Evolution and structure of Tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes

Diana Tentchev, Eric Verdin, Cécile Marchal, Monique Jacquet, Juan M. Aguilar, Benoît Moury

Evolution and structure of *Tomato spotted wilt virus* populations: evidence of extensive reassortment and insights into emergence processes

Diana Tentchev[1,2,3], Eric Verdin[1], Cécile Marchal[2], Monique Jacquet[3], Juan M. Aguilar[4] and Benoît Moury[1].
[1] *INRA, UR407 Pathologie Végétale, 84140 Montfavet, France.*
[2] *Clause, 13210 St Rémy de Provence, France*
[3] *Vilmorin, 30210 Lédenon, France*
[4] *HM-Clause, Paraje La Reserva s/n, 04745 La Mojonera, Almería, Spain*

Running title : TSWV evolution and emergence processes

Correspondence: Benoît Moury
Address : INRA., Unité de Pathologie Végétale, BP94, F-84143 Montfavet Cedex, France.
Tel: +33(0)4 32 72 28 16. Fax : +33(0)4 32 72 28 42.
E-mail : Benoit.Moury@avignon.inra.fr

The GenBank/EMBL/DDBJ accession numbers for the sequences reported in this paper are FR692373 to FR693268.

Summary
Tomato spotted wilt virus (TSWV, genus *Tospovirus*, family Bunyavirideae) genetic diversity was evaluated by sequencing parts of the three RNA genome segments of 224 isolates, mostly from pepper and tomato crops in southern Europe. Eighty-three percent of the isolates showed consistent clustering into three clades, corresponding to their geographic origin, Spain, France or the USA, for the three RNA segments. In contrast, the remaining 17% of isolates did not belong to the same clade for the three RNA segments and were shown to be reassortants. Among them, eight different reassortment patterns were observed. Further phylogenetic analyses provided insights into the dynamic processes of the worldwide resurgence of TSWV that, since the 1980s, has followed the worldwide dispersal of the western flower thrips (*Frankliniella occidentalis*) tospovirus vector. For two clades composed essentially of Old World (OW) isolates, tree topology suggested a local re-emergence of indigenous TSWV populations following *F. occidentalis* introductions, while it could not be excluded that the ancestors of two other OW clades were introduced from northern America contemporarily with *F. occidentalis*. Finally, estimation of the selection intensity that has affected the evolution of the NSs and nucleocapsid proteins encoded by RNA S of TSWV suggests that the former could be involved in breakdown of the resistance conferred by the *Tsw* gene in pepper.

Keywords: *Capsicum annuum*, pepper, *Solanum lycopersicum*, tomato, Solanaceae, *Tospovirus*, *Tomato spotted wilt virus*, TSWV, reassortment, recombination, virus emergence, resistance breakdown

Introduction

Tomato spotted wilt virus (TSWV) is the type member of the genus *Tospovirus*, the only plant virus group within the family Bunyaviridae. After its first descriptions in the beginning of the 20th century (Brittlebank, 1919; Samuel *et al.*, 1930), TSWV decreased in prevalence between the 1940s and the 1980s. A worldwide resurgence of TSWV occurred since 1980 in northern America, 1987 in western Europe and 1993 in Australia, among other examples (Allen & Broadbent, 1986; Greenough *et al.*, 1985; Dietzgen *et al.*, 2005; Kirk & Terry, 2003). The worldwide spread of the western flower thrips (*Frankliniella occidentalis*), an efficient TSWV vector, from the western part of the USA during the 1980s has certainly played an important role in TSWV emergence (Kirk & Terry, 2003).
Yet, another kind of TSWV emergence has been observed more recently, as the virus was shown to adapt rapidly to plant resistances. For example, the *Tsw* resistance gene was broken down only a few years after its deployment in pepper crops (Roggero *et al.*, 2002; García-Arenal & McDonald, 2003; Thomas-Carroll & Jones 2003; Margaria *et al.*, 2004; Sharman & Persley, 2006). The genome of TSWV consists of three single-stranded RNA segments referred to as L, M and S (large, medium and small, respectively) (Whitfield *et al.*, 2005). This genome arrangement allows the exchange of entire genome segments between variants coinfecting a single plant, a process known as 'reassortment' to distinguish it from recombination, *i.e.* the exchange of parts of genome segments during replication. Experimental selection of reassortants from plants coinfected by two TSWV isolates with contrasted biological properties allowed mapping the determinant for breakdown of the *Tsw* resistance to RNA S (Jahn *et al.* 2000; Margaria *et al.*, 2007). Further studies have suggested contradictorily that the NSs non-structural protein (Margaria *et al.*, 2007) or alternatively the nucleocapsid protein (NP) (Lovato *et al.*, 2008) coding regions of RNA S were involved in breakdown.
In this article, we present an analysis of the molecular diversity of TSWV in southern Europe to get deeper insights into the mechanisms of these two kinds of TSWV emergence.

Results

Genetic diversity of sampled TSWV isolates

Sequences of four genome regions, covering parts of the three RNA segments, were determined for 224 TSWV isolates collected mainly in Spain (Fig. 1; Supplementary Table S1) but also in France, Italy, Algeria and the USA (Supplementary Table S2). The sequences are available in GenBank (accession numbers FR692373 to FR692596 for RNA L, FR692597 to FR692820 for RNA M, FR692821 to FR693044 for the NSs gene and FR693045 to FR693268 for the NP gene). No gaps were observed in any of the four regions after alignment with ClustalW. The four alignments were concatenated in a single alignment to search for recombination (or reassortment) events. We found 38 sequences (17% of isolates) with a significant recombination signal both using RDP (Table 1) and classical phylogenetic studies (Fig. 2). For these 38 isolates, no recombination signal was detected within each of the four sequenced regions or between the NSs and NP genes corresponding to RNA S, which suggests that these isolates are reassortants but not true recombinants. As recombinant and parental groups could be misidentified in recombination analyses, we used the consensus score implemented in RDP v. 3.44 for this purpose. A consensus score above 60% identified 14 isolates as being reassortant rather than parental with strong confidence (Table 1). The other detected isolates showed a medium consensus score between 43 and 60%, indicating a fair likelihood of being reassortants.
Maximum likelihood and neighbour joining phylogenetic methods allowed distinguishing five, four and three major clades, respectively, for RNAs S, L and M (Fig. 2). For each sequence alignment, the 186 non-reassortant TSWV isolates clustered in three clades. The first group of non-reassortant isolates corresponded to clade 1 for RNA L, to clade *a* for RNA M and to clade α for RNA S and contained Spanish isolates and the single Algerian isolate (Fig. 2). The second group of non-reassortant isolates (corresponding to clades 2, *b* and β for the three RNAs) contained the three isolates from the USA and the third group (corresponding to clades 3, *c*, and γ for the three RNAs) comprised French isolates and an Italian isolate (Fig. 2). For the three RNAs, these three clades of non-reassortant

isolates were consistent, with bootstrap values varying from 75% to 100% (data not shown), while addition of the reassortant isolates decreased slightly the bootstrap values associated to some of them, especially clade 1 of RNA L (Fig. 2).

Depending on the genome region examined, the 38 putative reassortant isolates clustered with one or another of these major clades, confirming their reassortant nature (Fig. 2; Table 1). Considering together the tree topologies obtained for the three genome segments, eight different groups of reassortants could be distinguished (Table 1; Fig. 2). For RNA L, reassortant groups #3, #4 and #5 belonged to clade 4, which was very distantly related to the other clades defined for this RNA. In contrast, for the other two RNAs, isolates corresponding to this distant clade were genetically close to some non-reassortant isolates (Fig. 2). This particular clade of RNA L did not contain any non-reassortant isolates, suggesting that reassortant groups #3, #4 and #5 were the results of reassortment events with a parental group not represented in our dataset. This distant clade decreased the confidence level associated to the topology of RNA L tree, as attested by the rather low bootstrap value obtained for clade 1 (58%; Fig. 2). Excluding these 19 distant sequences provided a much more robust tree (bootstrap values >83% for the three clades; data not shown). Concerning RNA M, the reassortant TSWV isolates were included into, or closely related to non-reassortant isolates of clades *a*, *b* or *c* (Fig. 2). Finally, concerning RNA S, the reassortants were also included into, or closely related to non-reassortant isolates of either clade α or γ, except one group of isolates. This group included reassortant groups #1 and #5, which built together clade δ and clustered with the two clades γ and ε with a rather low bootstrap value (58%; Fig. 2). Nonetheless, their clustering with different groups of non-reassortant isolates for the other two RNAs (L and M), proves their reassortant nature (Fig. 2; Table 1). Among the eight reassortant groups, three were from the Spanish province of Almería, three were from France and two were from Italy or the closeby French department of Alpes Maritimes (Table 1). No reassortant isolate was observed in the Spanish province of Murcia, despite extensive sampling.

To root the TSWV trees, several outgroups were used. *Tomato chlorotic spot virus* (TCSV), the closest species related to TSWV, was used to root the trees corresponding to RNA M and to the NP gene. For these two genome regions, the root of the TSWV tree was consistently found between clades *c* or γ+ε of French and Italian isolates and the other clades (Fig. 2). *Groundnut ringspot virus* (GRSV), which is more distant, provided similar results (data not shown). No TCSV or GRSV sequences are available for the L RNA or for the NSs gene. *Impatiens necrotic spot virus*, another tospovirus, was found to be too distant from TSWV to estimate reliably the location of the root of the trees (data not shown).

We compared our TSWV isolates with additional sequences available in databanks. Most TSWV sequences in GenBank were from RNA S, especially the NP gene, while only four sequences of RNA L were available and RNA M showed an intermediate situation. Only results obtained with the NP gene, which comprised the largest number of sequences, are presented (Fig. 3). Similar but less extensive results were obtained with the other genome regions and there was no evidence of recombination or reassortment after analyzing these additional genome regions (data not shown). The sequences of the NP gene belonged to six major clades, five of which comprised some of our isolates and corresponded to the clades revealed by the previous analysis of the NSs and NP genes (Fig. 2 and 3). The sixth clade contained three Brazilian isolates. The clade composed of clade β and additional isolates from the USA is probably polyphyletic since the bootstrap value associated to this group was low (Fig. 3).

Structure of pepper TSWV populations in relation to geographical origin and host genotype

To get deeper insight into the dispersal processes of TSWV, we examined the genetic structure of the virus populations in six administrative communes of the provinces of Murcia and Almería sampled quite extensively. Genetic differentiation between these TSWV populations was estimated by pairwise F_{ST}. Twelve of 15 pairwise F_{ST} comparisons revealed a significant (P<0.05) differentiation, including the eight pairwise F_{ST} comparisons between populations of Murcia and Almería provinces (Table 2). Differenciation of TSWV populations was significant even for populations distant of less than 35 kilometres (four of seven pairwise F_{ST} comparisons).

Analyses of the molecular variance (AMOVAs) were conducted to evaluate the contribution of various factors to the genetic differentiation of the TSWV populations. The total variation observed among sequences was shared (i) among the different groups considered (province; host plants carrying or not the *Tsw* gene), (ii) between communes within the same group and (iii) within the communes. The contribution of the province level to the total variation was only marginally significant (24.2% of total variation; *P*-value=0.08) (Table 3A). 7.3% of the variation was contributed by the commune level within provinces (*P*-value<0.001), while the largest part of the variance (68.5%) was at the within-commune level (Table 3A). There was no effect of the presence of the *Tsw* resistance gene in the pepper plant on the TSWV genetic diversity (Table 3B) but the strong genetic differentiation contributed by the communes could have hampered the detection of an effect of the plant genotype. Consequently, we performed also an AMOVA with a partition of the TSWV genetic variation (i) between the different communes, (ii) between the host plants carrying or not the *Tsw* resistance gene within the same commune and (iii) within each group of plants in each commune. Again, the AMOVA did not reveal any effect of the resistance of the host plant (Table 3C).

Analysis of positive selection in the NSs and NP genes of TSWV

Since independent studies suggested alternatively that the NSs or the NP genes were involved in the breakdown of the *Tsw* resistance in pepper (Margaria *et al.*, 2007; Lovato *et al.*, 2008), we analysed the patterns of selection in these two genes with the aim to identify the true breakdown determinant. For this, we used only sequences of Spanish isolates that were collected in the provinces of Murcia and Almería. Indeed, these isolates were exposed to similar environmental conditions and their NSs and NP genes share a common evolutionary history since all of them belong to clade α (Fig. 2; Table 1). Consequently, the selection patterns observed in sequences of these isolates are more likely attributable to factors differentiating the isolates and, notably, the presence or absence of the *Tsw* resistance gene in the sampled plants. We first compared the overall selection intensity in the NSs and NP genes. The distribution of the dN/dS ratios estimated from pairwise comparisons between sequences showed that the selection intensity was significantly higher in the NSs than in the NP gene (Fig. 4; average pairwise dN/dS were 0.288 and 0.042 for the NSs and the NP genes, respectively). Then, using four different methods implemented in the HyPhy sofware, we detected eleven codon positions in the NSs gene and one codon position in the NP gene potentially undergoing positive selection (Table 4).

Discussion

The emergence dynamics of TSWV

Severe outbreaks of TSWV have been reported worldwide since 1980. They were preceded by a rapid and worldwide expansion of the thrips *F. occidentalis*, native to western North America, since the late 1970s (Kirk & Terry, 2003). Owing to these coincidental events and given the fact that *F. occidentalis* was shown to be a more efficient vector of TSWV than other thrips species indigenous to the Old World (Wijkamp *et al.*, 1995), *F. occidentalis* was invoked as a major cause of the resurgence of TSWV. Moreover, TSWV is transmitted in a persistent manner and multiplies in its thrips vectors (Ullman *et al.*, 1993; Wijkamp *et al.*, 1993), two properties favourable to long-distance TSWV dispersal. Considering these data, two scenarios for TSWV resurgence can be drawn: (i) Both *F. occidentalis* and TSWV populations from western North America have undergone a worldwide spread, and *F. occidentalis* allowed subsequent local outbreaks of these TSWV populations ('emergence' scenario) or (ii) the worldwide spread of *F. occidentalis* favoured local outbreaks of indigenous TSWV populations ('re-emergence' scenario). A mixture of these two scenarios is also possible. According to the emergence scenario, one would expect (i) that northern American TSWV populations had undergone an earlier diversification than populations resulting from worldwide outbreaks and (ii) a higher genetic richness of northern American populations compared to others due to bottlenecks during their worldwide spread. According to the re-emergence scenario, one would expect neither an earlier diversification of northern American TSWV isolates, nor a higher TSWV richness in northern

America than in other parts of the world. Our phylogenetic analyses allow assessing the relevance of these two scenarios.

Based on the position of the root of the tree (Fig. 2 and 3), the two clades γ and ε (Fig. 2 and 3) clearly diverged from the other TSWV isolates (node "*x*" in Figs. 2 and 3) before the diversification of northern American isolates of clade β and related isolates (node "*z*"). Consequently, we exclude their introduction from northern American TSWV populations and, instead, favour a scenario of local re-emergence. These two clades γ and ε may have been affected by more recent trans-continental dispersal. For example, the two American isolates that cluster with clade γ (accession numbers AY744478 and AY744476; Fig. 3) diverged recently in the tree and were presumably the result of recent introductions from the Old World. Also, the two French isolates France81 and SO46 (clade ε) could have been introduced from Asia (Fig. 3). The tree topology also suggests that the group of Brazilian isolates diverged (node "*y*" in Fig. 3) before the diversification of northern American isolates (node "*z*") and fits also with a re-emergence scenario for this group (Fig. 3).

The situation is less clear for clades α and δ. Because of the low bootstrap values associated to the internal nodes that link these clades and the group of northern American isolates in the phylogenetic tree (Fig. 2 and 3), we cannot establish reliably which of these clades get diversified earlier. The low bootstrap values and short internal branches of the group of northern American isolates suggest that it could be polyphyletic and therefore more diversified than clades α and δ. In conclusion, these data are not incompatible with a scenario where ancestors of clades α and δ were introduced from America and undergone a foundation effect, which led to their rapid diversification. This is also suggested by the rather long internal branches that link these two clades to the northern American group of isolates (Fig. 3).

Frequent reassortment during TSWV evolution

Until now, no evidence of reassortment or recombination had been obtained for natural TSWV isolates (Tsompana *et al.*, 2005), even though laboratory experiments have shown a great potential of TSWV to reassort in plants infected by mixtures of TSWV isolates (Qiu *et al.*, 1998; Qiu & Moyer, 1999; Jahn *et al.*, 2000; Hoffmann *et al.*, 2001; Margaria *et al.*, 2007). Detection of reassortment in natural TSWV populations may have been hampered by the lack of sequence data, especially for RNA L. Most studies of plant or animal viruses with a segmented genome showed that reassortment occurs at rather low rates in viruses of the family Bunyaviridae (Henderson *et al.*, 1995; Nemirov *et al.*, 1999) or of other families (Fraile *et al.*, 1997; Lin *et al.*, 2004; Roossinck, 2002; Bonnet *et al.*, 2005; Miranda *et al.*, 2000; Iturriza-Gomara *et al.*, 2001; Watanabe *et al.*, 2001; McDonald *et al.*, 2009), which could be due to strong selection acting against reassortants (Brown *et al.*, 2002; Fraile *et al.*, 1997; Escriu *et al.*, 2007). Some viral taxa seem however to show more extensive reassortment (Silander *et al.*, 2005; Lindstrom *et al.*, 2004; Nelson *et al.*, 2008; Khiabanian *et al.*, 2009).

We found both a high diversity and a relatively high frequency of reassortant TSWV isolates in Spain, France and Italy. In Spain, 25 of 191 (13%) isolates were reassortants and belonged to three different groups; in France, reassortants represented 5 of 20 (25%) isolates and belonged to four different groups, while in Italy reassortants represented 8 of 9 (89%) isolates and belonged to two different groups (Table 1). An alternative interpretation of our results could be that these virus samples were composed of mixtures of non-reassortant virus strains and that RT-PCR amplified preferentially the genome of one or the other strain, depending on the RNA segment. We consider this alternative view as extremely unlikely for the following two reasons. First, we found no evidence of plants being infected by mixtures of several TSWV strains when checking the sequence chromatograms, whatever the RNA segment. This is also suggested by the extreme genetic similarity of independent samples collected from the same plant. In almost all cases, leaf and fruit samples collected on the same plant differed by one or two nucleotides only among a total of 3525 nucleotides sequenced. Consequently, under this alternative assumption, one should imagine that widespread and drastic RT-PCR biases have affected not only one, but simultaneously the three RNA segments for all the detected reassortant isolates in Table 1. Also, since for at least two reassortant groups (#3 and #5), the three RNAs cluster with three different groups, this alternative assumption would imply that the sampled plants were

infected by a mixture of three non-reassortant strains and that RT-PCR biases would have allowed amplification of only one of them for each RNA segment.
Second, for each of the countries where reassortants have been detected (*i.e.* Spain, France and Italy), only one non-reassortant TSWV clade has been found. Then, the alternative assumption would imply that the putative additional non-reassortant virus groups occurred systematically in co-infection with strains from other non-reassortant groups and were never detected singly within plants. Consequently, from three to five non-reassortant groups would have been missed as in single infections: one corresponding to clade 4, one corresponding to clade ε and from one to three corresponding to the different subgroups of detected reassortants in clade b.
The fact that only one non-reassortant TSWV group was observed in each country suggests either (i) that reassortment occurred in pepper or tomato plants in these respective countries but some of the non-reassortant parental TSWV groups which contributed to the generation of the reassortants became extinct or (ii) that these reassortants were introduced from other geographical areas or from other host plants. The fact that some reassortants show high similarity with isolates from distant geographical origins supports the second scenario. Strongest evidence comes from reassortant group #5, from Italy, which appears to have undergone at least two reassortment events since its three RNAs cluster with three different non-reassortant groups (Table 1). For each RNA of these Italian isolates, sequences from Japanese isolates are most similar (accession number AB198742 for RNA L and accession number AB277581 for the NP gene) or among the most similar (accession number AB010996 for RNA M) (Fig. 3 and data not shown). RNAs M and S of isolates from another Italian reassortant group, group #1, also show high similarity with Japanese and Korean isolates (Fig. 3 and data not shown). Finally, RNA S of French isolates from reassortant group #7 are most closely related to a group of South Korean isolates (Fig. 3). These data suggest genetic exchanges between Asiatic and European TSWV populations. Though only few sequences of Asiatic isolates are available, none of them possess RNA segments that belong to the three non-reassortant groups shown in our study. Consequently, this suggests that Asiatic isolates may have been introduced into (or may have migrated to) Europe rather than the reverse. If this scenario is correct, the observed genetic diversity in Spanish TSWV populations indicates at least two independent introductions, corresponding to groups #2 and #4 of reassortants. In contrast, group #3 may be the result of subsequent local reassortment by exchange of RNA L or M between isolates from groups #2 and #4 (Table 1).

Insights into the low durability of the* Tsw *resistance in pepper

The *Tsw* gene used to protect pepper crops from TSWV is emblematic of the 'boom and bust' cycle of resistance genes, where resistant cultivars are first deployed on large geographical scales (the boom part of the cycle) which is followed by resistance breakdowns (the bust part). The first breakdowns of the *Tsw* resistance were observed only one year after their initial deployment in 1999 in Italy and Spain (García-Arenal & McDonald, 2003). Several factors could be responsible for such low resistance durability. The durability of plant resistances to viruses was shown to be linked to (i) the capacity of the virus to evolve resistance breaking properties, which is linked to the number of mutations required for such properties (Harrison, 2002; Ayme *et al.*, 2006, 2007; Fabre *et al.*, 2009), and to (ii) the impact of these mutations on the virus competitiveness (Jenner *et al.*, 2002; Desbiez *et al.*, 2003; Fabre *et al.*, 2009; Janzac *et al.*, 2010). Concerning TSWV, the trees of the NP and NSs genes (Fig. 2 and 3), which correspond to the RNA segment which carries the factor responsible for breakdown of the *Tsw* resistance, show that multiple acquisitions of the resistance breakdown properties have affected TSWV history. The NP gene of isolates collected on pepper plants carrying *Tsw* corresponded to four of the six TSWV clades, *i.e.* clades α, γ, δ and ε (Fig. 2 and 3; Table 1). No data is available concerning the resistance-breaking nature of natural isolates from the additional two clades, *i.e.* clade β and the clade comprising Brazilian isolates (Fig. 3). These data suggest that the *Tsw*-adapted isolates were the results of multiple (at least four) independent evolutions and emphasizes the facility of acquisition of the resistance-breaking properties. In addition, if the *Tsw*-adapted isolates had suffered from a fitness cost (Jenner *et al.*, 2002; Janzac *et al.*, 2010), they would be less competitive than wild-type isolates in plants that do not carry the *Tsw* resistance gene and a strong

effect of the nature of the sampled plants (*Tsw*-carrying or not) would have been expected on the genetic differentiation of the TSWV populations. However, AMOVA did not reveal this kind of effect, even after taking into account the strong geographical effect on the virus genetic differentiation (Table 3). This suggests either that the mutations involved in resistance breakdown are not costly for the virus or that the fitness cost was compensated for since the appearance of the *Tsw*-adapted TSWV variants.

There is presently a controversy about the TSWV factor involved in breakdown of the *Tsw* resistance in pepper. Margaria *et al.* (2007) compared sequences of RNA S from wild-type and resistance-breaking isolates and found that nonsynonymous substitutions in the NSs gene distinguished these categories of isolates, while isolates from these two groups could share the same NP amino acid sequence and the same nucleotide sequence for the nontranslated regions. Therefore, they concluded that the NSs gene was the factor involved in *Tsw* breakdown. In contrast, Lovato *et al.* (2008) used a *Potato virus X* vector to express the NSs or the NP of wild-type TSWV isolates in plants carrying *Tsw*. While the NP was able to induce necroses similar to the hypersensitive reactions induced typically by TSWV in plants carrying *Tsw*, the NSs was not, and Lovato *et al.* assumed that the NP was involved in *Tsw* breakdown. The lack of a reverse genetics system in TSWV precludes the formal validation of these hypotheses. Examination of the evolution pattern of these genes may help identify the true resistance breakdown factor. Given the rapid breakdown of the *Tsw* resistance, the corresponding viral factor has been the target of strong selection and signatures of positive selection can be expected in this factor (Moury *et al.*, 2004; Schirmer *et al.*, 2005; Janzac *et al.*, 2009). Also, the fact that *Tsw* breakdown appeared several times independently, potentially involving multiple nucleotide positions or different substitutions at a given nucleotide position, should increase the evidence of positive selection in the corresponding viral factor. Under the assumption that this factor is of proteic nature, we found that more codon positions were under positive selection in the NSs than in the NP gene (eleven *vs*. one; Table 4). The NP was also significantly more constrained than the NSs, *i.e*. amino acid substitutions are more likely to be deleterious for the virus in the NP than in the NSs. These results converge toward the NSs being the TSWV factor involved in *Tsw* breakdown rather than the NP and the positively-selected codon sites in the NSs gene are candidates for resistance-breaking mutations.

Methods

TSWV isolates

TSWV isolates were collected in Spain from August 2005 to January 2008 (Supplementary Table S1). From June 2007 to January 2008, isolates were collected according to a hierarchical sampling design. Indeed, samples were collected on pepper (isolate names beginning with a P) or tomato (names beginning with a T) plants grown in greenhouses in two provinces (Almería and Murcia), in different administrative communes per province (Fig. 1), and in the greenhouses of 3 to 7 growers per commune, except Dalia and Roquetas where few samples were collected. In June and October 2007, both leaf and fruit samples have been collected on five symptomatic plants for each grower, while afterwards only fruit samples have been collected on 10 different plants for each grower, because viral sequences were nearly identical for samples collected on the same plant (data not shown). Pepper genotypes carrying or not the *Tsw* gene were sampled. Presence of TSWV in the samples was assessed by DAS-ELISA and positive samples were inoculated onto *Nicotiana benthamiana* plants to get high viral titers. About 8 days after inoculation, leaves of *N. benthamiana* plants showing symptoms were collected for molecular analyses. In total, 191 samples have been collected from pepper (95% of samples) and tomato (5% of samples) plants in the provinces of Almería (63% of samples) and Murcia (37% of samples). Thirty-three additional TSWV isolates of different plant and geographical origins have also been analysed (Supplementary Table S2).

TSWV partial genome sequencing

Total RNAs were purified from *N. benthamiana* leaf samples using the Tri Reagent kit (Molecular Research Center, USA) and used for reverse transcription-polymerase chain reaction (RT-PCR) with AMV reverse transcriptase (Promega, USA) and Taq DNA polymerase (Promega). TSWV genome

Version définitive du manuscrit publié dans / Final version of the manuscript published in : Tentchev, *et al.* 2011. Evolution and structure of Tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes. Journal of General Virology 92: 961-973. DOI 10.1099/vir.0.029082-0

regions of highest polymorphism level in RNAs L and M were defined using TSWV sequences available in GenBank in August 2007 and the software DnaSP version 4.50.3 (Rozas *et al.*, 2003). TSWV-polyvalent primers were then defined to amplify these parts of RNAs M and L and the entire NSs and NP genes (Supplementary Table S3). Sequencing reactions were performed directly on RT-PCR products by Genome Express (France).

Sequence analyses

Nucleotide sequences corresponding to the four genome regions were aligned separately using ClustalW (Thompson *et al.*, 1994). For recombination (and reassortment) analyses, the RDP version 2 software (Martin & Rybicki 2000) was used (i) separately for the four alignments, (ii) for an overall alignment resulting from the concatenation of the four previous alignments and (iii) for an alignment resulting from the concatenation of the alignments of the NSs and NP genes, both harboured by RNA S. RDP2 implements six different algorithms to detect recombination events. Only recombination events detected by two or more of the six methods with a probability value threshold of 0.05 were considered. RDP version 3.44 (Martin *et al.*, 2010) provided a consensus recombinant score to distinguish the recombinant and parental sequences for each recombination event. A score above 60% indicates that the identified sequence is almost certainly the recombinant while a score between 40% and 60% indicates a fair likelihood that the recombinant was truly identified (RDP3 Instruction Manual available at http://darwin.uvigo.es/rdp/rdp.html).

For all sequence alignments, the most appropriate nucleotide substitution model was selected by MODELTEST (Posada & Crandall 1998) implemented in the HyPhy software (Kosakovsky Pond & Frost, 2005a) using the Akaike Information Criterion (Akaike, 1974) and the hierarchical Likelihood Ratio Test. The best-fit model was then used for phylogenetic analyses using the maximum likelihood method implemented in PhyML version 3.0 (Guindon & Gascuel, 2003) with the appropriate nucleotide substitution model and all other optional parameters estimated by the program. To evaluate the reliability of the phylogenetic trees, bootstrap analysis was applied using 100 bootstrap resampling. The neighbour-joining method implemented in MEGA (Tamura *et al.*, 2007) was also used for confirmation.

Analysis of the genetic structure of TSWV populations collected on pepper plants in Spain

Genetic differentiation between TSWV populations collected on pepper plants was estimated with the F statistic (Weir & Cockerham 1984) for the six different communes in Spain (four in the province of Almería and two in the province of Murcia) for which sequences of more than 10 isolates were available. We used the software Arlequin version 3.11 (Excoffier *et al.*, 2005) to test for population differentiation (exact test; 100,000 steps in Markov chain; 10,000 dememorization steps) using the overall alignment resulting from the concatenation of the alignments of the four genome regions.

The effects of the province level (Almería and Murcia), of the commune level and of the presence of the *Tsw* resistance gene in the sampled pepper plants were estimated by hierarchically partitioning the populations among these factors, and determining the contribution of each separate factor to the observed genetic variation. For this purpose, AMOVAs (Excoffier *et al.*, 1992) were performed with Arlequin and contrasted against 10,000 permutations to assess the significance of each factor.

Positive selection analysis

To estimate the selection intensity acting on amino acid substitutions in the putative virus factors involved in breakdown of the *Tsw* resistance gene in pepper, the dN/dS ratio between the non-synonymous and the synonymous substitution rates (Kimura, 1983) was estimated for the NSs and NP genes of the Spanish isolates. To compare the selection intensities on the NSs and NP genes, we used the distributions of the dN/dS ratio calculated with the Yang & Nielsen (2000) method implemented in the yn00 program of PAML version 4.2 software (Yang 1997) for pairs of aligned sequences. Since dN/dS estimation is less accurate at low divergence rates, we selected the pairs of sequences showing a divergence greater than 2%. This divergence threshold allowed also excluding pairs of sequences in which the estimation of dS was zero (leading to infinity value for dN/dS). To compare the dN/dS

distributions obtained with the NSs and NP genes, a program written in R (Ihaka & Gentleman 1996) allowed sampling randomly one element from each dN/dS distribution and comparing them. Ten thousand random samplings were performed to calculate the probability that the selection intensities in the NSs and NP genes were not significantly different. Detection of positive selection on particular codon positions in the NSs and NP genes was performed with the SLAC, FEL, IFEL and REL methods implemented in the HyPhy software (Kosakovsky Pond & Frost 2005a, 2005b; Kosakovsky Pond *et al.*, 2006).

Acknowledgements

We thank Mireille Jacquemond, Josselin Montarry and Cécile Desbiez for fruitful discussions about previous versions of the manuscript.

References

Akaike, H. (1974). The new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716-723.

Allen, W. R. & Broadbent, A. B. (1986). Transmission of tomato spotted wilt virus in Ontario greenhouses by *Frankliniella occidentalis*. *Can J Plant Pathol* **8**, 33–38.

Aramburu, J. & Marti, M. (2003). The occurrence in north-east Spain of a variant of *Tomato spotted wilt virus* (TSWV) that breaks resistance in tomato (*Lycopersicon esculentum*) containing the *Sw-5* gene. *Plant Pathol* **52**, 407.

Ayme, V., Petit-Pierre, J., Souche, S., Palloix, A. & Moury, B. (2007). Molecular dissection of the potato virus Y VPg virulence factor reveals complex adaptations to the *pvr2* resistance allelic series in pepper. *J Gen Virol* **88,** 1594-1601.

Ayme, V., Souche, S., Caranta, C., Jacquemond, M., Chadoeuf, J., Palloix, A. & Moury, B. (2006). Different mutations in the genome-linked protein VPg of *Potato virus Y* confer virulence on the *pvr2*^{3} resistance in pepper. *Mol Plant-Microbe Interact* **19,** 557-563.

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.

Bouloy, M. & Flick, R. (2009). Reverse genetics technology for *Rift Valley fever virus*: Current and future applications for the development of therapeutics and vaccines. *Antivir Res* **84,** 101-118.

Brittlebank, C. C. (1919). Tomato diseases. *J Agric* **17,** 231–235.

Brown, J. K., Idris, A. M., Alteri, C. & Stenger, D. C. (2002). Emergence of a new Cucurbit-infecting begomovirus species capable of forming viable reassortants with related viruses in the *Squash leaf curl virus* cluster. *Phytopathology* **92**, 734–742.

Ciuffo, M., Finetti-Sialer, M. M., Gallitelli, D. & Turina, M. (2005). First report in Italy of a resistance-breaking strain of *Tomato spotted wilt virus* infecting tomato cultivars carrying the *Sw-5* resistance gene. *Plant Pathol* **54,** 564.

Desbiez, C., Gal-On, A., Girard, M., Wipf-Scheibel, C. & Lecoq, H. (2003). Increase in *Zucchini yellow mosaic virus* symptom severity in tolerant zucchini cultivars is related to a point mutation in P3 protein and is associated with a loss of relative fitness on susceptible plants. *Phytopathology* ***93***, 1478-1484.

Dietzgen, R. G., Twin, J., Talty, J., Selladurai, S., Carroll, M. L., Coutts, B. A., Berryman, D. I. & Jones, R. A. C. (2005). Genetic variability of *Tomato spotted wilt virus* in Australia and validation of real time RT-PCR for its detection in single and bulked leaf samples. *Ann Appl Biol* **146,** 517–530.

Escriu, F., Fraile, A. & Garcia-Arenal, F. (2007). Constraints to genetic exchange support gene coadaptation in a tripartite RNA virus. *PLoS Pathog* **3,** 67-74.

Excoffier, L., Laval, G. & Schneider, S. (2005). Arlequin ver. 3.0**:** An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Excoffier, L., Smouse, P. & Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes**,** Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Fabre, F., Bruchou, C., Palloix, A. & Moury, B. (2009). Key determinants of resistance durability to plant viruses**,** Insights from a model linking within- and between-host dynamics. *Virus Res* **141,** 140-149.

Flick, R., Flick, K., Feldmann, H. & Elgh, F. (2003). Reverse genetics for Crimean-Congo hemorrhagic fever virus. *J Virol* **77,** 5997-6006.

Fraile, A., Alonso-Prados, J. L., Aranda, M. A., Bernal, J. J., Malpica, J. M. & García-Arenal, F. (1997). Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *J Virol* **71,** 934-940.

García-Arenal, F. & McDonald B. A. (2003) An analysis of the durability of resistance to plant viruses. *Phytopathology*, **93**, 941–952.

Greenough, D. R., Black, L. L., Story, R. N, Newsom, L. D. & Bond, W. P. (1985). Occurrence of *Frankliniella occidentalis* in Louisiana**,** a possible cause for the increased incidence of tomato spotted wilt virus. *Phytopathology* **75**, 1362.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

Harrison, B. D. (2002). Virus variation in relation to resistance breaking in plants. *Euphytica* **124,** 181–192.

Henderson, W. W., Monroe, M. C., Jeor, S. C. S., Thayer, W. P., Rowe, J. E., Peters, C. J. & Nichol, S. T. (1995). Naturally occurring *Sin Nombre virus* genetic reassortants. *Virology* **214**, 602–610.

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.

Ihaka, R. & Gentleman, R. (1996). R**,** A language for data analysis and graphics. *J Comput Graph Stat* **5,** 299-314.

Iturriza-Gomara, M., Isherwood, B., Desselberger, U. & Gray, J. (2001). Reassortment in vivo**,** Driving force for diversity of human rotavirus strains isolated in the United Kingdom between 1995 and 1999. *J Virol* **75**, 3696–3705.

Jahn, M., Paran, I., Hoffmann, K., Radwanski, E. R., Livingstone, K. D., Grube, R. C., Aftergoot, E., Lapidot, M. & Moyer, J. (2000). Genetic mapping of the *Tsw* locus for resistance to the tospovirus *Tomato spotted wilt virus* in *Capsicum* spp. and its relationship to the *Sw-5* gene for resistance to the same pathogen in tomato. *Mol Plant-Microbe Interact* **13,** 673-682.

Janzac, B., Fabre, F., Palloix, A. & Moury, B. (2009). Constraints on evolution of virus avirulence factors predict the durability of corresponding plant resistances. *Mol Plant Pathol* **10,** 599–610.

Janzac, B., Montarry, J., Palloix, A., Navaud, O. & Moury B. (2010). A point mutation in the polymerase of *Potato virus Y* confers virulence towards the *Pvr4* resistance of pepper and a high competitiveness cost in susceptible cultivar. *Mol Plant-Microbe Interact* **23,** 823-830.

Jenner, C. E., Wang, X., Ponz, F. & Walsh, J. A. (2002). A fitness cost for *Turnip mosaic virus* to overcome host resistance. *Virus Res* **86,** 1-6.

Khiabanian, H., Trifonov, V. & Rabadan, R. (2009). Reassortment patterns in swine influenza viruses. *PLoS One* **4,** e7366.

Kimura, M. (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press.

Kirk W. D. J. & Terry I. L. (2003). The spread of the western flower thrips *Frankliniella occidentalis* (Pergande). *Agr Forest Entomol* **5**, 301–310.

Kosakovsky Pond, S. L. & Frost, S. D. W. (2005a). Datamonkey**,** rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**, 2531–2533.

Kosakovsky Pond, S. L. & Frost, S. D. W. (2005b). Not so different after all**,** a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* **22,** 1208–1222.

Kosakovsky Pond, S. L., Frost, S. D. W., Grossman, Z., Gravenor, M. B., Richman, D. D., Leigh Brown, A. J. (2006). Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Comput Biol* **2,** e62.

Lin, H. X., Rubio, L., Smythe, A. B., Falk, B. W. (2004). Molecular population genetics of *Cucumber mosaic virus* in California**,** Evidence for founder effects and reassortment. *J Virol* **78,** 6666-6675.

Lindord, M. B. (1932). Transmission of the pineapple yellow-spot virus by *Thrips tabaci*. *Phytopathology* **22,** 301–324.

Lindstrom, S. E., Cox, N. J. & Klimov, A. (2004). Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972**,** evidence for genetic divergence and multiple reassortment events. *Virology* **328**, 101–119.

Lovato, F. A., Inoue-Nagata, A. K., Nagata, T., de Avila, A. C., Pereira, L. A. R. & Resende, R.O. (2008). The N protein of *Tomato spotted wilt virus* (TSWV) is associated with the induction of programmed cell death (PCD) in *Capsicum chinense* plants, a hypersensitive host to TSWV infection. *Vir Res* **137,** 245-252.

Margaria, P., Ciuffo, M. & Turina, M. (2004). Resistance breaking strain of *Tomato spotted wilt virus* (*Tospovirus*; Bunyaviridae) on resistant pepper cultivars in Almería, Spain. *Plant Pathol* **53,** 795.

Margaria, P., Ciuffo, M., Pacifico, D. & Turina, M. (2007). Evidence that the nonstructural protein of *Tomato spotted wilt virus* is the avirulence determinant in the interaction with resistant pepper carrying the *Tsw* gene *Mol Plant-Microbe Interact* **20,** 547–558.

Martin, D., Lemey, P., Lott, M., Moulton, V., Posada, D., & Lefeuvre, P. (2010). RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.

Martin, D. & Rybicki, E. (2000). RDP**,** detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-563.

McDonald, S. M., Matthijnssens, J., McAllen, J. K., Hine, E., Overton, L., Wang, S., Lemey, P., Zeller, M., Van Ranst, M., Spiro, D. J. & Patton, J. T. (2009). Evolutionary dynamics of human rotaviruses**,** balancing reassortment with preferred genome constellations. *PLoS Pathog* **5,** e1000634.

Miranda, G.J., Azzam, O. & Shirako, Y. (2000). Comparison of nucleotide sequences between northern and southern Philippine isolates of rice grassy stunt virus indicates occurrence of natural genetic reassortment. *Virology* **266,** 26-32.

Moury, B., Morel, C., Johansen, E., Guilbaud, L., Souche, S., Ayme, V., Caranta, C., Palloix, A. & Jacquemond, M. (2004). Mutations in Potato virus Y genome-linked protein determine virulence toward recessive resistances in *Capsicum annuum* and *Lycopersicon hirsutum*. *Mol Plant-Microbe Interact* **17,** 322-329.

Nelson, M. I., Viboud, C., Simonsen, L., Bennett, R. T., Griesemer, S. B., George, K. S., Taylor, J., Spiro, D. J., Sengamalay, N. A., Ghedin, E., Taubenberger, J. K. & Holmes, E. C. (2008). Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathog* **4,** e1000012.

Nemirov, K., Vapalahti, O., Lundkvist, A., Vasilenko, V., Golovljova, I., Plyusnina, A., Niemimaa, J., Laakkonen, J., Henttonen, H., Vaheri, A. & Plyusnin, A. (1999). Isolation and characterization of Dobrava hantavirus carried by the striped field mouse (Apodemus agrarius) in Estonia. *J Gen Virol* **80**, 371–379.

Parrella, G., Gognalons, P., Gebre-Selassie, K., Vovlas, C. & Marchoux, G. (2003). An update of the host range of tomato spotted wilt virus. *J Plant Pathol* **85**, 227-264.

Posada, D. & Crandall, K. A. (1998). MODELTEST**:** testing the model of DNA substitution. *Bioinformatics* **14,** 817-818.

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.

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.

Roggero, P., Masenga, V. & Tavella, L. (2002). Field isolates of *Tomato spotted wilt virus* overcoming resistance in pepper and their spread to other hosts in Italy. *Plant Dis* **86,** 950-954.

Roossinck, M. J. (2002). Evolutionary history of *Cucumber mosaic virus* deduced by phylogenetic analyses. *J Virol* **76,** 3382-3387.

Rozas, J., Sánchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19,** 2496-2497.

Samuel, G., Bald, J. G. & Pittman, H. A. (1930). Investigations on "spotted wilt" of tomatoes, Australia. *Commonw Counc Sci Ind Res Bull* **44,** 8–11

Schirmer, A., Link, D., Cognat, V., Moury, B., Beuve, M., Meunier, A., Bragard, C., Gilmer, D. & Lemaire, O. (2005). Phylogenetic analysis of isolates of beet necrotic yellow vein virus collected worldwide. *J Gen Virol* **86,** 2897-2911.

Sharman, M. & Persley, D. M. (2006). Field isolates of *Tomato spotted wilt virus* overcoming resistance in capsicum in Australia. *Austral Plant Pathol* **35**, 123–128.

Silander, O. K., Weinreich, D. M., Wright, K. M., O'Keefe, K. J., Rang, C. U., Turner, P. E. & Chao, L. (2005). Widespread genetic exchange among terrestrial bacteriophages. *Proc Natl Acad Sci USA* **102**, 19009–19014.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4**:** Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24,** 1596-1599.

Thomas-Carroll, M. L. & Jones, R. A. C. (2003). Selection, biological properties and fitness of resistance-breaking strains of *Tomato spotted wilt virus* in pepper. *Ann Appl Biol* **142,** 235-243.

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., Abad, J., Purugganan, M. & Moyer, J. W. (2005). The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. *Mol Ecol* **14,** 53– 66.

Ullman, D. E., German, T. L., Sherwood, J. L., Westcot, D. M. & Cantone, F. A. (1993). Tospovirus replication in insect vector cells**,** Immunocytochemical evidence that the non-structural protein encoded by the S RNA of tomato spotted wilt tospovirus is present in thrips vector cells. *Phytopathology* **83,** 456-463.

van de Wetering, F., Goldbach, R. & Peters, D. (1996). Tomato spotted wilt tospovirus ingestion by first instar larvae of *Frankliniella occidentalis* is a prerequisite for transmission. *Phytopathology* **86,** 900–905

Watanabe, M., Nakagomi, T., Koshimura, Y. & Nakagomi, O. (2001). Direct evidence for genome segment reassortment between concurrently-circulating human rotavirus strains. *Arch Virol* **146**, 557–570.

Weir, B. S. & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370.

Whitfield, A. E., Ullman, D. E. & German, T. L. (2005). Tospovirus-thrips interactions. *Ann Rev Phytopathol* **43,** 459-489.

Wijkamp, I., Almarza, N., Goldbach, R. & Peters, D. (1995). Distinct levels of specificity in thrips transmission of tospoviruses. *Phytopathology* **85,** 1069-1074.

Wijkamp, I., Van Lent, J., Kormelink, R., Goldbach, R., Peters, D. (1993). Multiplication of tomato spotted wilt virus in its insect vector, *Frankliniella occidentalis*. *J Gen Virol* **74,** 341-349.

Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13,** 555-556.

Yang, Z. & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* **17,** 32-43.

Manuscrit d'auteur / Author manuscript

Table 1. The 38 TSWV isolates showing evidence of reassortment between the three genome segments and their clustering with other isolates (Fig. 2).

Group	Isolate[a]	RDP analysis[b]	Recombinant score	Origin	Clustering for each RNA[c]		
					L	M	S
#1	P51-1, P51-2, P52-1, P53-1, P53-2	3/6; 2×10^{-2} to 6×10^{-4}	0.51	Campania (Italy)	Clusters with clade 3	Clusters with clade *b*	Clusters with clade γ[d]
	CAASO3C	6/6; 2×10^{-3} to 6×10^{-3}	0.43	Alpes Maritimes (France)			
#2	P140, P144, P157-3, P202, P318, P320, P325	2/6; 10^{-2} to 8×10^{-6}	0.70	Almería (Spain)	Within clade 1	Clusters with clade *b*	Within clade α
	P154, P263	4/6; 2×10^{-2} to 7×10^{-4}	0.54				
#3	P110	3/6; 10^{-2} to 8×10^{-3}	0.58	Almería	Clusters with clade 1[e]	Clusters with clade *b*	Within clade α
	P86-1	5/6; 2×10^{-2} to 2×10^{-5}	0.52				
#4	P68-1, P81-2, P165, P166, P169, P214	6/6; 2×10^{-2} to 6×10^{-5}	0.53	Almería	Clusters with clade 1[e]	Within clade *a*	Within clade α
	P190	5/6; 10^{-2} to 6×10^{-5}	0.67				
	P215, P223, P260, P234, P236, P324	4/6; 2×10^{-2} to 3×10^{-5}	0.45				
	P221	3/6; 5×10^{-3} to 3×10^{-7}	0.52				
#5	P60, P61	5/6; 10^{-2} to 6×10^{-8}	0.60	Piemonte (Italy)	Clusters with clade 1[e]	Clusters with clade *b*	Clusters with clade γ[d]
	P59	4/6; 10^{-2} to 8×10^{-4}	0.60				
#6	CMartin	6/6; 10^{-2} to 2×10^{-4}	0.63	Pyrénées Orientales (France)	Within clade 3	Within clade *a*	Within clade α
#7	France81	6/6; 10^{-2} to 10^{-5}	0.51	Bouches-du-Rhône (France)	Within clade 3	Within clade *a*	Clusters with clade γ
	SO46	5/6; 10^{-2} to 6×10^{-4}	0.65	Meuse (France)			
#8	France77	6/6; 10^{-2} to 10^{-5}	0.65	Bouches-du-Rhône	Within clade 3	Within clade *a*	Within clade γ

[a] Underlined sequences correspond to isolates sampled on pepper plants carrying the *Tsw* resistance gene.

[b] Number of methods detecting significant reassortment (*P*-value<0.05) among the six ones implemented in the software RDP (Martin & Rybicki 2000); range of support *P*-values for reassortment obtained with the different methods and isolates of each sequence or group of sequences.

[c] When the reassortant group corresponds to a separate clade from non-reassortant isolates, the closest clade of non-reassortant isolates is indicated and the terminology “clusters with” is used. In contrast, when the reassortant group does not build a separate clade from non-reassortant isolates, the common clade which they belong to is indicated.

[d] relatively low bootstrap support (58%; Fig. 2).

[e] These isolates represent a lineage distantly related to group 1 (Fig. 2).

Manuscrit d’auteur / Author manuscript

Version définitive du manuscrit publié dans / Final version of the manuscript published in : Tentchev, *et al.* 2011. Evolution and structure of Tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes. Journal of General Virology 92: 961-973. 10.1099/vir.0.029082-0

Table 2. F_{ST} values between pairs of TSWV populations from six administrative communes in the provinces of Almería and Murcia (southern Spain) estimated from genetic distances obtained with the concatenation of the sequence alignments of four regions of the virus genome. Values shaded in gray correspond to comparisons within the provinces of Almería or Murcia. Similar results were obtained after excluding the reassortant TSWV isolates (Table 1).

	Berja[a]	Adra[a]	El Ejido[a]	La Mojonera[a]	Torre Pacheco[b]
Adra[a]	0.059*** [c]				
El Ejido[a]	0.002 [ns]	0.095***			
La Mojonera[a]	-0.009 [ns]	0.064***	0.017 [ns]		
Torre Pacheco[b]	0.275***	0.187***	0.256***	0.268***	
San Javier[b]	0.534***	0.411***	0.443***	0.527***	0.221***

[a] Communes from the province of Almería.

[b] Communes from the province of Murcia.

[c] Significance tested by bootstrap analysis based on 10,000 samples. ***: *P*-value < 0.0001; [ns]: *P*-value > 0.05.

Version définitive du manuscrit publié dans / Final version of the manuscript published in : Tentchev, *et al.* 2011. Evolution and structure of Tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes. Journal of General Virology 92: 961-973. DOI 10.1099/vir.0.029082-0

Table 3. Analysis of molecular variance of TSWV genome diversity when the virus populations are grouped by provinces and administrative communes or by categories of pepper variety (with or without the *Tsw* resistance gene) and communes. Similar results were obtained after excluding the reassortant TSWV isolates (Table 1).

Source of variation	df [a]	Sum of squares	Variance component	Percent variation [b]	*F*-statistics [b]
A.					
Between provinces	1	425.3	5.1	24.2 [ns]	F_{CT} = 0.24[ns]
Between communes in each province	4	211.8	1.6	7.3***	F_{SC} = 0.10***
Within communes	151	2187.0	14.5	68.5***	F_{ST} = 0.32***
Total	156	2824.1	21.2		
B.					
Between resistant and susceptible pepper varieties	1	33.44	-0.7	-4.0 [ns]	F_{CT} = -0.04 [ns]
Between communes among each plant category	10	716.18	4.6	25.4***	F_{SC} = 0.24***
Within each commune for each plant category	145	2074.5	14.3	78.6***	F_{ST} = 0.21***
Total	156	2824.1	18.2		
C.					
Between communes	5	637.1	4.1	21.9***	F_{CT} = 0.22***
Between resistant and susceptible pepper varieties in each commune	6	112.5	0.4	2.3[ns]	F_{SC} = 0.03[ns]
Within the categories of plant varieties in each commune	145	2074.5	14.3	75.7***	F_{ST} = 0.24***
Total	156	2824.1	18.9		

[a] df: degrees of freedom.

[b] ***: Significant at $P < 0.001$; [ns]: Not significant at $P < 0.05$.

Version définitive du manuscrit publié dans / Final version of the manuscript published in : Tentchev, *et al.* 2011. Evolution and structure of Tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes. Journal of General Virology 92: 961-973. DOI 10.1099/vir.0.029082-0

Table 4. Codon positions in the NSs and NP genes significantly affected by positive selection according to three methods implemented in the HyPhy software (Kosakovsky Pond & Frost, 2005a). All codon sites were not detected with all methods due to differences in statistical power between methods.

Detection method [a]	NSs gene		NP gene	
	Codon	*P*-value [b]	Codon	*P*-value [b]
FEL (fixed effects likelihood)	130	0.034	none	
	389	0.042		
	390	0.043		
IFEL (internal fixed effects likelihood)	79	0.016	255	0.034
	389	0.050		
	438	0.049		
	462	0.033		
REL [c] (random effects likelihood)	79	0.044	none	
	138	0.040		
	262	0.001		
	264	0.036		
	288	0.033		
	389	0.049		
	438	0.039		
	459	0.032		
	462	0.004		

[a] For the SLAC (Single-Likelihood Ancestor Counting) method implemented in HyPhy, which was shown to have a low statistical power and a low type-I error, no codon site showed significant positive selection at *P-value* < 0.10 significance threshold.

[b] Significance threshold: *P-value* < 0.05.

[c] Due to software limitations, the REL program was run with two subsets of 70 sequences for the NSs and NP genes.

Version définitive du manuscrit publié dans / Final version of the manuscript published in : Tentchev, *et al.* 2011. Evolution and structure of Tomato spotted wilt virus populations: evidence of extensive reassortment and insights into emergence processes. Journal of General Virology 92: 961-973. DOI 10.1099/vir.0.029082-0

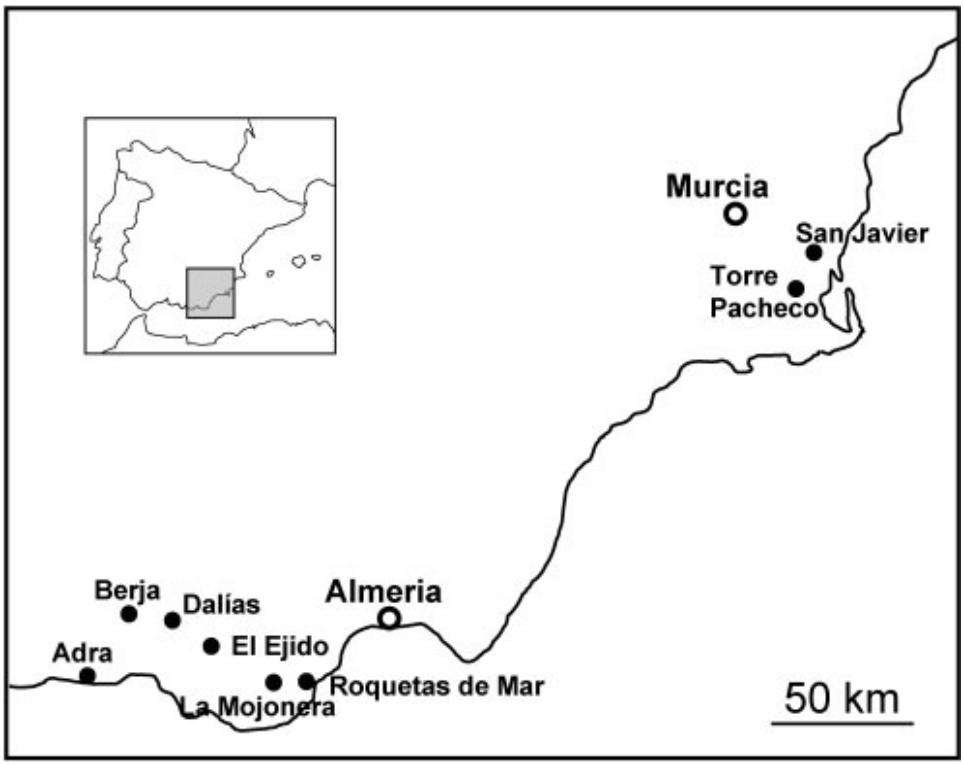

Figure 1. Map of administrative communes in Spain where TSWV samples were collected.

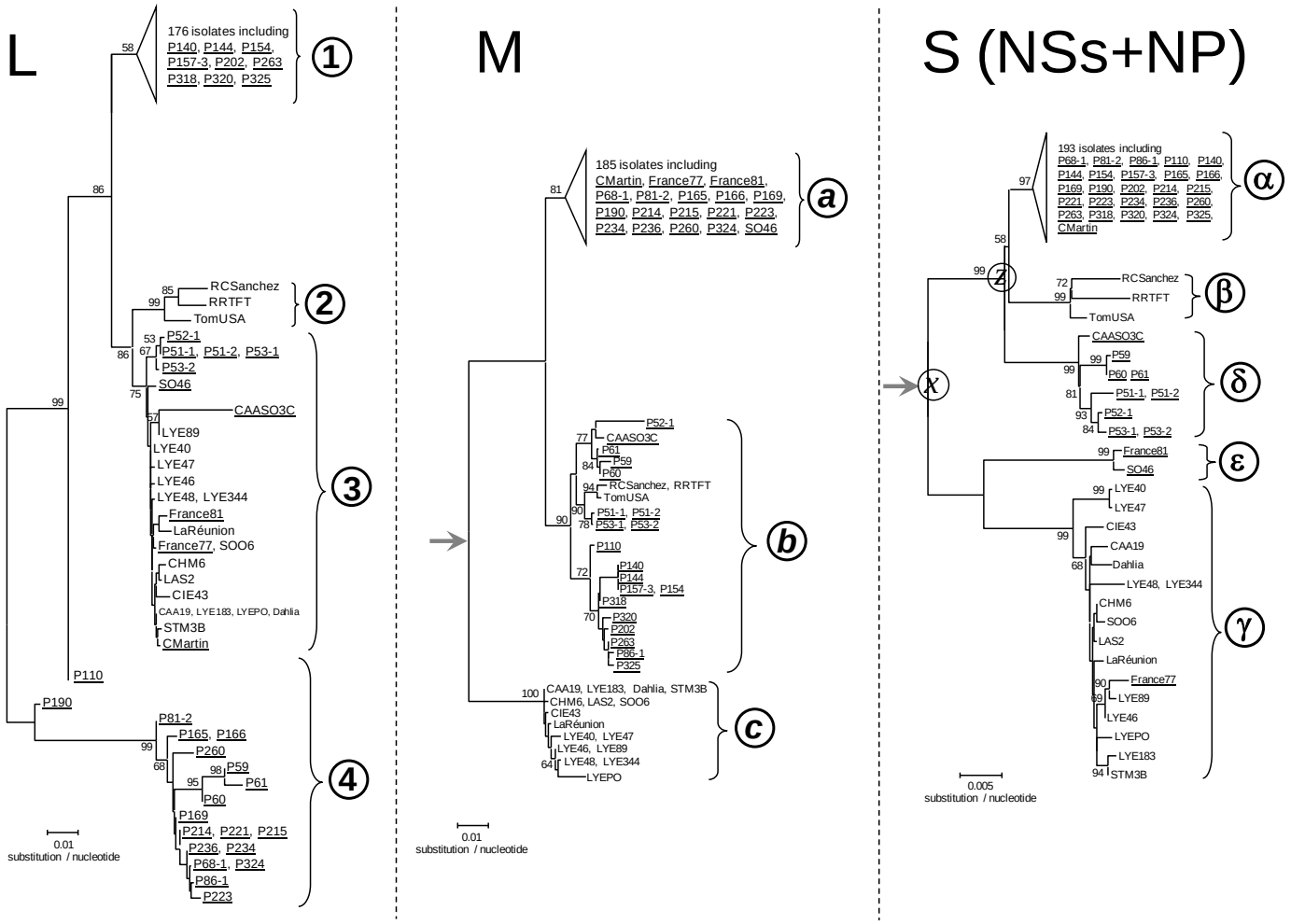

Figure 2. Maximum likelihood phylogenetic trees of the L, M and S genome segments of TSWV isolates sequenced in this study. For RNA S, the NSs and NP genes were concatenated in a single alignment since there was no evidence of recombination between them. Bootstrap percentages above 50% are shown. The roots of the M and S trees based on outgroups *Tomato chlorotic spot virus* (TCSV) and *Groundnut ringspot virus* (GRSV) are indicated by gray arrows and were associated to bootstrap values above 70%. For the RNA S tree, the position of the root was inferred from the NP gene only, since no sequence of the rest of RNA S is available for TCSV or GRSV. Reassortant isolates are underlined.

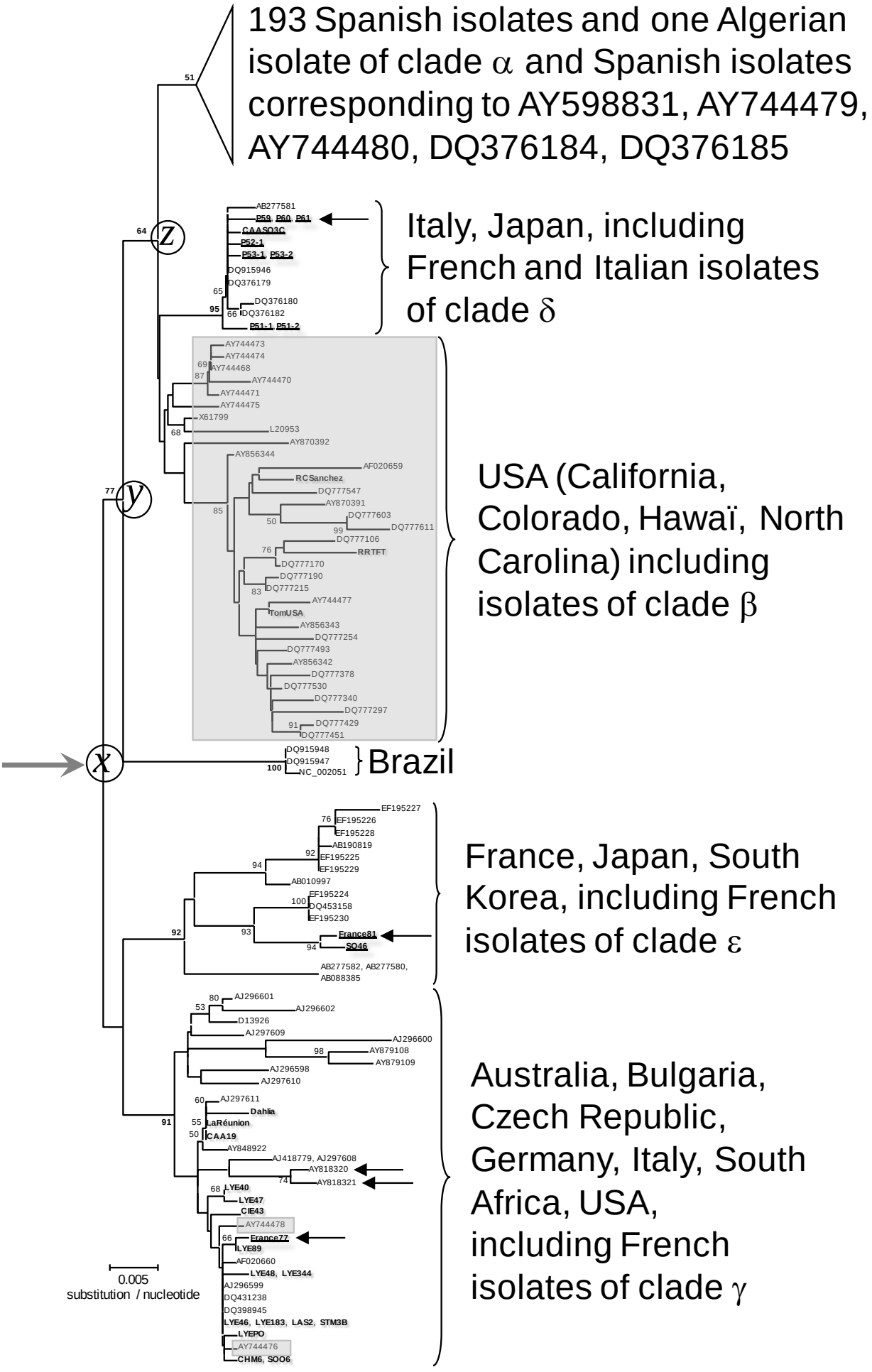

Figure 3. Maximum likelihood phylogenetic tree of the NP gene of TSWV isolates sequenced in this study and of additional isolates available in GenBank. Bootstrap percentages above 50% are shown. Black arrows indicate isolates sampled on pepper plants carrying the *Tsw* gene. The group of Spanish isolates contains also many isolates sampled on pepper plants carrying *Tsw* which were not indicated. The root of the tree based on outgroups *Tomato chlorotic spot virus* and *Groundnut ringspot virus* is indicated by a gray arrow. Underlined isolates are reassortants and isolates from the USA are shaded in gray.

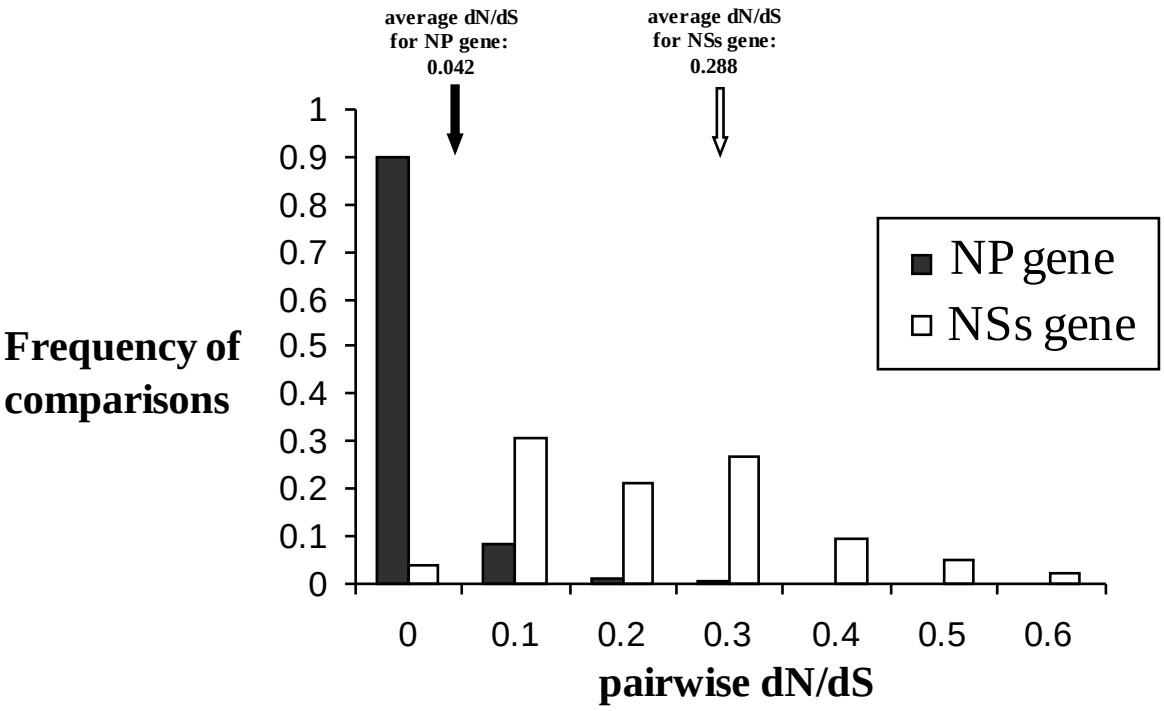

Figure 4. Distributions of pairwise dN/dS values among sequences of the NSs or the NP genes of Spanish TSWV isolates. Only sequences with dN+dS>0.02 were compared. Comparison of the two distributions was performed by 10,000 comparisons of one element sampled randomly in each distribution. In 97% of comparisons, dN/dS was higher for the NSs than for the NP gene.