mol. Biol. Rep.* 17 1:11-17.
65. Samuel G., Bald J. G., and Pittman H. A. 1930. Ivestigations on "spotted wilt" of tomatoes, Australia. *Commonw. Counc. Sci. Ind. Res. Bull.* No. 44.
66. Singh L., Hallan V., Jabeen N., Singh A. K., Ram R., Martin D. P., and Zaidi A. A. 2007. Coat protein gene diversity among Chrysanthemum virus B isolates from India. *Arch. Virol.* 152 2:405-413.
67. Sivparsad B. J., and Gubba A. 2008. Isolation and molecular characterization of Tomato spotted wilt virus (TSWV) isolates occurring in South Africa. *Afr. J. Agric. Res.* 3 6:428-434.

68. Soellick T., Uhrig J. F., Bucher G. L., Kellmann J., and Schreier P. H. 2000. The Movement Protein NSm of Tomato Spotted Wilt Tospovirus (TSWV): RNA Binding, Interaction with the TSWV N Protein, and Identification of Interacting Plant Proteins. *Proc. Natl. Acad. Sci. U. S. A.* 97 5:2373-2378.
69. Storms M. M. H., Kormelink R., Peters D., Van Lent J. W. M., and Goldbach R. W. 1995. The nonstructural NSm protein of *Tomato Spotted Wilt Virus* induces tubular structures in plant and insect cells. *Virology* 214:485-493.
70. Strimmer K., and von Haeseler A. 1997. Likelihood-mapping: A simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U. S. A.* 94 13:6815-6819.
71. Strimmer K., and von Haeseler A. 1996. Quartet Puzzling: A Quartet Maximum-Likelihood Method for Reconstructing Tree Topologies. *Mol. Biol. Evol.* 13 7:964-969.
72. Swofford D. L. 2003. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. 4th ed. Sinauer Associates, Sunderland, Massachusetts.
73. Tajima F. 1989. The Effect of Change in Population Size on DNA Polymorphism. *Genetics* 123 3:597-601.
74. Tajima F. 1989. Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics* 123 3:585-595.
75. Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105 2:437-460.
76. Takeda A., Sugiyama K., Nagaro H., Mori M., Kaido M., Mise K., Tsuda S., and Okuno T. 2002. Identification of a novel RNA silencing suppressor, NSs protein of *Tomato Spotted Wilt Virus*. *FEBS* 532:75-79.

77. Tamura K., Dudley J., Nei M., and Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol.* 24 8:1596-1599.
78. Tamura K., and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10 3:512-526.
79. Thompson J. D., Higgins D. G., and Gibson T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nuc. Acids Res.* 22:4673-4680.
80. Tomlinson J. A. 1987. Epidemiology and control of virus diseases of vegetables. *Ann. Appl. Biol.* 110 3:661-681.
81. Tsompana M. 2004. Molecular evolution and population genetics of Tomato Spotted Wilt virus (TSWV) [dissertation]. United States -- North Carolina: North Carolina State University.
82. Tsompana M., Abad J., Purugganan M., and Moyer J. W. 2005. The molecular population genetics of the Tomato spotted wilt virus (TSWV) genome. *Mol. Ecol.* 14 1:53-66.
83. Ullman D. E., German T. L., Sherwood J. L., Westcot D. M., and Cantone F. A. 1993. Tosspovirus replication in insect vector cells: immunocytochemical evidence that the nonstructural protein encoded by the S RNA of Tomato spotted wilt tospovirus is present in thrips vector cells. *Phytopathology* 83 4:456-463.
84. Van den Bosch F., Akudibilah G., Seal S., and Jeger M. 2006. Host resistance and the evolutionary response of plant viruses. *J. Appl. Ecol.* 43 3:506-516.
85. Van den Hurk J., Tas P. W. L., and Peters D. 1977. The ribonucleic acid of Tomato spotted wilt virus. *J. Gen. Virol.* 36 1:81-91.

86. Watterson G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7 2:256-276.
87. Whitfield A. E., Ullman D. E., and German T. L. 2005. Tospovirus-thrips Interactions. *Annu. Rev. Phytopathol.* 43:459-489.
88. Wijkamp I., van Lent J., Kormelink R., Goldbach R. W., and Peters D. 1993. Multiplication of tomato spotted wilt virus in its insect vector, *Frankliniella occidentalis*. *J. Gen. Virol.* 74 3:341-349.
89. Wright S. 1951. The genetical structure of populations. *Ann. Eugenics* 15 4:323-354.
90. Wynne J. C., Beute M. K., Bailey J., and Mozingo R. W. 1991. Registration of 'NC-VII' Peanut. *Crop Sci.* 31 2:484-485.
91. Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 24 8:1586-1591.
92. Yang Z., and Bielawski J. P. 2000. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* 15 12:496-503.
93. Yang Z., and Nielsen R. 2000. Estimating Synonymous and Nonsynonymous Substitution Rates Under Realistic Evolutionary Models. *Mol. Biol. Evol.* 17 1:32-43.
94. Zheng Y. X., Chen C. C., Yang C. J., Yeh S., and Jan F. J. 2008. Identification and characterization of a tospovirus causing chlorotic ringspots on *Phalaenopsis* orchids. *Eur. J. Plant Pathol.* 120 2:199-209.

**TABLE 1.** Year of collection, geographic area, peanut cultivar and number of amplified nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) fragments of 196 isolates of *Tomato spotted wilt virus*

Year collected	Geographic area	Peanut cultivar	Loci <sup>a</sup>		
			N	NSm	RdRP
2005	Bertie County, NC	Brantley	6	6	6
		NC_12C	3	3	3
		Unknown	13	12	10
2006	Bertie County, NC	NC_12C	22	22	17
	Suffolk County, VA	Gregory	5	7	7
		Perry	12	12	12
2007	Bertie County, NC	Perry	36	36	36
		NC_12C	8	8	8
	Suffolk County, VA	NC_12C	42	42	42
	Sampson County, NC	Gregory	32	32	32
		NC-V11	10	10	10
Total number of amplified genes for each locus			189	190	183

<sup>a</sup> Total number of genes amplified from isolates used in this study

**TABLE 2.** Summary of the number of isolates, length of gene fragment, haplotype number, haplotype diversity, and haplotype frequencies of the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of *Tomato spotted wilt virus* on peanut

Gene	n <sup>a</sup>	L <sup>b</sup>	H <sup>c</sup>	Hd <sup>d</sup>	H percentage <sup>e</sup>	# Singleton haplotypes <sup>f</sup>
N	189	720	150	.993	29.1%	135
NSm	190	699	151	.997	29.4%	135
RdRP	183	999	167	.999	14.7%	156

<sup>a</sup> Total number of isolates used for each locus

<sup>b</sup> Total length of the sequence used for each locus

<sup>c</sup> Total number of unique haplotypes in each locus

<sup>d</sup> Measurement of haplotype diversity,  $\hat{h} = n(1 - \sum x_i^2) / (n-1)$  (Nei, 1987)

<sup>e</sup> Total percentage of isolates that collapsed into a haplotype containing more than one isolate

<sup>f</sup> Total number of haplotypes that contained only a single isolate

**TABLE 3.** Frequency of haplotypes in the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of *Tomato spotted wilt virus* that were distributed across different peanut cultivars in different geographical areas from 2005 to 2007

		N						
		N1 <sup>d</sup>	N34	N60	N73	N95	N138	N145
Year <sup>a</sup>								
	2005	2	-	-	-	2	2	1
	2006	4	-	1	-	-	2	3
	2007	10	2	1	4	-	2	2
Area <sup>b</sup>								
	Bertie	8	-	-	1	2	5	6
	Suffolk	7	1	2	1	-	1	-
	Sampson	1	1	-	2	-	-	-
Cultivar <sup>c</sup>								
	NC_12C	12	1	1	1	1	3	3
	Unknown	-	-	-	-	1	2	1
	Gregory	1	1	-	-	-	-	-
	Perry	3	-	1	1	-	1	2
	NC-V11	-	-	-	2	-	-	-

**Table 3.** Continued

NSm									RdRP	
M49	M68	M70	M103	M105	M108	M131	M141	M144 <sup>d</sup>	R151	R166 <sup>d</sup>
-	-	-	-	-	-	1	-	3	-	-
-	3	2	1	-	-	1	1	6	-	-
4	1	1	1	2	2	4	3	4	2	4
-	-	1	1	1	-	2	2	10	1	4
1	4	1	1	-	-	1	2	2	1	-
3	-	1	-	1	2	3	-	1	-	-
1	1	2	2	-	-	2	1	6	1	1
-	-	-	-	-	-	1	-	3	-	-
2	-	1	-	1	1	2	1	1	-	-
-	3	-	-	1	-	-	2	3	1	3
1	-	-	-	-	1	2	-	-	-	-

<sup>a</sup> Isolate collection year

<sup>b</sup> Geographical collection area

<sup>c</sup> Peanut cultivar sampled

<sup>d</sup> Major haplotype for each locus

**TABLE 4.** Summary of the number of segregating sites, nucleotide diversity, and estimates of the population mutation rate of the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of *Tomato spotted wilt virus* on peanut

Gene	S <sup>a</sup>	k <sup>b</sup>	$\pi^c$	nN <sup>d</sup>	dN <sup>e</sup>	nS <sup>f</sup>	dS <sup>g</sup>	$\omega^h$	$\theta W^i$	$\theta W/\text{site}$
N	176	6.907	0.010	565	0.0015	155	0.0335	0.045	30.26	0.042
NSm	214	6.806	0.010	546	0.003	153	0.0303	0.099	36.76	0.053
RdRP	335	13.035	0.013	794	0.0019	205	0.0471	0.039	57.919	0.058

<sup>a</sup> Total number of segregating sites

<sup>b</sup> Average number of nucleotide differences between sequences; Tajima's estimate of the population mutation rate,  $\theta$

<sup>c</sup> Average number of nucleotide differences per site

<sup>d</sup> Number of nonsynonymous sites

<sup>e</sup> Average number of pairwise differences of nonsynonymous sites

<sup>f</sup> Number of synonymous sites

<sup>g</sup> Average number of pairwise differences of synonymous sites

<sup>h</sup> Ratio of dN/dS

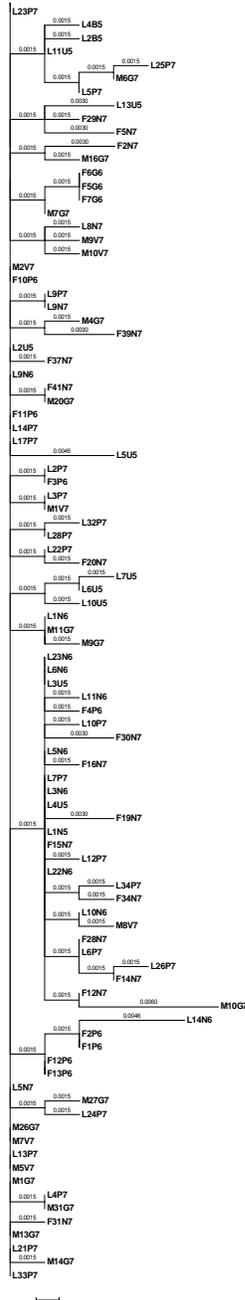
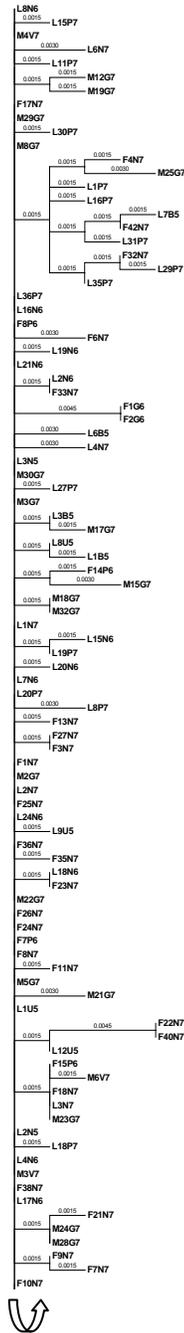
<sup>i</sup> Watterson's estimate of the population mutation rate,  $\theta$

**TABLE 5.** Summary of neutrality test statistic values for the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) loci of *Tomato spotted wilt virus* on peanut

Gene	Tajima's D	Fu and Li's $D^*$	Fu and Li's $F^*$	Fu's $F_s$
N	-2.732*	-7.183*	-5.925*	-20.614*
NSm	-2.739*	-8.612*	-6.776*	-19.938*
RdRP	-2.766*	-8.990*	-6.966*	-20.299*

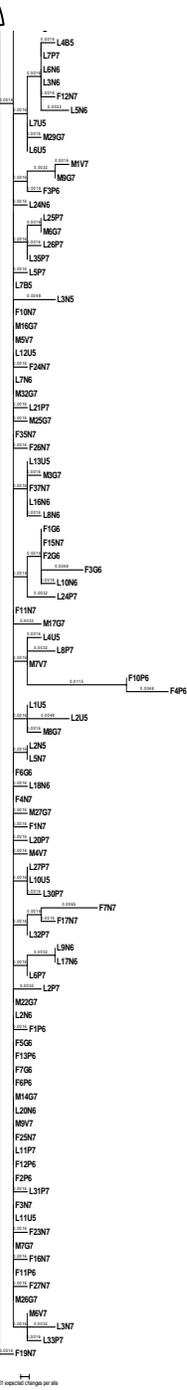
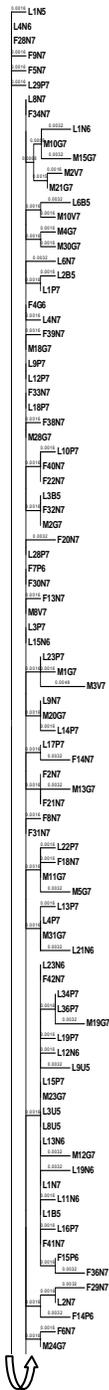
\* = indicates a significant result, ( $p < .001$ )

Figure. 1. Unrooted majority rules consensus tree of the N gene for isolates of *Tomato spotted wilt virus* obtained using nucleotide sequences without conflict. The tree was based on the General Time Reversible (GTR) model with a gamma distribution rate of 1.4805 across sites for quartet puzzling maximum-likelihood analysis. Tree length = 0.218003 and overall likelihood was  $\ln L = -1643.44$ .



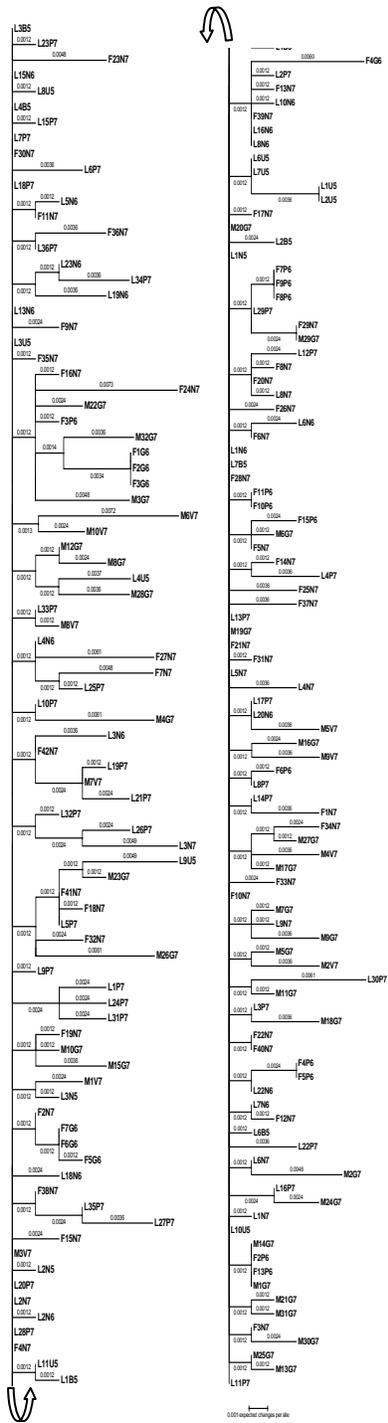
0.001 expected changes per site

Figure 2. Unrooted majority rules consensus tree of the NSm gene for isolates of *Tomato spotted wilt virus* obtained using nucleotide sequences without conflict. The tree was based on the Tamura-Nei model (TrN) with a gamma distributed rate variation was 0.9352 across sites for quartet puzzling maximum-likelihood analysis. Tree length = 0.266179 and overall likelihood was  $\ln L = -1648.33$ .



001 esperar cambios por día

Figure 3. Unrooted majority rules consensus tree of the RdRP gene for isolates of *Tomato spotted wilt virus* obtained using nucleotide sequences without conflict. The tree was based on the General Time Reversible (GTR) model with a 0.2936 proportion of invariable sites and a gamma distributed rate variation of 1.1497 across sites for quartet puzzling maximum-likelihood analysis. Tree length = 0.347142 and overall likelihood was  $\ln L = -2358.92$ .



## CHAPTER 2

### **Examining Recombination in the Nucleocapsid, Movement, and RNA-dependent RNA Polymerase Genes of Tomato Spotted Wilt Virus<sup>2</sup>**

Amanda C. Kaye

James W. Moyer

Elizabeth J. Parks

Ignazio Carbone

Marc A. Cubeta

*Department of Plant Pathology, North Carolina State University, Raleigh, North Carolina, 27695.*

---

<sup>2</sup> This chapter was formatted for submission to the journal Virus Research

## **Abstract**

Exploring the genetic diversity and evolutionary history of plant viruses is critical to understanding their ecology. High diversity in viruses has been attributed to the occurrence of recombination and mutation. Tomato spotted wilt virus (TSWV) is an ambisense RNA virus that infects over 1000 species of plants and has been shown to reassort its genome. Documenting recombination in this virus would allow quantification of recombination events, identification of areas prone to recombination, and to estimate population recombination rates. To investigate recombination in TSWV, matrix compatibility and ancestral recombination methods were used to detect, quantify, and reconstruct the history of recombination in three genes of TSWV in isolates collected from three cultivars of peanut (Gregory, NC\_12C, and Perry) sampled in multiple years. Site commonalities among the isolates collected from different cultivars were found in each gene where the majority of the recombination appeared to be occurring. These corresponded to sequence near the 3' terminus in the N and RdRP genes and to sequence near the 5' terminus in the NSm gene. Among isolates collected from different peanut cultivars, the Gregory isolates proportionally had the most recombination occurring in the N gene and the RdRP genes, while the Perry isolates had the most recombination occurring in the NSm. Recombination rates per site estimated for the N, NSm, and RdRP genes in the NC\_12C isolates were 0.00934, 0.01559, and 0.00495, respectively.

*Keywords:* Recombination; Ancestral recombination graph; TSWV

## 1. Introduction

RNA viruses are genetically heterogeneous and can exhibit wide intraspecies diversity (Moury et al., 2006). The source of this inherent genetic diversity is most commonly attributed to mutation due to the errors in replication by RNA dependent RNA polymerase and reassortment between genetic segments of the viral genome (Jeger et al., 2006). Previous studies have provided compelling evidence for the occurrence of recombination in positive-sense RNA plant viruses such as *Brome mosaic virus*, *Cucumber mosaic virus*, and *potato virus Y* (Bonnet et al., 2005; Bruyere et al., 2000; Chare and Holmes, 2006). However, in a survey of 79 sequence alignments, Chare et al. (2003) used phylogenetic incongruence, Sawyer's runs tests, and informative sites tests to determine that homologous recombination is uncommon in negative-sense RNA viruses. This study also included sequence analysis of the nucleocapsid gene of the plant virus, Tomato spotted wilt virus (TSWV), the type member of the genus Tospovirus in the family Bunyaviridae (Moyer, 2000). TSWV is a single-stranded, tripartite, ambisense RNA virus (Best, 1968; Van den Hurk et al., 1977) which has previously been shown to reassort its genomic segments (Qiu et al., 1998). While the phylogenetic methods employed by Chare et al. (2003) determined that there was no evidence of homologous recombination in TSWV, another member of Bunyaviridae, the animal virus *Hantaan virus*, was found to exhibit characteristics of recombination within genome segments. Given this evidence and because it is known that

TSWV adapts to host plant resistance by genome reassortment (Qiu and Moyer, 1999), we investigated the occurrence and frequency of homologous recombination in this virus.

In this paper, matrix incompatibility methods and ancestral recombination methods were used to detect and quantify recombination events in three gene regions of TSWV isolated from three different peanut cultivars. The first objective of this research was to document recombination in the nucleocapsid (N) gene, the movement (NSm) gene and the RNA-dependent RNA polymerase (RdRP) gene, and determine if there were specific regions within each gene that would exhibit more recombination than other regions. A second objective was to quantify the recombination in the isolates collected from cultivars with varying degrees of resistance, to determine if host resistance increased recombination occurring in the isolates. The third objective of this research was to estimate an overall population recombination rate of each gene region in a single cultivar from a single geographic area to determine if there were different rates for each gene region.

## **2. Materials and methods**

### *2.1 Isolates of Tomato spotted wilt virus*

A sample of 97 isolates of Tomato spotted wilt virus (TSWV) from peanut (*Arachis hypogaea* L.) leaves exhibiting typical symptoms of TSWV (e.g., ring spots, "rusting", stunting, shortened internodes, and bud necrosis) collected from 2005-2007 were used in this study (Table 1). The isolates represented a subsample of 196 isolates described previously

(Kaye et al., 2006; Kaye et al., 2008). The sample examined in this study consisted of 26 isolates collected from a single location in Bertie County, NC (77.1755° W, 36.1324° N) from a single cultivar of peanut (NC\_12C) over 2005-2007; 44 isolates collected from Suffolk County, VA (76.6097° W, 36.7411° N) and Bertie County, NC from a single cultivar of peanut (Perry) over 2006-2007; and 37 isolates collected from Suffolk County, VA and Bertie County, NC from a single cultivar of peanut (Gregory) over 2006-2007. The cultivars Perry (Isleib et al., 2003) and NC\_12C (Isleib et al., 1997) exhibit greater susceptibility to TSWV than Gregory (Isleib et al., 1999) when other crop production factors are considered (Hurt et al., 2003). The data was partitioned by peanut cultivar.

## *2.2 RNA Extraction, Reverse Transcription and cDNA Synthesis*

Total plant RNA was extracted from each leaf sample with either TRIzol Reagent (Invitrogen Corporation, Carlsbad, CA), RNeasy Plant Mini Kit (Qiagen, Valencia, CA) with 1% polyvinyl pyrrolidone in RLT buffer (Salzman et al., 1999), or directly from leaf sap-saturated Immunostrips using the RNeasy Plant Mini Kit with 1% polyvinyl pyrrolidone according to the manufacturer's protocol. Extracted RNA from each sample was used as a template for reverse transcription-polymerase chain reaction (RT-PCR) to generate complementary DNA (cDNA) from the nucleocapsid (N), movement (NSm), and RNA dependent RNA polymerase (RdRP) genomic regions of TSWV (Tsompana, 2004). A 720 bp region of the N gene region was amplified with the S1976 forward primer (5'-ACAGCTG(ACT)TTTTAAGCAAGTTCTG-3') and the S2847 reverse primer (5'-

AAGAAACGACTGCGG(AG)ATACA-3'). A 699 bp region of the NSm gene region was amplified with the M2 forward primer (5'- GAGCAATCAGTGC(AG)TCAGA-3') and the M824 reverse primer (5'- CTTGCA(AG)GC(CT)TCAATGAATG-3'). A 999 bp fragment of the RdRP gene region was obtained in two independent amplifications with the L7541 forward primer (5'-CTCTGATAAATGATGCAT-3') and L8250 reverse primer (5'- (AG)GG(AG)CAATTGACATTCTT-3'); and the L7109 forward primer (5'- T(AG)AGGCT(CT)CCTGAAGTAGG-3') and L7750Y reverse primer (5'- ATGTGGCATATGT(CT)CTGC-3') All oligonucleotide primers used in this study were synthesized by Integrated DNA Technologies, Inc.. (Coralville, IA).

cDNA was synthesized with AMV Reverse Transcriptase (Promega) using the reverse primer by the method of Law and Moyer (1990). PCR amplification was performed in a total volume of 25 µl with a solution of 1X PCR buffer (Promega), 2.5 µM MgCl<sub>2</sub>, 250 µM dNTPs (Promega), GoTaq<sup>®</sup> polymerase (Promega) and 10 pmoles of each forward and reverse primer for each locus, with 2.5 µl of cDNA. The reaction was initially denatured at 94°C for 10 min, followed by 40 cycles of the following: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min, then extended at 72°C for 10 min. Amplified DNA was subjected to electrophoresis in 1% agarose and visualized with UV transillumination. A 1 kb+ ladder (Promega) was used to determine size of the amplified product. Amplified fragments were cleaned with either a QIAquick PCR Purification (Qiagen) or a PerfectPrep PCR Cleanup 96 Kit (Eppendorf, Germany) according to the manufacturer's protocol. Sequencing reactions were performed with 3730 xl DNA Analyzers (Applied Biosystems, Foster City, CA) using Big Dye chemistry (version 3) at the Duke University Institute for Genomic Sciences and

Policy DNA Sequencing Facility (Durham, NC). Forward and reverse chromatograms were aligned using Sequencher™ version 4.6 (Gene Codes Corporation, Ann Arbor, MI) and with visual examination. All sequences for each locus were submitted to GenBank and are presented in Table 1.

### *2.3 Compatibility and recombination analysis*

A suite of nucleotide analyses programs, SNAP Workbench, was used in this study for analyses of the sequence data (Price and Carbone, 2005). Briefly, the sequences from each locus were aligned with ClustalW version 1.7 (Thompson et al., 1994) and collapsed into haplotypes after removing insertions, deletions or infinite sites violations with SNAP Map (Aylor et al., 2006). Site compatibility matrices for each locus of all isolates partitioned by cultivar were generated using SNAP Clade (Bowden et al., 2008). Compatibility matrices were graphically illustrated using SNAP Matrix (Bowden et al., 2008) and Inkscape version 0.45.1. Matrices were examined for incompatible sites (Carbone et al., 2004; Jakobsen et al., 1997). For each locus of the NC\_12C isolates only, we calculated  $\gamma$ , an estimator of the population mean recombination rate parameter,  $2N_e c$ , and the mean recombination rate per base<sup>3</sup>, using the program SITES version 1.1 (Hey and Wakeley, 1997). This parameter assumes an infinite sites model for the data. These isolates used for this calculation were collected from a single geographical area over three consecutive years. The estimates of  $\theta$  for each gene fragment were used with a base estimate of the neutral mutation rate,  $\mu$ , of  $1 \times$

---

<sup>3</sup> This rate was calculated by dividing the estimate of  $\gamma$  by the number of base pairs included in the analysis.

$10^{-3}$  substitutions per site per year (Jenkins et al., 2002) to estimate the effective population size,  $N_e$ , as  $\theta = 2N_e\mu$ . The estimate of  $N_e$  was used to algebraically estimate a value for  $c$ , the recombination rate per site per generation (Posada et al., 2002). The mutation and recombination histories for the separate gene regions of the isolates from each cultivar were estimated using the heuristic algorithm kwarg in the program beagle (Lyngsø et al., 2005). Kward assumes both mutation and recombination coalesce backward in time, and results in an estimated minimal recombination history. The resulting ancestral recombination graphs (ARGs) were visualized using GT MINER version 1.27, a component of SNAP Workbench (Carbone et al., 2007). In order to compare the amount of recombination across each gene region and cultivars, we calculated the ratio of the number of inferred recombination events to the number of possible recombination sites given by KWARG output. The ratios were all normalized to a lowest common denominator of 2.86 recombination sites to compare the number of recombination events between gene regions and between isolates collected from different cultivars.

### **3. Results**

#### *3.1 Compatibility and estimates of rates of recombination*

For the isolates collected from each cultivar, the site compatibility matrices for each gene region indicated that there were incompatible sites in the consensus sequence (data not shown). There were also no distinct patterns of recombination block partitioning in each of

the three genomic regions examined. In the isolates collected from NC\_12C, there were 2, 4 and 8 incompatible sites in the N, NSm and RdRP gene regions, respectively. In the isolates collected from Perry, there were 15, 12 and 21 incompatible sites in the N, NSm and RdRP gene regions, respectively. In the isolates collected from Gregory, there were 15, 10 and 25 incompatible sites in the N, NSm and RdRP gene regions, respectively.

For isolates sampled from NC\_12C, the estimated overall population mean recombination rate ( $\gamma$ ) per gene region ( $2N_e c$ ) was 6.726, 10.898, and 4.945 for the N, NSm, and RdRP gene regions, respectively. The mean recombination rate per site for the N, NSm, and RdRP gene region regions was 0.00934, 0.01559, and 0.00495, respectively. This corresponds to a rate ( $c$ ) of  $1.81 \times 10^{-3}$ ,  $1.70 \times 10^{-3}$ , and  $4.98 \times 10^{-4}$  per site per generation for the N, NSm, and RdRP gene region regions of the NC\_12C isolates, respectively.

### *3.1.1 Recombination in the N region*

The ARG for the N region of the NC\_12C isolates (Fig. 1) indicated that haplotype 3 (H3) was the sequence most closely related to the inferred ancestral haplotype, comprised of an isolate collected in 2007. Recombination was deep in the evolutionary history of this gene in this set of isolates, with the deepest recombination event between sites 603 and 604 being common to 15 of the 19 haplotypes; subsequent mutational events resulted in the formation of 14 of these haplotypes. The remaining haplotype with this recombination event in its history (H16) possessed both mutation and three subsequent recombination events; one at 603-604, one at 712-713, and finally one at 567-568. Four haplotypes had no recombination

in their histories (H3, H4, H6 and H18). For this N gene region, all three of the recombination sites occurred in the last third of the sequence near the 3' terminus.

Similarly, the ARG for the N gene region of the Perry isolates indicated that recombination was deep in the evolutionary history, with the deepest recombination event between sites 399-400 being common to 39 of the 41 haplotypes; subsequent mutational events resulted in the formation of 20 of these haplotypes (data not shown). This site was the source of a total of three recombination events that occurred further back in the history of the isolates. The next deepest recombination event between sites 282-283 was common to 5 of the 41 haplotypes, which all underwent at least two further recombination events before differentiating into haplotypes. The recombination site 531-532 is the source of five recombination events that occurred more recently in the history of seven of the haplotypes. Seven of the other recombination sites occurred in the history only once and four occurred twice. There was only one of 41 haplotypes that formed by mutation alone (H3), which was also the most closely related to the inferred ancestral haplotype and comprised of an isolate collected in 2007. In total, there were 25 recombination events that were estimated to have occurred over 14 different sites, which equal a ratio of 5.10 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). A recombination event at site 712-713 occurred once in the isolates from NC\_12C and twice in the isolates from Perry. For this N gene region region, six of the fourteen recombination sites occurred in the last third of the sequence near the 3' terminus.

Analysis of the ARG for the N gene region of the Gregory isolates indicated that none of the 34 haplotypes were differentiated solely by mutation (data not shown). The ARG

indicated that the deepest recombination event occurring between sites 627-628 formed the haplotype structure of all 37 Gregory isolates in this analysis. This site is the source of five recombination events throughout the haplotype history. The recombination site 636-634 and site 394-395 were each the source of four recombination events throughout the history of the isolates. The recombination sites 258-259, 266-267, and 558-559 were each the source of three recombination events. Three of the other recombination sites each occurred in the history twice and six sites occurred only once. There were 39 recombination events estimated to have occurred over 16 different sites, which equal a ratio of 6.97 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). A recombination event occurred at site 282-283 once in isolates from both Perry and Gregory. A recombination event occurred at site 531-532 five times in isolates from Perry and once in isolates from Gregory. For this N gene region region, six of the sixteen recombination sites occurred in the last third of the analyzed sequence data near the 3' terminus.

### *3.1.2 Recombination in the NSm region*

The ARG for the NSm gene region of the NC\_12C isolates (Fig. 2) indicated that H20 was the most closely related to the inferred ancestral sequence with no mutations and comprised of an isolate collected in 2007. Sixteen of the 23 haplotypes were formed through mutation alone. The deepest recombination event that occurred between sites 312-313 was common to 5 of 23 haplotypes and subsequent mutation resulted in the formation of two of

these haplotypes. This site is the source of two recombination events in the history of these isolates. The sites 156-157, 183-184, and 517-518 were each the source of a single recombination event. There were a total of five recombination events that occurred over four recombination sites, which equal a ratio of 3.57 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). Two of the four recombination sites occurred in the first third of the analyzed sequence data near the 5' terminus.

In comparison, the ARG of the NSm gene region of the Perry isolates indicated that 5 of the 41 haplotypes had no recombination in their histories, but were differentiated by mutation (data not shown). The haplotype most closely related to the inferred ancestral sequence had no mutations differentiating it and was comprised of an isolate collected in 2007. The deepest recombination event between sites 491-492 was common to 36 of 41 haplotypes. This site is the source of two recombination events. The recombination site 231-232 and site 429-430 were each the source of three recombination events throughout the history of the Perry isolates. Three of the other recombination sites each occurred in the history twice and eight sites occurred only once. There were a total of 23 recombination events estimated to have occurred over 14 sites in the Perry isolates, which equal a ratio of 4.70 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). Five of the fourteen recombination sites in the Perry isolates occurred in the first third of the analyzed sequence data near the 5' terminus.

The ARG of the NSm gene region of the Gregory isolates indicated that 17 of the 33 haplotypes had no recombination in their histories, but were differentiated by mutation (data

not shown). The haplotype most closely related to the inferred ancestral sequence had three mutations differentiating it and was comprised of an isolate collected in 2007. The deepest recombination event between sites 429-430 was common to 12 of the 33 haplotypes. This site is the source of two recombination events deep in the history of the Gregory isolates. The recombination site between sites 675-676 was the source of two recombination events in the history of two haplotypes. The other 12 recombination sites were each the source of a single recombination event. A total of 16 recombination events were estimated to have occurred over 14 sites in the Gregory isolates, which equal a ratio of 3.27 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). A recombination event occurred once at sites 183-184 in both the NC\_12C and Gregory isolates. Another recombination event occurred at sites 312-313 twice in the NC\_12C and once in the Gregory isolates. A recombination event occurring at sites 429-430 occurred three times in the Perry isolates and twice in the Gregory isolates. A recombination event occurred once at sites 447-448 in both the Perry and Gregory isolates. Four of the fourteen recombination sites occurred in the first third of the analyzed sequence data near the 5' terminus in the Gregory isolates.

### *3.1.3 Recombination in the RdRP region*

The ARG for the RdRP region of the NC\_12C isolates (Fig. 3) showed that H5 was the most closely related to the inferred ancestral sequence and differed by a single mutation. This haplotype was comprised of an isolate collected in 2007. Fourteen of the 25 haplotypes

had no recombination in their histories. The deepest recombination event between sites 258-259 was common to 9 of 25 haplotypes. Three of the 9 haplotypes were subsequently differentiated by mutation and the rest resulted from at least one further recombination event. The recombination site between sites 858-859 was the source of recombination in the history of two haplotypes. The other 6 recombination sites were each the source of a single recombination event. A total of 10 recombination events were estimated to have occurred over 7 recombination sites, which equal a ratio of 4.08 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). Three of the seven recombination sites occurred in the last third of the analyzed sequence data near the 3' terminus.

The ARGs for the Perry and Gregory isolates showed a higher level of evolutionary complexity compared to the NC\_12C isolates (data not shown). Both ARGs indicated that many recombination events were occurring throughout the evolution of the haplotypes, instead of a single major ancestral event. For the Perry isolates, 2 of the 41 haplotypes have no recombination in their histories. The haplotype most closely related to the inferred ancestral sequence had eleven mutations differentiating it and was comprised of an isolate collected in 2007. There was no single recombination site that was common to all isolates; however, the recombination site between sites 777-778 was the source of six recombination events. The recombination sites between sites 396-397, 468-469, 496-497 and 858-859 were each the source of three recombination events, with the rest of the sites each occurring only once or twice. There were 44 recombination events estimated to have occurred over 25 sites which equal a ratio of 5.03 recombination events to 2.86 recombination sites when compared

to all other isolates of the cultivars in all gene regions (Table 2). A recombination event occurring at sites 789-790 and 858-859 was present in both the NC\_12C and Perry isolates.

For the Gregory isolates, 5 of the 35 haplotypes have no recombination in their histories. The haplotype most closely related to the inferred ancestral sequence had two mutations differentiating it and was comprised of an isolate collected in 2007. There was no single recombination event that was common to all isolates; however, the recombination site between sites 477-478 and 870-871 were each the source of five recombination events. The recombination site between sites 918-919 was the source of four recombination events. The recombination sites between sites 192-193, 489-490, 822-823, and 894-895 were each the source of three recombination events, with the rest of the sites each occurring only once or twice. There were 59 recombination events estimated to have occurred over 31 sites which equal a ratio of 5.44 recombination events to 2.86 recombination sites when compared to all other isolates of the cultivars in all gene regions (Table 2). A recombination event occurring at sites 489-490 occurred once in the NC\_12C isolates and three times in the Gregory isolates. A recombination event occurring at sites 268-269 occurred once in both the Perry and Gregory isolates. A recombination event occurring at sites 783-784 occurred twice in the Perry isolates and once in the Gregory isolates. A recombination event at sites 870-871 occurred twice in the Perry isolates and five times in the Gregory isolates. Finally, a recombination event occurring at sites 918-919 occurred once in the Perry isolates and four times in the Gregory isolates. For the RdRP gene region of the Perry and Gregory isolates, fourteen sites out of 25 and twelve sites out of 31, respectively, of the recombination sites occurred in the last third of the 3' terminus of the analyzed sequence.

#### **4. Discussion**

The primary objective of this research was to detect, quantify, and reconstruct the history of homologous recombination in three different genes of a sample of TSWV isolates. For this study, isolates sampled from different peanut cultivars with varying levels of host plant resistance were partitioned to minimize confounding factors that may originate from cultivar effects. This experimental approach provided us with a unique opportunity to estimate the number and placement of recombination events within a phylogenetic framework without a larger dataset that would be computationally intensive and prohibitive for currently available computer algorithms. Our secondary objective was to compare the levels of TSWV recombination in two susceptible and one field-resistant cultivar of peanut to identify any correlations between the amount of recombination in the virus and expected level of disease resistance of the cultivar. Our third objective was to estimate the recombination rate population parameter for TSWV in a single cultivar from a single geographic area over multiple years to ensure uniformity of isolate characteristics and reduce the effects of local ecological variables on rate estimates. While recombination in other RNA plant viruses has been examined previously (Bonnet et al., 2005; Bruyere et al., 2000; Hu et al., 2007), this question has not been addressed fully for TSWV.

In this study, three previously well-characterized and physically unlinked regions of the TSWV genome were sequenced and examined for homologous recombination. Phylogenetic analyses also showed that the growth rates and topologies of maximum-

likelihood trees were not the same between genes and the differing haplotype groupings in current analysis suggest different evolutionary histories for each of the genes (data not shown). For these reasons, the gene regions were analyzed separately, as it has been shown that combining datasets with different histories can lead to erroneous phylogenetic inferences (Posada and Crandall, 2002; Wiens, 1998). To detect recombination in the gene regions, we used matrix incompatibility methods, which highlight sites with different evolutionary histories, to find specific sites in the sequence data that provided some evidence that recombination had occurred (Jakobsen et al., 1997; Posada and Crandall, 2001). Although no explicit recombination blocks were identified within any gene region, we were able to distinguish and quantify the incompatible sites in each gene region matrix.

For the first objective, we examined the sequence data to evaluate whether there were regions that exhibited more recombination than others. In the N gene region, all recombination events exhibited by the NC\_12C isolates occurred in a specific region of the sequence near the 3' terminus. While the Perry isolates had a recombination site in common with the NC\_12C isolates, both Perry and Gregory isolates exhibited other recombination sites in different regions of the sequence. However, 36% of all sites for the Perry isolates and 25% of all sites for the Gregory isolates were located in the same 3' terminus region as the NC\_12C sites, which may suggest that this area of the sequence is a common area for recombination to occur. In the NSm gene region of the NC\_12C isolates, 50% of the recombination is occurring in the region specifically located in the first third of the analyzed sequence data near the 5' terminus. In the Perry and Gregory isolates, the recombination sites of the NSm gene region occurred along the entire region; however, 29% of the sites

were located in the same region as the NC\_12C isolates in the first third of the analyzed sequence data near the 5' terminus for both cultivars. This sequence area has also previously been shown to contribute to rapid “resistance-breaking” in commercial tomato crops (Hoffmann et al., 2001) as a direct result of reassortment in the TSWV genome (Qiu et al., 1998; Qiu and Moyer, 1999). In the RdRP gene region, although many recombination events are occurring throughout the sequence of all isolates from each cultivar, 56%, 60%, and 39% of the recombination sites for the NC\_12C, Perry and Gregory isolates, respectively, were located in the last third of the sequence near the 3' terminus. Furthermore, the highest number of shared recombination sites among the three cultivars (7), was found in this gene region, with five of those sites located near the 3' terminus. The differences in amount and location of recombination sites could be influenced by local ecological effects on the TSWV genome, but the trend supports our hypothesis that there are specific regions in each gene where recombination occurs.

For the second objective, we compared the amount of recombination found in isolates collected from three cultivars by using a ratio of the number of recombination events to a common lowest number of recombination sites. The Gregory and Perry isolates had the most recombination occurring in the N gene compared to their respective NSm and RdRP genes. Furthermore, the Gregory isolates had the most recombination occurring in this region of all the other isolates, which would suggest that more recombination occurs in cultivars with higher levels of field resistance. However, in the NSm gene region, the Perry isolates showed the most recombination occurring in this gene region compared to the NSm gene region of both the NC\_12C and Gregory isolates. A plausible explanation for the lowest

amount of recombination in the NSm gene region for the Gregory isolates could be that the particular sequence in this gene region was already successful in overcoming cultivar field resistance. Any further recombination in the gene region may have lost the advantageous sequence, leading to decreased representation of disadvantageous sequences in the population sampled in this field-resistant cultivar. Finally, the RdRP gene region exhibited the most complex reconstructed history for all isolates. All of the ARGs for the RdRP gene region show numerous mutations after recombination events separating the haplotypes of the isolates of each cultivar which suggests that the incompatible sites arising in the RdRP gene region could be a product of older recombination events. Interestingly, this gene region was the proportionately most recombined gene region out of all three gene regions for the NC\_12C isolates. Furthermore, this gene region is also the proportionately most recombined RdRP gene region in the Gregory isolates out of all isolates. This evidence supports that the most recombination is occurring in the most field-resistant cultivar, Gregory.

To further characterize the recombination in each gene region, our third objective was to estimate a separate rate of recombination for each gene region. We used isolates of a single cultivar (NC\_12C) collected over multiple years from a single geographical region (Bertie County, NC) to estimate the population recombination rate for each region. In this study, rates for each of the three gene regions were estimated based on conservative estimates of the neutral mutation rate,  $\mu$ , of  $1 \times 10^{-3}$  substitutions per site per generation (Jenkins et al., 2002). The neutral mutation rate for all types of viruses has been estimated from values of  $1 \times 10^{-3}$  to  $1 \times 10^{-8}$  (Duffy and Holmes, 2008), but an estimate for riboviruses is, on average, 0.76 per genome per replication (Drake and Holland, 1999). The factors that

influence these rates have been attributed to viral polymerase errors, nucleotide substitutions from host cellular enzymes, and recombination (Duffy and Holmes, 2008). Rates of recombination for TSWV have not been previously reported but are necessary for comparing to other plant viruses, such as *Cauliflower mosaic virus*, a dsDNA plant virus which was estimated to have a recombination rate of  $2 \times 10^{-5}$  to  $4 \times 10^{-5}$  per genome per replication (Froissart et al., 2005). Also, estimated rates of recombination frequency per base for *Brome mosaic virus* and *Cucumber mosaic virus*, both ss(+)RNA plant viruses, range from  $8.1 \times 10^{-2}$  to  $18.5 \times 10^{-2}$  and from  $1.0 \times 10^{-4}$  to  $1.0 \times 10^{-8}$ , respectively (García-Arenal et al., 2001). For the NC\_12C subsample of TSWV, our estimates of the recombination rate were an order of magnitude or two lower than the dsDNA virus, but in the same range as the ss(+)RNA viruses. However, the analyzed sequence length and similarity, as well as areas of frequent recombination, may affect this estimate (García-Arenal et al., 2001).

While the estimates of the overall population mean recombination rate and mean recombination rate per site suggest that the NSm gene region would have the highest incidence of recombination for the NC\_12C isolates, the estimated recombination rate per site per generation was lower than the N gene region estimate, which had the highest. This corresponds to the N and NSm gene region observed ratios of recombination events to recombination sites in the NC\_12C isolates. It could be that the previously reported lower selection pressure on the NSm gene region (Kaye *et al.*, unpublished), while still significant, allowed more recombination to take place and for its effects to remain in the population, although recombination may have taken place more slowly over the course of viral replication. The RdRP gene region of the NC\_12C isolates had the lowest estimates of the

overall population mean recombination rate, mean recombination rate per site, and estimated recombination rate per site per generation of the three gene regions, which suggests that most of the haplotype diversity in these isolates is due to recent mutation instead of recombination. This does not correspond to the RdRP gene region observed ratio of recombination events to recombination sites in the NC\_12C isolates. Although the RdRP gene region is under a similar selection pressure as the N gene region (Kaye *et al.*, in preparation), recombination incidence appears to be severely constrained by selection and the configuration of the ARG suggests that recombination is maintained in the RdRP sequence when it does occur. This suggests that there may have been a shift in selection pressure upon this gene region over time and could explain the results of the difference in ratio.

Examination of isolates collected from three different peanut cultivars has led to the detection and quantification of recombination occurring in three genes of TSWV. We have identified a common area in each gene where the majority of the recombination is occurring. These areas corresponded to sequence near the 3' terminus in the N and RdRP genes and to sequence the 5' terminus in the NSm gene. Future studies may also examine entire genomes of multiple strains for a more expanded accounting of mutation and recombination sites and patterns. Also, because it is known that these cultivars exhibit different levels of resistance to TSWV in the field (Hurt *et al.*, 2003), we examined the incidence of recombination in isolates of each cultivar for comparison. The Gregory isolates proportionally had the most recombination occurring in the N and RdRP gene regions out of all isolates of the three cultivars, while the NC\_12C isolates had the most recombination occurring in the NSm gene region. This suggests that the most recombination is occurring in the most field-resistant

cultivar. Future studies may want to include isolates from a range of other cultivars of peanut that exhibit a range of susceptibility to determine whether trends of recombination frequency may occur. Finally, we estimated recombination rates for each of the three gene regions in a single cultivar, NC\_12C. We found that although the recombination in the N gene region may be affected by strong negative selection (Kaye *et al.*, unpublished) that constrains which regions of the sequence recombine, recombination may be occurring in this gene region the most frequently per generation. Recombination in the NSm gene region may also be constrained to certain areas by negative selection, but it may be occurring at the highest overall rate per site. The recombination in the RdRP appears to be the most constrained and the least frequent of all the gene regions for the NC\_12C isolates, but ARG results suggest that the recombined sequence may be retained in the isolate population. This study examined selected regions of the genome based on phylogenetic optimality criterion (Tsompana *et al.*, in review); future studies may want to use an entire gene or RNA sequence to reflect a rate estimate that considers all parts of the genomic regions.

## **5. Conclusions**

Our analyses support the following four conclusions: 1) homologous recombination occurs in each of the three genes of TSWV, 2) there are specific regions for recombination in each gene, 3) the amount of recombination differs in each gene for isolates sampled from each peanut cultivar, and 4) each gene has a unique incidence and rate of recombination in the NC\_12C isolates.

## REFERENCES

- Aylor, D.L., Price, E.W., Carbone, I., 2006. SNAP: Combine and map modules for multilocus population genetic analysis. *Bioinformatics* 22, 1399-1401.
- Best, R.J., 1968. Tomato spotted wilt virus. *Adv. virus res.* 13, 65-145.
- Bonnet, J., Fraile, A., Sacristán, S., Malpica, J.M., García-Arenal, F., 2005. Role of recombination in the evolution of natural populations of cucumber mosaic virus, a tripartite RNA plant virus. *Virology* 332, 359-368.
- Bowden, L.C., Price, E.W., Carbone, I., 2008. SNAP clade and matrix. Department of Plant Pathology, North Carolina State University 2.0.
- Bruyere, A., Wantroba, M., Flasiński, S., Działłott, A., Bujarski, J.J., 2000. Frequent homologous recombination events between molecules of one RNA component in a multipartite RNA virus. *J. virol.* 74, 4214-4219.
- Carbone, I., Jakobek, J.L., Ramirez-Prado, J.H., Horn, B.W., 2007. Recombination, balancing selection and adaptive evolution in the aflatoxin gene cluster of *Aspergillus parasiticus*. *Mol. ecol.* 16, 4401-4417.
- Carbone, I., Yir-Chung, L., Hillman, B.I., Milgroom, M.G., 2004. Recombination and migration of *Cryphonectria hypovirus 1* as inferred from gene genealogies and the coalescent. *Genetics* 166, 1611-1629.
- Chare, E.R., Holmes, E.C., 2006. A phylogenetic survey of recombination frequency in plant RNA viruses. *Arch. virol.* 151, 933-946.
- Drake, J.W., Holland, J.J., 1999. Mutation rates among RNA viruses. *Proc. natl. acad. sci. U.S.A.* 96, 13910-13913.

- Duffy, S., Holmes, E.C., 2008. Phylogenetic evidence for rapid rates of molecular evolution in the single-stranded DNA begomovirus tomato yellow leaf curl virus. *J. virol.* 82, 957-965.
- Froissart, R., Roze, D., Uzest, M., Galibert, L., Blanc, S., Michalakis, Y., 2005. Recombination every day: Abundant recombination in a virus during a single multi-cellular host infection. *PLoS Biol.* 3, 389-395.
- García-Arenal, F., Fraile, A., Malpica, J.M., 2001. Variability and genetic structure of plant virus populations. *Annu.rev.phytopathol.* 39, 157.
- Hey, J., Wakeley, J., 1997. A coalescent estimator of the population recombination rate. *Genetics* 145, 833-846.
- Hoffmann, K., Qiu, W.P., Moyer, J.W., 2001. Overcoming host- and pathogen-mediated resistance in tomato and tobacco maps to the M RNA of *tomato spotted wilt virus*. *Mol. plant microbe interact.* 14, 242-249.
- Hu, J., Fu, H., Lin, C., Su, H., Yeh, H., 2007. Reassortment and concerted evolution in banana bunchy top virus genomes. *J.virol.* 81, 1746-1761.
- Hurt, C., Brandenburg, R., Jordan, D., Shew, B., Isleib, T., Linker, M., et al, 2003. Managing Tomato spotted wilt virus in peanuts in North Carolina and Virginia AG-638.
- Isleib, T.G., Rice, P.W., Bailey, J.E., Mozingo, R.W., Pattee, H.E., 1997. Registration of 'NC\_12C' peanut. *Crop sci.* 37, 1976.
- Isleib, T.G., Rice, P.W., Mozingo, R.W., II, Bailey, J.E., Mozingo, R.W., Pattee, H.E., 2003. Registration of 'Perry' peanut. *Crop sci.* 43, 739-740.
- Isleib, T.G., Rice, R.W., Mozingo, R.W., Mozingo, R.W.I., Pattee, H.E., 1999. Registration of 'Gregory' peanut. *Crop sci.* 39, 1526.

- Jakobsen, I.B., Wilson, S.R., Easteal, S., 1997. The partition matrix: Exploring variable phylogenetic signals along nucleotide sequence alignments. *Mol. biol. evol.* 14, 474-484.
- Jeger, M.J., Seal, S.E., Van den Bosch, F., 2006. Evolutionary epidemiology of plant virus disease. *Adv. virus res.* 67, 163-2003.
- Jenkins, G.M., Rambaut, A., Pybus, O.G., Holmes, E.C., 2002. Rates of molecular evolution in RNA viruses: A quantitative phylogenetic analysis. *J. mol. evol.* 54, 156-165.
- Kaye, A.C., Abad, J., Shew, B.B., Cubeta, M.A., Moyer, J.W., 2006. Exploring the role of *tomato spotted wilt virus* and soilborne fungi in yellowing syndrome of peanut. *Phytopathology* 96.
- Kaye, A.C., Kennedy, G.G., Shew, B.B., Cubeta, M.A., Moyer, J.W., 2008. Population genetic data analysis of *tomato spotted wilt virus* on peanut in north carolina and virginia. *Phytopathology* 98.
- Lyngsø, R.B., Song, Y.S., Hein, J., 2005. Minimum recombination histories by branch and bound. *Proceedings of the 5th international workshop on algorithms in bioinformatics*, 239-250.
- Moury, B., Desbiez, C., Jacquemond, M., Lecoq, H., 2006. Genetic diversity of plant virus populations: Towards hypothesis testing in molecular epidemiology. *Adv. virus res.* 67, 49-87.
- Moyer, J.W., 2000. Encyclopedia of microbiology. In: Lederberg, J. (Ed.), *Tospoviruses*. Academic Press, San Diego, pp. 592-597.
- Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc.natl.acad.sci.U.S.A.* 98, 13757-13762.
- Posada, D., Crandall, K.A., 2002. The effect of recombination on the accuracy of phylogeny estimation. *J.mol.evol.* 54, 396-402.

- Posada, D., Crandall, K.A., Holmes, E.C., 2002. Recombination in evolutionary genomics. *Annu.rev.genet.* 36, 75-97.
- Price, E.W., Carbone, I., 2005. SNAP: Workbench management tool for evolutionary population genetic analysis. *Bioinformatics* 21, 402-404.
- Qiu, W.P., Geske, S.M., Hickey, C.M., Moyer, J.W., 1998. Tomato spotted wilt tospovirus genome reassortment and genome segment-specific adaptation. *Virology* 244, 186-194.
- Qiu, W.P., Moyer, J.W., 1999. Tomato spotted wilt tospovirus adapts to the TSWV N gene-derived resistance by genome reassortment. *Phytopathology* 89, 575-582.
- Salzman, R.A., Hasegawa, P.M., Bressan, R.A., Fujita, T., Zhu-Salzman, K., 1999. An improved RNA isolation method for plant tissues containing high levels of phenolic compounds or carbohydrates. *Plant mol. biol. rep.* 17, 11-17.
- Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nuc. acids res.* 22, 4673-4680.
- Tsompana, M., 2004. Molecular evolution and population genetics of Tomato spotted wilt virus (TSWV). DISSERTATION (North Carolina State University), 1-189.
- Van den Hurk, J., Tas, P.W.L., Peters, D., 1977. The ribonucleic acid of tomato spotted wilt virus. *J. gen. virol.* 36, 81-91.
- Wiens, J.J., 1998. Combining data sets with different phylogenetic histories. *Syst.biol.* 47, 568-581.

**Table 1.** Year of collection, cultivar, isolate number and accession numbers of amplified nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) fragments of 97 isolates of *Tomato spotted wilt virus*.

Year collected	Cultivar	Isolate	AN N <sup>a</sup>	AN Nsm <sup>b</sup>	AN RdRP <sup>c</sup>
2005	NC_12C				
		L1N5	FJ234541	FJ234640	FJ262056
		L2N5	FJ234482	FJ234709	FJ262002
		L3N5	FJ234461	FJ234690	FJ262077
2006					
	NC_12C				
		L1N6	FJ234562	FJ234714	FJ262074
		L2N6	FJ234487	FJ234744	FJ262076
		L3N6	FJ234538	FJ234669	FJ261998
		L4N6	FJ234485	FJ234641	FJ262013
		L5N6	FJ234532	FJ234671	FJ261991
		L6N6	FJ234531	FJ234668	FJ262014
		L7N6	FJ234577	FJ234693	FJ262070
		L8N6	FJ234451	FJ234773	FJ262154
		L10N6	FJ234537	FJ234739	FJ262149
		L15N6	FJ234493	FJ234817	FJ262005
		L16N6	FJ234465	FJ234772	FJ262153
		L18N6	FJ234474	FJ234713	FJ262025
		L19N6	FJ234459	FJ234682	FJ262084
		L20N6	FJ234466	FJ234819	FJ262042
		L23N6	FJ234529	FJ234678	FJ262083
	Gregory				
		F1G6	FJ234581	FJ234735	FJ262138
		F2G6	FJ234582	FJ234737	FJ262139
		F5G6	FJ234527	FJ234750	FJ262092
		F6G6	FJ234525	FJ234751	FJ262091
		F7G6	FJ234526	FJ234752	FJ262090
	Perry				
		F2P6	FJ234607	FJ234768	FJ262096
		F3P6	FJ234560	FJ234774	FJ262152
		F4P6	FJ234551	FJ234829	FJ262051
		F7P6	FJ234476	FJ234659	FJ262028
		F10P6	FJ234519	FJ234828	FJ262017
		F11P6	FJ234518	FJ234827	FJ262016
		F13P6	FJ234606	FJ234770	FJ262097
		F15P6	FJ234569	FJ234683	FJ262062
2007					
	NC_12C				
		L1N7	FJ234464	FJ234746	FJ262089
		L2N7	FJ234471	FJ234748	FJ262003

**Table 1. Continued**

		L3N7	FJ234601	FJ234824	FJ262162
		L4N7	FJ234467	FJ234699	FJ262068
		L5N7	FJ234559	FJ234711	FJ262079
		L6N7	FJ234453	FJ234649	FJ262088
		L8N7	FJ234578	FJ234647	FJ262032
		L9N7	FJ234624	FJ234776	FJ262135
	Perry				
		L1P7	FJ234634	FJ234762	FJ262099
		L2P7	FJ234516	FJ234740	FJ262148
		L3P7	FJ234574	FJ234661	FJ262050
		L4P7	FJ234609	FJ234790	FJ262114
		L5P7	FJ234506	FJ234688	FJ262170
		L6P7	FJ234548	FJ234813	FJ262027
		L7P7	FJ234534	FJ234670	FJ261993
		L8P7	FJ234490	FJ234708	FJ262069
		L9P7	FJ234509	FJ234652	FJ262048
		L10P7	FJ234530	FJ234657	FJ262082
		L11P7	FJ234499	FJ234766	FJ262054
		L12P7	FJ234543	FJ234653	FJ262031
		L13P7	FJ234572	FJ234732	FJ262066
		L14P7	FJ234580	FJ234777	FJ262058
		L15P7	FJ234452	FJ234686	FJ262040
		L16P7	FJ234635	FJ234733	FJ262101
		L17P7	FJ234515	FJ234663	FJ262023
		L18P7	FJ234483	FJ234715	FJ262075
		L19P7	FJ234600	FJ234734	FJ262163
		L20P7	FJ234468	FJ234717	FJ262004
		L21P7	FJ234595	FJ234694	FJ262165
		L22P7	FJ234523	FJ234665	FJ262085
		L23P7	FJ234539	FJ234662	FJ261989
		L24P7	FJ234633	FJ234745	FJ262106
		L25P7	FJ234504	FJ234720	FJ262142
		L26P7	FJ234565	FJ234722	FJ262161
		L27P7	FJ234462	FJ234725	FJ262124
		L28P7	FJ234561	FJ234821	FJ262006
		L29P7	FJ234546	FJ234644	FJ262057
		L30P7	FJ234455	FJ234726	FJ262157
		L31P7	FJ234637	FJ234783	FJ262100
		L32P7	FJ234517	FJ234814	FJ262036
		L33P7	FJ234591	FJ234825	FJ262008
		L34P7	FJ234553	FJ234684	FJ262087
		L35P7	FJ234638	FJ234723	FJ262111
		L36P7	FJ234457	FJ234685	FJ262046

**Table 1. Continued**

	Gregory				
		M1G7	FJ234590	FJ234710	FJ262102
		M2G7	FJ234507	FJ234701	FJ262117
		M3G7	FJ234588	FJ234704	FJ262141
		M4G7	FJ234557	FJ234795	FJ262144
		M5G7	FJ234613	FJ234755	FJ262131
		M6G7	FJ234505	FJ234721	FJ262063
		M7G7	FJ234592	FJ234789	FJ262049
		M8G7	FJ234627	FJ234803	FJ262118
		M9G7	FJ234564	FJ234676	FJ262134
		M10G7	FJ234585	FJ234765	FJ262119
		M11G7	FJ234563	FJ234718	FJ262160
		M12G7	FJ234579	FJ234800	FJ262110
		M13G7	FJ234626	FJ234796	FJ262125
		M14G7	FJ234632	FJ234757	FJ262093
		M15G7	FJ234583	FJ234764	FJ262120
		M16G7	FJ234552	FJ234691	FJ262107
		M17G7	FJ234614	FJ234792	FJ262109
		M18G7	FJ234619	FJ234760	FJ262105
		M19G7	FJ234586	FJ234712	FJ262021
		M20G7	FJ234617	FJ234778	FJ262086
		M21G7	FJ234478	FJ234786	FJ262094
		M22G7	FJ234602	FJ234782	FJ261997
		M23G7	FJ234571	FJ234729	FJ262167
		M24G7	FJ234587	FJ234780	FJ262127
		M25G7	FJ234636	FJ234788	FJ262121
		M26G7	FJ234568	FJ234823	FJ262169
		M27G7	FJ234629	FJ234805	FJ262122
		M28G7	FJ234625	FJ234793	FJ262123
		M29G7	FJ234494	FJ234677	FJ262159
		M30G7	FJ234622	FJ234798	FJ262104
		M31G7	FJ234610	FJ234791	FJ262115
		M32G7	FJ234620	FJ234759	FJ262113

<sup>a</sup> Accession numbers for the nucleocapsid (N) gene; <sup>b</sup> Accession numbers for the movement (NSm) gene; <sup>c</sup> Accession numbers for the RNA-dependent RNA polymerase (RdRP) gene

**Table 2.** The ratios of the number of inferred recombination events to the number of possible recombination sites<sup>a</sup> for the nucleocapsid (N), movement (NSm), and RNA-dependent RNA polymerase (RdRP) fragments of NC\_12C, Perry, and Gregory isolates of *Tomato spotted wilt virus*.

Gene	NC_12C	Perry	Gregory
N	3.81	5.10	6.97
NSm	3.57	4.70	3.27
RdRP	4.08	5.03	5.44

<sup>a</sup> The numbers displayed in the table correspond to the number of inferred recombination events. All are against the lowest common denominator of 2.86 recombination sites

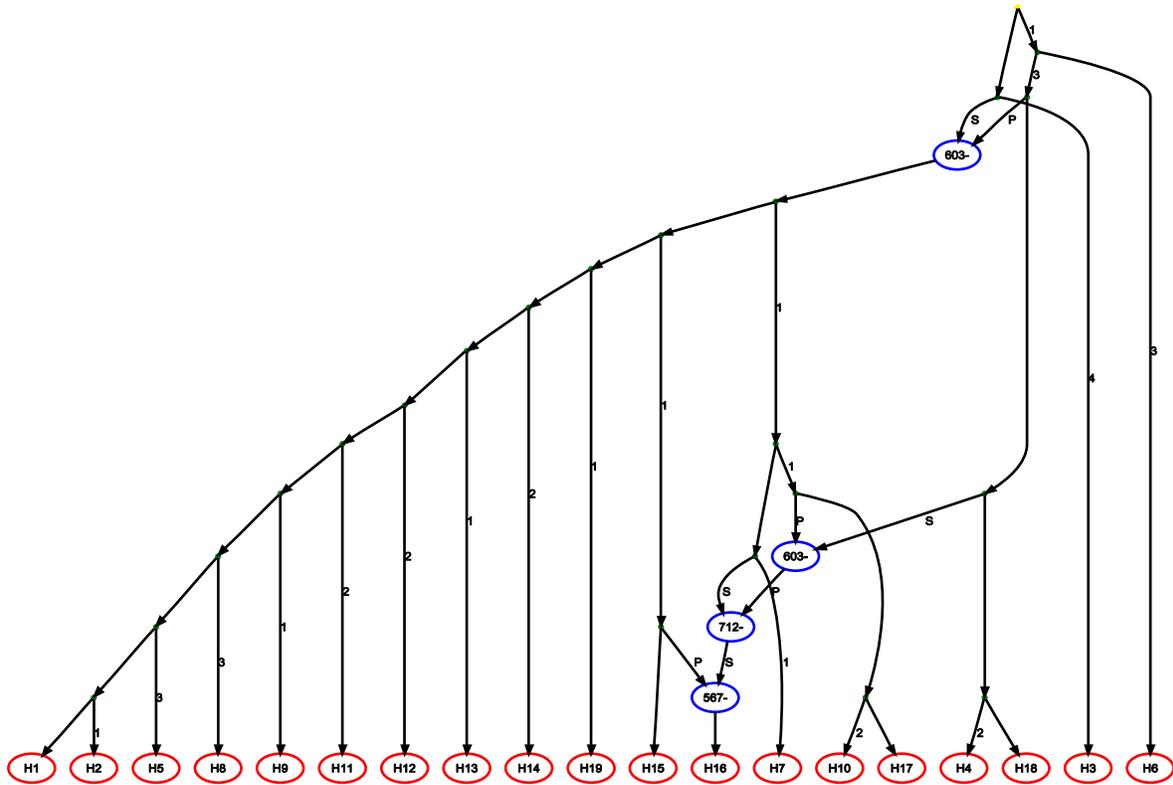


Figure 1. Ancestral recombination graph (ARG) for the N gene region. The yellow dot at the top represents the inferred ancestral sequence for the sample isolates. The numbers along the branches indicate the number of mutational events that occurred in the formation of each haplotype. The ‘S’ and ‘P’ correspond to the suffix and prefix of each recombination block that form the subsequent recombination event in the blue oval. The numbers in the blue ovals represent the site where the recombination break occurred in the sequence. The red ovals indicate the relative positions of each haplotype in the inferred history.

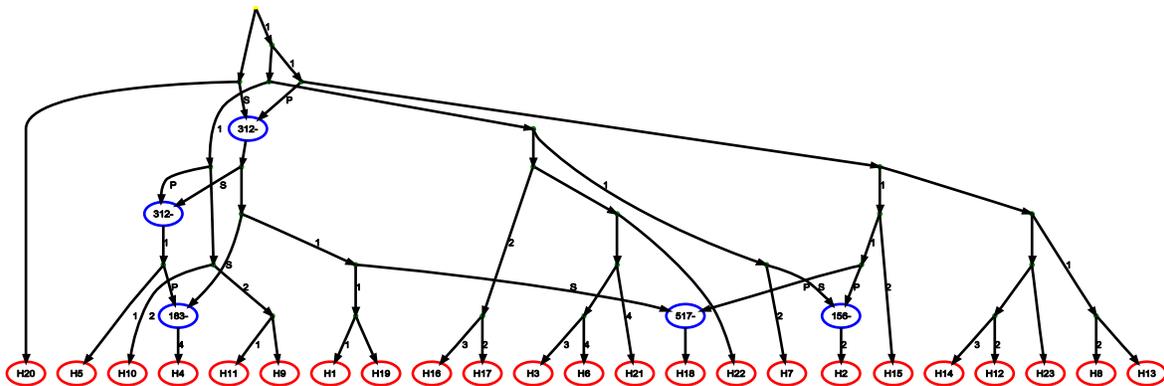


Figure 2. Ancestral recombination graph (ARG) for the NSm gene region of the NC\_12C isolates. The yellow dot at the top represents the inferred ancestral sequence for the sample isolates. The numbers along the branches indicate the number of mutational events that occurred in the formation of each haplotype. The ‘S’ and ‘P’ correspond to the suffix and prefix of each recombination block that form the subsequent recombination event in the blue oval. The numbers in the blue ovals represent the site where the recombination break occurred in the sequence. The red ovals indicate the relative positions of each haplotype in the inferred history.

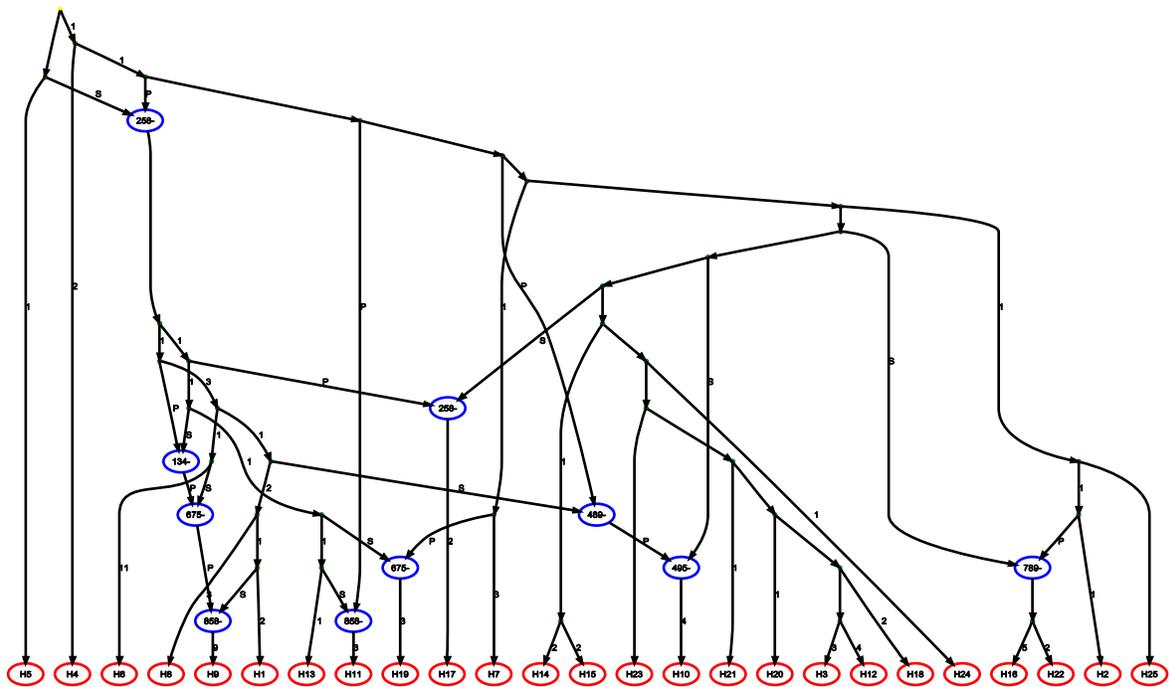


Figure 3. Ancestral recombination graph (ARG) for the RdRP gene region of the NC12C isolates. The yellow dot at the top represents the inferred ancestral sequence for the sample isolates. The numbers along the branches indicate the number of mutational events that occurred in the formation of each haplotype. The ‘S’ and ‘P’ correspond to the suffix and prefix of each recombination block that form the subsequent recombination event in the blue oval. The numbers in the blue ovals represent the site where the recombination break occurred in the sequence. The red ovals indicate the relative positions of each haplotype in the inferred history.