Received: 4 January 2021 | Accepted: 19 March 2021

DOI: 10.1111/ppa.13382

Plant Pathology WILEY

ORIGINAL ARTICLE

# Genetic variability and evolutionary dynamics of tomato black ring virus population

**Daria Budzyńska**[1] | **Beata Hasiów-Jaroszewska**[1] | **Santiago F. Elena**[2,3]

[1]Department of Virology and Bacteriology, Institute of Plant Protection - National Research Institute, Poznań, Poland

[2]Instituto de Biología Integrativa de Sistemas (I2SysBio), Consejo Superior de Investigaciones Científicas-Universitat de València, València, Spain

[3]The Santa Fe Institute, Santa Fe, NM, USA

**Correspondence**
Beata Hasiów-Jaroszewska, Department of Virology and Bacteriology, Institute of Plant Protection - National Research Institute, Wł. Węgorka 20, 60-318 Poznań, Poland.
Email: B.Hasiow@iorpib.poznan.pl

**Funding information**
Spain Agencia Estatal de Investigación - FEDER, Grant/Award Number: PID2019-103998GB-I00; Narodowe Centrum Nauki, Grant/Award Number: 2015/17/B/NZ8/02407

## Abstract

Tomato black ring virus (TBRV) is an important pathogen infecting a wide range of plant species worldwide. Phylogenetic studies of TBRV have already been conducted, although limited by the use of short genomic regions or a reduced amount of isolates. In the present study, we carried out an exhaustive phylogenetic and population genetic analysis based on the coat protein gene (*CP*) sequence of 57 TBRV isolates originating from different host plants and European geographic regions (47 isolates from Poland, 8 from Lithuania, one from the UK, and one from Hungary). Moreover, the selective pressure acting on particular codons and coevolution of amino acid residues in the CP were analysed. The results clearly showed that the TBRV population is being shaped by recombination and both positive and purifying selection. The analyses revealed that the placement of TBRV isolates in the phylogenetic trees was nonrandom, with isolates clustering according to host plant families and geographic origin.

**KEYWORDS**
evolutionary dynamics, molecular epidemiology, recombination, selective pressure, TBRV

## 1 | INTRODUCTION

The genus *Nepovirus* (family *Secoviridae*) includes numerous plant viruses, and due to their ability to infect a broad range of plant species, including both horticultural crops and ornamental plants, some of them have considerable economic relevance. Currently, the genus includes 40 viruses, which have been separated into three subgroups, A, B, and C (https://talk.ictvonline.org/taxonomy/), based on the size of RNA2, the phylogenetic relationships inferred using the coat protein gene (*CP*), and the cleavage site specificity of the protease (Fuchs et al., 2017).

Tomato black ring virus (TBRV; species *Tomato black ring nepovirus*) is the type member of subgroup B and is widely distributed in temperate regions. Since its first detection in tomato (Smith, 1946), the presence of the virus has been confirmed in a wide range of important crop species. Rymelska et al. (2013) showed that TBRV isolates were biologically diverse and the variability of infection symptoms strongly depended on the combination of viral isolate and the host plant. The infection symptoms include chlorosis, leaf malformation or stunting, and necrotic rings noticeable on infected leaves and fruits (Rymelska et al., 2013). As for other nepoviruses, TBRV is transmitted mechanically and through nematode species from the *Longidorus* genus (Harrison et al., 1961). However, the most considerable challenge from the perspective of controlling TBRV spread is the fact that it is transmitted through seeds of over 30 plant species (Lister & Murant, 1967), and therefore the virus may be easily spread over long distances.

TBRV has non-enveloped, isometric virions about 28 nm in diameter built of 60 CP subunits. The genome structure consists of two linear single-stranded positive-sense RNAs: RNA1 and RNA2. Both RNAs show some flexibility in size (depending on particular isolates), with approximately 7,400 and 4,500 nucleotides (nt) for RNA1 and RNA2, respectively (Zarzyńska-Nowak et al., 2020). The TBRV genome contains two open reading frames (ORFs, one per RNA strand)

  

surrounded by 5′ and 3′ untranslated regions (UTRs) with the small VPg protein attached at the 5′ end and a poly(A) tail at the 3′ end of the RNA molecules. ORFs are translated as polyproteins that then undergo cleavage to functional proteins. Previous research indicated that ORF1 (RNA1) encodes mostly nonstructural peptides responsible for virus replication and genome expression, whereas ORF2 (RNA2) encodes structural proteins involved in virion formation and movement (Digiaro et al., 2015). For some of the TBRV isolates, the presence of additional, subgenomic particles like satellite RNAs (satRNA) or defective RNAs (D RNA) has been observed (Budzyńska et al., 2020; Hasiów-Jaroszewska et al., 2012, 2018; Rymelska et al., 2013; Zarzyńska-Nowak et al., 2020). These particles are relatively short, noninfectious and their replication, encapsidation, and spread depend on the helper virus. Satellite RNAs share little sequence similarity with the viral genomic RNAs, whereas defective RNAs are derived from the genome of the helper virus by further deletions or rearrangements. Short D RNAs (about 335–570 nt in length) obtained so far originated from TBRV genomic RNA1 or RNA2 during prolonged passages in one host (Budzyńska et al., 2020; Hasiów-Jaroszewska et al., 2012; Rymelska et al., 2013). It has been shown that the presence of D RNA has a largely negative impact on TBRV replication and therefore should be rather referred to as defective interfering (DI) RNA. Moreover, the dynamics of accumulation were different in each host, and were affected in a different manner by the presence of DI RNAs, and this effect was host-dependent (Hasiów-Jaroszewska et al., 2018). Pospieszny et al. (2020) investigated the impact of DI RNAs on TBRV vertical transmission through *Chenopodium quinoa* seeds. The analysis revealed that the presence of DI RNAs made the TBRV seed transmission 44.8% more efficient. The presence of satRNAs has also been confirmed for some of the TBRV isolates. TBRV satRNAs consisted of 1,374 nt and did not show sequence similarity with the helper virus genome (Zarzyńska-Nowak et al., 2020). The effect of satRNAs on TBRV replication and accumulation has not been established yet.

Phylogenetic analyses of TBRV conducted to date have revealed a high level of genetic variability of the virus; however, the amount of data included in these analyses was rather limited. Phylogenetic analyses were performed for the Polish and Lithuanian isolates, based on part of the polymerase and *CP* genes, respectively. Polish TBRV isolates collected from different hosts have been grouped in two separate phylogenetic clusters with an average nucleotide similarity ranging from 82.9% to 99.6% (Rymelska et al., 2013). Similarly, Lithuanian isolates tend to cluster into two groups and the level of genetic diversity among them ranges from 78.2% to 97.6% (Šneideris & Staniulis, 2014). The latest study performed by Zarzyńska-Nowak et al. (2020) included the full genomes of nine Polish isolates as well as sequence of TBRV-Mirs isolate available in GenBank (Digiaro et al., 2015). The analyses showed that nucleotide sequence identity of ORF1 and ORF2 ranged from 87.5% to 100% and from 89.4% to 99.9%, respectively. Moreover, the occurrence of potential recombination events was confirmed among TBRV isolates.

Here, we have extended these previous studies of the TBRV population circulating in Poland and Lithuania, increasing knowledge on the genetic variability and evolutionary processes involved in TBRV diversification. Analyses were performed on 57 TBRV sequences, including 47 obtained in this study, thus representing the largest data set that has been analysed for this virus to date. The evolutionary dynamics of the virus was investigated using Bayesian approaches. The association between genetic diversity, host species, and geographic origin of the different TBRV isolates was characterized. Our analytical approach has allowed us to analyse the degree of variability and population genetic structure of TBRV in perennial crops, where the virus inhabits the same host for many years, in comparison with annual plants. To our knowledge, this is the first time that such a study has been conducted for TBRV.

## 2 | MATERIALS AND METHODS

### 2.1 | Plant material

During 2017–2019, surveys were performed and 801 samples originating from different host species displaying symptoms of viral infection were collected. Among them, 548 were collected from *Robinia pseudoacacia* (black locust). TBRV infection for all these samples was confirmed as follows. Approximately 500 mg of plant material was ground with 0.05 M phosphate buffer (pH 7.2) and the obtained sap was used for mechanical inoculation of carborundum-dusted test plants (*C. quinoa*, *Nicotiana tabacum* 'Xanthi', and *Solanum lycopersicum* 'Bethalux'). Subsequently, test plants were maintained under greenhouse conditions at a temperature of 22–23 °C and a photoperiod of 16 hr, and symptoms development was observed. At 10 days postinoculation (dpi), plant material with visible symptoms of infection was collected and the total RNAs were isolated using the RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's protocol. To confirm the presence of TBRV, reverse transcription (RT)-PCR was performed using specific primers amplifying part of the *CP* (Hasiów-Jaroszewska et al., 2015) and Transcriptor One-Step RT-PCR Kit (Roche) according to the manufacturer's instructions. The RT-PCR products were analysed in 1% agarose gel. In addition to these new isolates, other TBRV isolates that were deposited in the Department of Virology and Bacteriology of Institute of Plant Protection – National Research Institute in Poland were also included to the study (Table 1).

### 2.2 | *CP* amplification and cloning

To obtain the complete nucleotide sequences of the *CP* gene, cDNAs were first prepared using 1 μl of previously isolated total RNAs (after the confirmation of virus presence), oligo(dT) primer (200 nM), and the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) according to the manufacturer's protocol. Complete sequences of the *CP* were amplified with two different pairs of specific primers, one previously described by Šneideris and Staniulis (2014) and one designed by us (CPF 5′-TTTTGGGGAAGAGAAACAAC-3′/

**TABLE 1** TBRV isolates used in this study

| No. | Isolate | Host | Region | Collection date | Acc. number |
|---|---|---|---|---|---|
| 1 | TBRV-MJ[a] | *Robinia pseudoacacia* | Mazowieckie, PL | 2002 | MW381200 |
| 2 | TBRV-M1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381195 |
| 3 | TBRV-L1[a] | *R. pseudoacacia* | Mazowieckie, PL | 2001 | MW381194 |
| 4 | TBRV-R2[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2008 | MW381205 |
| 5 | TBRV-SM1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381207 |
| 6 | TBRV-SM2[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381196 |
| 7 | TBRV-NR3[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381202 |
| 8 | TBRV-POB3[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381204 |
| 9 | TBRV-BIS1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381185 |
| 10 | TBRV-B1[a] | *Sambucus nigra* | Wielkopolskie, PL | 2003 | MW381193 |
| 11 | TBRV-B2[a] | *S. nigra* | Mazowieckie, PL | 2011 | MW381191 |
| 12 | TBRV-CK[a] | *Cucurbita pepo* var. *giromontiina* | Kujawsko-Pomorskie, PL | 2004 | MW381187 |
| 13 | TBRV-K5[a] | *C. pepo var. giromontiina* | Wielkopolskie, PL | 2005 | MW381190 |
| 14 | TBRV-K8[a] | *C. pepo var. giromontiina* | Wielkopolskie, PL | 2008 | MW381192 |
| 15 | TBRV-Pi[a] | *Solanum lycopersicum* | Wielkopolskie, PL | 2011 | MG458221.1 |
| 16 | TBRV-PM[a] | *S. lycopersicum* | Wielkopolskie, PL | 2000 | MW381211 |
| 17 | TBRV-Z1[a] | *Solanum tuberosum* | Mazowieckie, PL | 1998 | MW381210 |
| 18 | TBRV-S1[a] | *Lactuca sativa* | Wielkopolskie, PL | 2013 | MW381206 |
| 19 | TBRV-AG[a] | *Tagetes patula* | Wielkopolskie, PL | 2014 | MW381183 |
| 20 | TBRV-AY[a] | *Tagetes erecta* | Wielkopolskie, PL | 2014 | MW381184 |
| 21 | TBRV-O1[a] | *Cucumis sativus* | Wielkopolskie, PL | 2002 | MW381203 |
| 22 | TBRV-CH[a] | *Armoracia rusticana* | Mazowieckie, PL | 2017 | MW381189 |
| 23 | TBRV-Rab[a] | *Rheum rhabarbarum* | Mazowieckie, PL | 2017 | MW381186 |
| 24 | TBRV-Gost2[a] | *R. pseudoacacia* | Mazowieckie, PL | 2017 | MW381188 |
| 25 | TBRV-M2[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2016 | MW381198 |
| 26 | TBRV-Nar6[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2017 | MW381201 |
| 27 | TBRV-Szp6[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2017 | MW381208 |
| 28 | TBRV-Wloc3[a] | *R. pseudoacacia* | Kujawsko-Pomorskie, PL | 2017 | MW381199 |
| 29 | TBRV-Wloc4[a] | *R. pseudoacacia* | Kujawsko-Pomorskie, PL | 2017 | MW381209 |
| 30 | TBRV-Nar1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2017 | MW381212 |
| 31 | TBRV-Buk10[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381213 |
| 32 | TBRV-Byd4[a] | *R. pseudoacacia* | Kujawsko-Pomorskie, PL | 2018 | MW381214 |
| 33 | TBRV-Byd7[a] | *R. pseudoacacia* | Kujawsko-Pomorskie, PL | 2018 | MW381215 |
| 34 | TBRV-Byd8[a] | *R. pseudoacacia* | Kujawsko-Pomorskie, PL | 2018 | MW381216 |
| 35 | TBRV-Byd11[a] | *R. pseudoacacia* | Kujawsko-Pomorskie, PL | 2018 | MW381217 |
| 36 | TBRV-CH3[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381218 |
| 37 | TBRV-OB2[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381219 |
| 38 | TBRV-OP1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381220 |
| 39 | TBRV-OP3[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381221 |
| 40 | TBRV-Pila11[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381222 |
| 41 | TBRV-Pila1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381223 |
| 42 | TBRV-Sier1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381224 |
| 43 | TBRV-WM1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381225 |
| 44 | TBRV-WM2[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381226 |
| 45 | TBRV-WM3[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381227 |

(Continues)

TABLE 1 (Continued)

| No. | Isolate | Host | Region | Collection date | Acc. number |
|---|---|---|---|---|---|
| 46 | TBRV-WS1[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381229 |
| 47 | TBRV-WS2[a] | *R. pseudoacacia* | Wielkopolskie, PL | 2018 | MW381228 |
| 48 | Lt-1[a] | *Phlox* sp. | Širvintos, LT | 2013 | KF678365.1 |
| 49 | Lt-2[a] | *Viola* sp. | Klaipėda, LT | 2013 | KF678366.1 |
| 50 | Lt-3[a] | *Phlox* sp. | Kaunas, LT | 2013 | KF678367.1 |
| 51 | Lt-4[a] | *S. lycopersicum* | Kėdainiai, LT | 2013 | KF678368.1 |
| 52 | Lt-5[a] | *Hosta* sp. | Vilnius, LT | 2013 | KF678369.1 |
| 53 | Lt-6[a] | *Lamprocapnos spectabilis* | Širvintos, LT | 2013 | KF678370.1 |
| 54 | Lt-7[a] | *R. rhabarbarum* | Anykščiai, LT | 2013 | KF678371.1 |
| 55 | Lt-8[a] | *Clematis* sp. | Kelmė, LT | 2013 | KF678372.1 |
| 56 | ED | *Daucus carota* | GB | – | X80831.1 |
| 57 | Mirs | *Vitis vinifera* | HU | – | HG939488.1 |

*Note:* The sequences of isolates 1 to 47 were obtained in this study, whereas sequences of isolates 48 to 57 were retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/).

[a]TBRV isolates used in phylogenetic analyses performed with BEAST v. 2.6.

CPR 5′-TAAGAAATGCCTAAGAAACTA-3′), and Expand Long Range dNTP Pack (Roche). Subsequently, the obtained PCR products were separated in 1% agarose gel and analysed in terms of the appropriate size (c.1,537 and c.1,747 base pairs, respectively). The resulting fragments were ligated into the pCR4-TOPO vector (Invitrogen) and then transformed into One Shot TOP10 chemically competent *Escherichia coli* cells (Invitrogen). For each isolate, at least three independent plasmid DNAs were sequenced using M13F and M13R primers by an external company (Genomed S.A.).

## 2.3 | Sequence alignment

Consensus *CP* sequences of the Polish TBRV isolates were obtained and then compiled with other TBRV *CP* sequences available in GenBank (www.ncbi.nlm.nih.gov/genbank/) using BioEdit (Hall, 1999). Overall, a data set of 57 full *CP* sequences was used, including 47 sequences generated in this study. The sequences were aligned by codons using MUSCLE as implemented in MEGA X (Kumar et al., 2018). The information about the particular isolates (name, host, year of collection, region of origin, and accession number) is presented in Table 1.

## 2.4 | Recombination analysis

To investigate the presence of recombination events within the analysed cistron, two approaches were used. First, a parsimony phylogenetic network was constructed using SplitsTree v. 4.0 (Huson & Bryant, 2006). The statistical confidence of particular nodes was verified based on 1,000 bootstrap pseudoreplicates. Secondly, multiple-sequence alignment was screened for recombination breakpoints using the GARD algorithm as implemented in the Datamonkey Adaptive Evolution Server (www.datamonkey.org; Weaver et al., 2018). According to the GARD results, the full *CP* alignment was divided into two recombination-free partitions near the detected recombination breakpoint (with the correct codon arrangement). The obtained partitions are hereafter referred to as P.1 and P.2, and used as separated data sets for further analyses.

## 2.5 | Phylogenetic and evolutionary dynamics analyses

For these analyses, the partitioned *CP* sequences of 55 TBRV isolates from Poland and Lithuania with known collection date, host plant species, and geographic origin were used (Table 1). First, the temporal time-stamp in our data was investigated using TempEst v. 1.5.3 (Rambaut et al., 2016). jModelTest (Posada, 2008) was used to find the best-fitting substitution model and, therefore, the models with the lowest Bayesian information criterion (BIC) scores were chosen as the most adequate. For partition P.1, the second-best model, GTR+G, was used because the first one with the lowest BIC, SYM+G, was not available in further analyses. For partition P.2, the GTR+G+I model was chosen. Furthermore, for each partition, maximum-likelihood (ML) trees were constructed in MEGA X and then used as an input in TempEst. An analysis of linear regression between the genetic distance from the root and the collection date of the sequences revealed a lack of temporal signal.

Bayesian phylogenetic analyses were performed using BEAST v. 2.6 as implemented on CIPRES Science Gateway (https://www.phylo.org; Bouckaert et al., 2019; Miller et al., 2010). The input XML files were prepared in BEAUTi (BEAST associated program). Data was analysed with the substitution

models described in the previous paragraph. Combinations of strict and relaxed clock models with different tree priors were tested. Finally, the best combination of the molecular clock and tree prior was chosen based on Aikaike's information criterion (AIC; calculated using particular likelihood scores). The main estimation was done with unlinked clock models and trees (accounting for the possibility of different rates of evolution and evolutionary history for P.1 and P.2), under a relaxed clock log-normal model and coalescent constant population prior (same for both partitions). Posterior distributions of parameters were evaluated under Markov chain Monte Carlo (MCMC) of $10^9$ steps with sampling every 10,000 steps and the first 10% of samples discarded as burn-in. The sampling from posterior and the convergence of the obtained results were checked using Tracer v. 1.7.1 software (Rambaut et al., 2018). Subsequently, after burning the first 10% of the sampled trees, the maximum clade credibility (MCC) trees were constructed in TreeAnnotator v. 2.6.3 (separately for each partition). The summarized and annotated MCC trees were visualized with FigTree v. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

### 2.6 | Association between isolate segregation and host plant species or geographic origin

The association between the host plant family of TBRV isolates or their geographic origin (Table 1) and their distribution along phylogenetic trees was evaluated using the TreeBreaker package (Ansari & Didelot, 2016). Two independent analyses using different MCC trees (constructed based on P.1 and P.2) were performed. Each tree (MCC1 and MCC2) was tested in terms of the contribution of two traits: host plant species (grouped in terms of plant family: Fabaceae, Solanaceae, Cucurbitaceae, Brassicaceae, Asteraceae, Adoxaceae, Polygonaceae, Polemoniaceae, Papaveraceae, Violaceae, Asparagaceae, and Ranunculaceae) and geographic origin of isolates.

### 2.7 | Estimation of selective pressure

The selective pressure acting on particular codons was evaluated based on the relative rates of synonymous ($d_S$) and nonsynonymous ($d_N$) substitutions, $\omega = d_N/d_S$ using four different approaches implemented in the Datamonkey Adaptive Evolution Server. Analyses were performed using the *CP* sequences of 57 TBRV isolates divided into recombinant-free P.1 and P.2 regions (in accordance with the above GARD results). For three ML-based methods: fixed effects likelihood (FEL), single likelihood ancestor counting (SLAC), and mixed effects model of evolution (MEME), the sites under positive/purifying selection were estimated under significance value $p < .05$ (Weaver et al., 2018). The results of fast unconstrained Bayesian approximation (FUBAR) method, based on a Bayesian approach, were accepted when Bayesian posterior probabilities (BPP) were at least 0.85 (Weaver et al., 2018).

Finally, for each partition, as well as full-length CP alignment, the residue covariation analysis was performed using the algorithms implemented in the CAPS2 server (http://caps.tcd.ie/; Fares & McNally, 2006) and Spidermonkey-BGM analytical pipeline implemented in Datamonkey (Poon et al., 2008).

### 2.8 | 3D structure prediction of TBRV CP

The 3D CP structure was predicted using the Robetta server (https://robetta.bakerlab.org/; Yang et al., 2020). As input template, the amino acid sequence of TBRV-MJ isolate was used. The resulting 3D structure was visualized with EzMol v. 2.1 (http://www.sbg.bio.ic.ac.uk/ezmol/; Reynolds et al., 2018).

## 3 | RESULTS

### 3.1 | TBRV isolates

In this study, the full *CP* sequences of 47 TBRV isolates were obtained (Table 1). The isolates were collected during field surveys in three Polish geographic regions: Wielkopolskie (33), Kujawsko-Pomorskie (7), and Mazowieckie (7). All of the obtained *CP* sequences are 1,533 nt long and do not contain any deletions or insertions. The sequences were aligned and compared with other TBRV *CP* nucleotide sequences retrieved from the GenBank database (https://www.ncbi.nlm.nih.gov/genbank/).

A comparison of the sequences revealed that the sequence identity of *CP* ranged from 76.9% to 99.8% and 86.4% to 100% at nucleotide and amino acid levels, respectively.

### 3.2 | Recombination analysis

The recombination between TBRV isolates was investigated using SplitsTree v. 4.0 (Figure 1). The numbers near particular branches represent bootstrap support values. The constructed network mostly revealed a star-radiation shape; however, for some of the isolates, a reticulated connection was presented. Three groups of isolates can be distinguished. First, the analyses clearly show the separation of Lithuanian isolates, which together comprise a strong geographic group. The second group consists of the English TBRV-ED isolate, from carrot, together with TBRV-OP3, TBRV-WS1, and TBRV-WS2 Polish isolates from black locust. The third one groups the Polish isolates TBRV-O1 (from cucumber), TBRV-WM1 and TBRV-WM3 (from black locust), and Hungarian isolate TBRV-Mirs (from grapevine).

Furthermore, the existence of at least one recombination event in the TBRV population was confirmed by the GARD method. The significant recombination breakpoint was located between nucleotides 475 and 476 of the *CP* sequences ($p < .01$). This result suggests that *CP* can be divided into two fragments and each of them may have a different evolutionary history. Therefore, for further analyses, the nucleotide

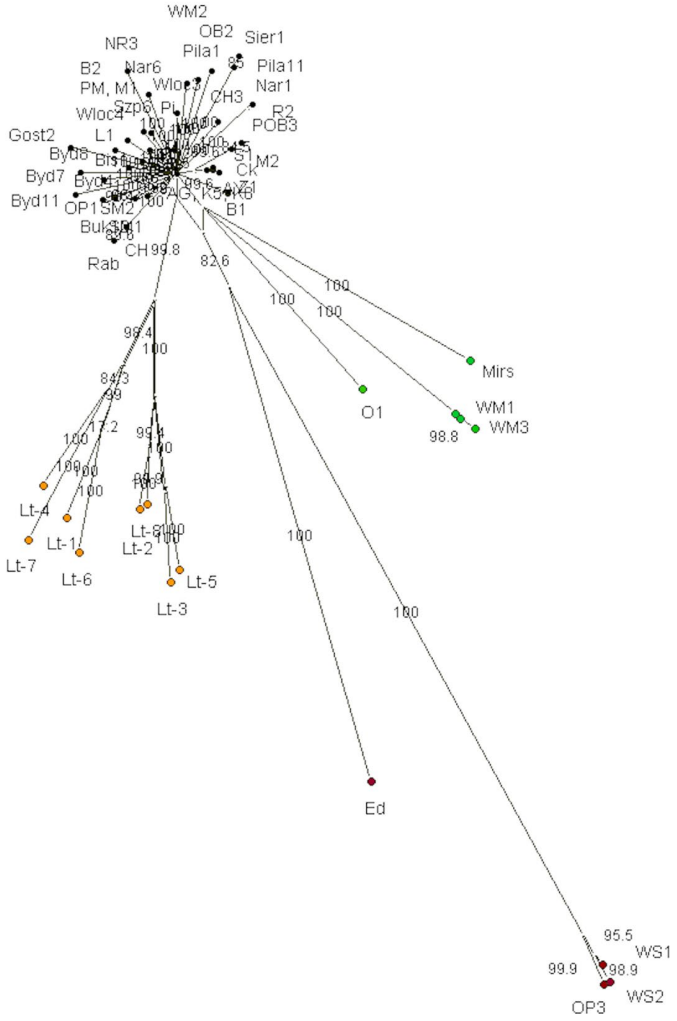


**FIGURE 1** The parsimony splits network constructed in SplitsTree v. 4.0 based on 57 TBRV *CP* sequences. The confidence of particular branches was confirmed by bootstrap resampling (1,000 pseudoreplicates). Three different groups of TBRV isolates are marked with colours: orange, red, and green [Colour figure can be viewed at wileyonlinelibrary.com]

sequences of the *CP* were divided into two partitions near the indicated breakpoint (by maintaining the correct amino acid sequence). The partitions, P.1 and P.2, contain the nucleotides from positions 1 to 477 and 478 to 1,533, respectively. Partitions P.1 and P.2 were retained in all the analyses described in the next sections.

### 3.3 | The phylogenetic reconstruction of TBRV

The sequences of isolates with unknown collection date, origin, or host species were removed, reducing the data set to 55 TBRV *CP* sequences. First, the presence of a time-stamp in the data was tested with TempEst. For P.1, P.2, and the entire protein, the $R^2$ values were 0.05, 0.0003, and 0.0005, respectively, showing the lack of temporal structure. Therefore, the calibration of molecular clock with collection dates was not justified.

Two MCC trees were constructed, each based on a set of 90,001 sampled trees generated by BEAST. Trees MCC1 and MCC2 were constructed based on P.1 and P.2 data sets, respectively (Figures 2 and 3). The numbers above the branches indicate the BPP values. Strongly supported branches were taken as those with BPP ≥ 0.85. Due to the low support of the inner clades, we focused only on the major phylogroups.

On both MCC trees, two main clades can be noticed, one well supported (BPP = 0.86 for MCC1 and BPP = 1 for MCC2) and one less supported with BPP values of 0.63 and 0.59 for MCC1 and MCC2, respectively. Overall, in each tree, a split into three different phylogenetic groups (named A, B, and C) can be proposed. For MCC1, phylogroup A1 consists of 44 Polish isolates originating from different hosts, phylogroup B1 consists of three Polish isolates collected from black locust (TBRV-OP3, TBRV-WS1, and TBRV-WS2), whereas phylogroup C1 only includes isolates from Lithuania. In the case of MCC2, phylogroup A2 contains both Lithuanian isolates and 41 isolates from Poland, and phylogroup B2 contains Polish isolates TBRV-O1 from cucumber and TBRV-WM1 and TBRV-WM3, both from black locust. Phylogroup C2 is formed by three isolates from black locust (TBRV-OP3, TBRV-WS1, and TBRV-WS2) collected in Poland.

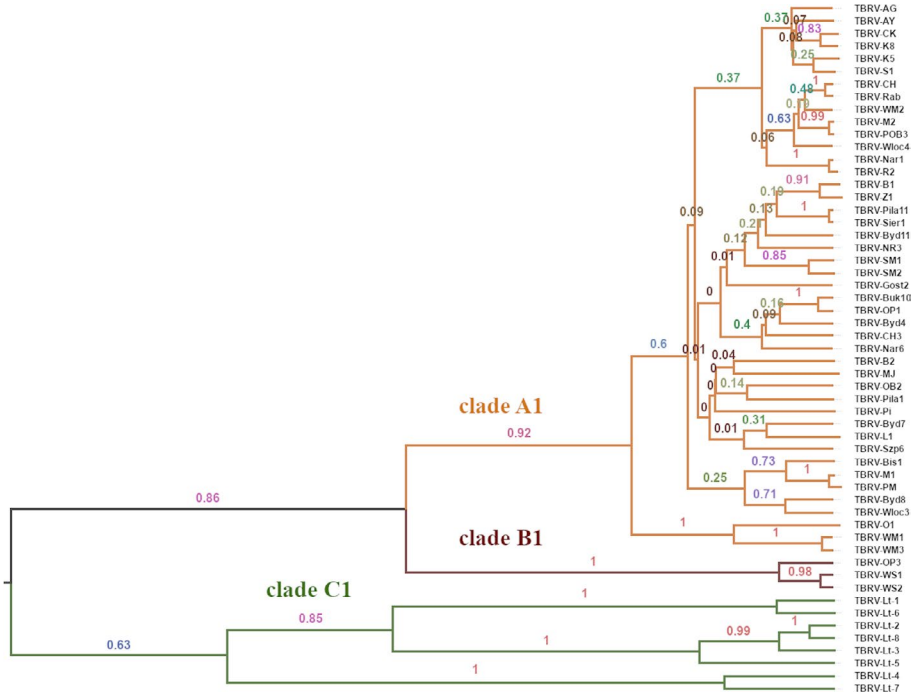


**FIGURE 2** A maximum credibility clade (MCC) tree constructed for the P.1 partition of 55 TBRV *CP* sequences. The numbers above the branches represent the Bayesian posterior probabilities (BPPs) [Colour figure can be viewed at wileyonlinelibrary.com]

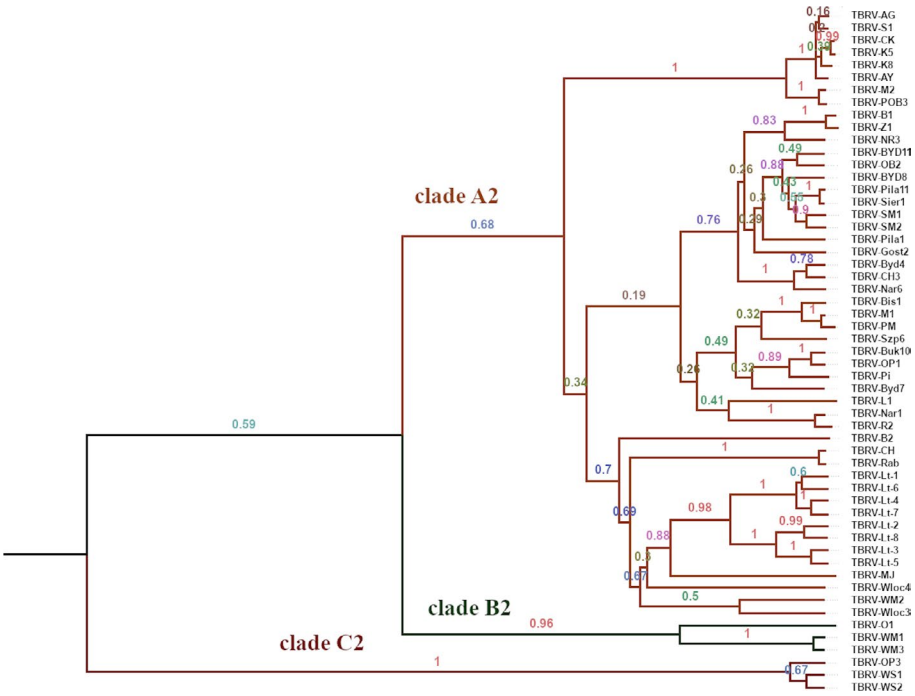


**FIGURE 3** A maximum credibility clade (MCC) tree constructed for the P.2 partition of 55 TBRV *CP* sequences. The numbers above the branches represent the Bayesian posterior probabilities (BPPs) [Colour figure can be viewed at wileyonlinelibrary.com]

### 3.4 | Correlation between distribution of TBRV isolates on MCC trees, their geographic origin, and host plants

Four different trees presenting the correlation between segregation of TBRV isolates on MCC tree and host plant or geographic origin were analysed (Figures 4 and 5). The null hypothesis assuming a random distribution of host plant family/geographic origin (considered as different location in Poland and Lithuania) along the phylogenetic trees was unambiguously rejected. In the case of MCC1, there are signs of strong changes in region (Figure 4a) and host (Figure 4b) distribution from the beginning of isolate diversification in the particular clades. The analyses

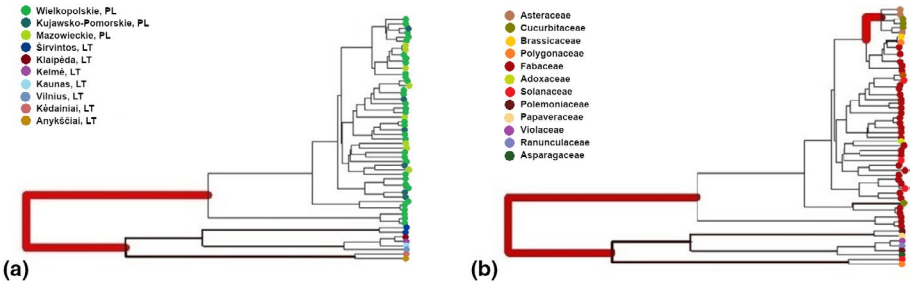


**FIGURE 4** Analyses of switch of (a) geographic distribution and (b) host plant family distribution on the maximum credibility clade (MCC) tree constructed for the P.1 partition of 55 TBRV CP sequences. The thickness and colour of the branches are proportional to the posterior probability of having a changepoint. Different regions of origin or host plant families are indicated with different colours. The distribution of isolates on trees is consistent with the distribution presented in Figure 2 [Colour figure can be viewed at wileyonlinelibrary.com]

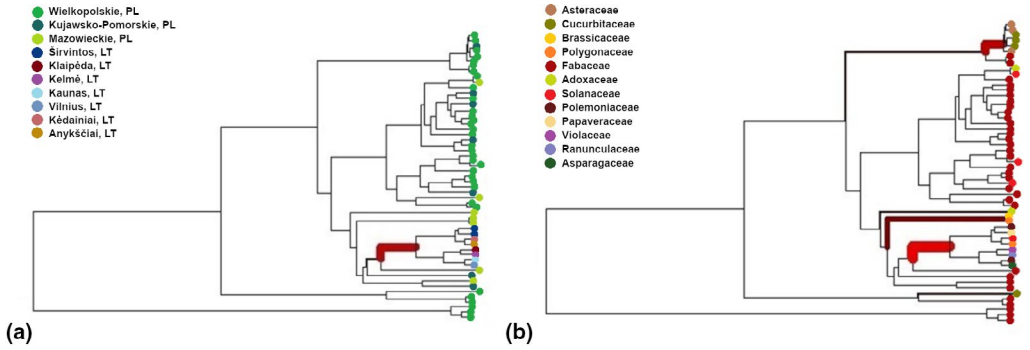


**FIGURE 5** Analyses of switch of (a) geographic distribution and (b) host plant family distribution on maximum credibility clade (MCC) tree constructed for the P.2 partition of 55 TBRV CP sequences. The thickness and colour of the branches are proportional to the posterior probability of having a changepoint. Different regions of origin or host plant families are indicated with different colours. The distribution of isolates on trees is consistent with the distribution presented in Figure 3 [Colour figure can be viewed at wileyonlinelibrary.com]

showed that Polish and Lithuanian isolates clearly constitute two separate main clades (Figure 4a). Moreover, the significant switch of geographic origin distribution for the branch occupied by Lithuanian isolates is consistent with results obtained for MCC2 (Figure 5a). In terms of host plant families, some branches indicate the nonrandom distribution of TBRV isolates. For both MCC1 and MCC2 trees, there is strong evidence of host families switches in the inner branches, which group isolates TBRV-AG, TBRV-AY, TBRV-CK, TBRV-K8, TBRV-K5, and TBRV-S1 from French marigold (*Tagetes patula*), Mexican marigold (*Tagetes erecta*), zucchini (*Cucurbita pepo*), and lettuce (*Lactuca sativa*; Figures 4b and 5b). The host plants of these isolates represent the Asteraceae (TBRV-AG, TBRV-AY, TBRV-S1) and Cucurbitaceae (TBRV-CK, TBRV-K5, TBRV-K8) families. Moreover, for MCC2 the branch grouping the Lithuanian isolates was identified as one with strong switches in host family distribution. The thickness and intensity of the red colour of particular branches on the obtained trees are proportional to the posterior probability of the occurring switches (Figures 4 and 5).

## 3.5 | Selective pressure and coevolving sites

The strength of selective pressure operating on particular codons in the analysed data set was assessed by four different approaches: FEL, FUBAR, MEME, and SLAC. The results indicated that most of the codons had $\omega < 1$, indicative of purifying selection operating on *CP* (283, 384, and 183 sites identified by the FEL, FUBAR, and SLAC methods, respectively); but in some cases $\omega > 1$ values were also obtained, compatible with the action of positive selection at specific amino acid residues. Table 2 shows the localization of codons under positive selection estimated for the P.1 and P.2 regions as well as summarized for full-length *CP*. In total, 17 codons under positive selection were detected (Table 2). The action of positive selection on codon 105 was confirmed by three of the four approaches, whereas in the case of purifying selection, up to 179 sites were identified with at least three methods. All positively selected codons were mapped into the 3D structure of the CP (Figure 6).

Next, we sought for coevolution among different amino acid sites within the CP, that is, coordinated changes. First, the CAPS2 server was used. In the case of P.1, two coevolving groups with a total of five amino acids were detected, whereas for P.2 seven coevolving groups, with a total 28 amino acids, were found. In the case of the complete CP sequences, coevolution was found for a total 51 amino acid sites, grouping into 21 different coevolving groups. Furthermore, some covarying sites were consistent with those assigned by MEME as codons under positive selective pressure (positions 12, 362, 407, 440, and 442). Networks of interacting sites, together with overlapping groups of coevolving amino acids, are presented in Figures S1–S3. In a second stage, covariation analyses were performed using the SpiderMonkey-BGM method. This method found evidence of two coevolving pairs for partition P.1 (Table 3). In the case of partition P.2, as well as complete *CP* sequences, no covarying amino acids were found by this second method.

**TABLE 2** The results of selective pressure analyses performed with four different approaches: FEL, FUBAR, MEME, and SLAC

| | FEL | | FUBAR | | MEME | SLAC | | |
|---|---|---|---|---|---|---|---|---|
| **Region** | $\omega < 1$ | $\omega > 1$ | $\omega < 1$ | $\omega > 1$ | $\omega > 1$ | $\omega < 1$ | $\omega > 1$ | **Amino acid position** |
| P.1 (nucleotides 1–477) | 88 | 1 | 122 | 1 | 3 | 65 | 0 | 12, 84, 105[a] |
| P.2 (nucleotides 478–1,533) | 195 | 0 | 262 | 0 | 14 | 118 | 0 | 9, 32, 36, 69, 145, 146, 203, 246, 248, 270, 281, 283, 318, 331 |
| Full-length coat protein | 283 | 1 | 384 | 1 | 17 | 183 | 0 | 12, 84, 105[a], 168, 191, 195, 228, 304, 305, 362, 405, 407, 429, 440, 442, 477, 490 |

*Note:* Significance threshold was set to $p < .05$ (SLAC, FEL, MEME) and BPP >0.9 (FUBAR). Numbers of codon positions detected under positive ($\omega > 1$) or negative ($\omega < 1$) selection are presented.

[a]Codon in position 105 was confirmed by three of the four approaches.

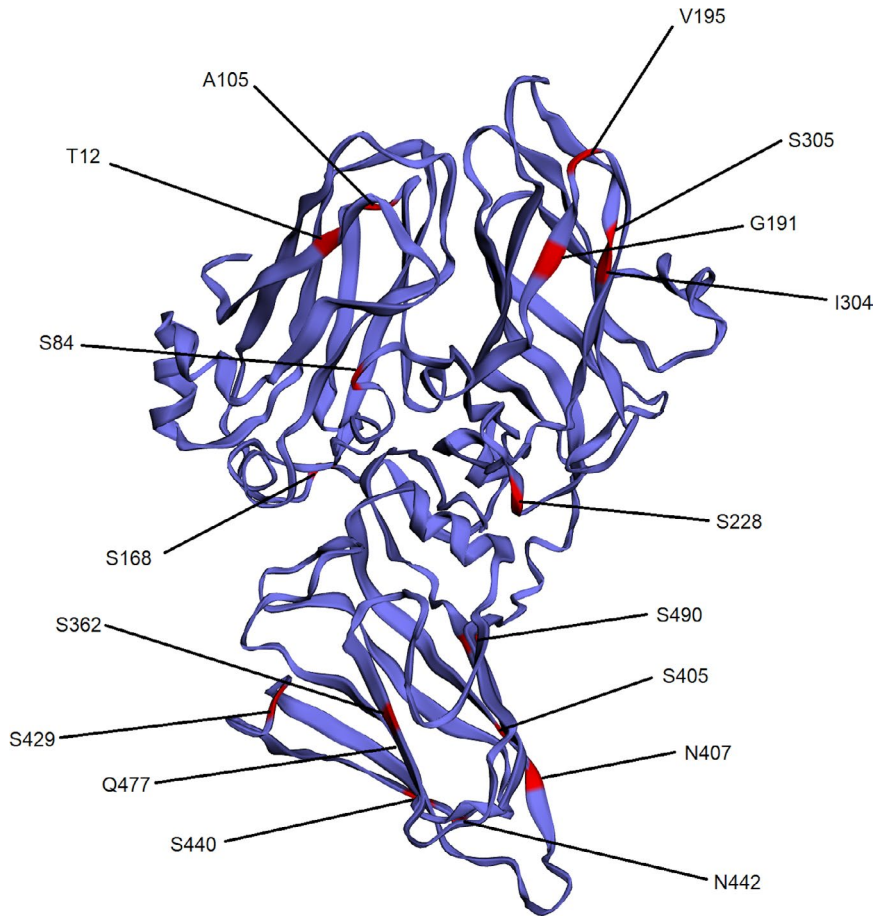


**FIGURE 6** The 3D structure of coat protein (CP) constructed based on TBRV-MJ *CP* sequence using Robetta protein structure prediction server. The codons under positive selection are indicated in red [Colour figure can be viewed at wileyonlinelibrary.com]

## 4 | DISCUSSION

RNA viruses have a great potential for genetic variation, rapid evolution, and adaptation as a consequence of high mutation rates, large population sizes, and very short generation times. Numerous factors have been described as potential drivers of the emergence of viral outbreaks, including pathogen introduction through global trading, changes in vector populations, genetic recombination, new farming techniques, and changes in weather conditions (Wu et al., 2011). The fast evolution and small genomes of RNA viruses make them excellent models for evolutionary studies, and expanding the knowledge about the mechanisms of viral evolution has become an indispensable basis for developing disease control strategies and the risk assessment of new viral emergence events or outbreaks.

In the present study, the *CP* sequences of 57 TBRV isolates originating from different host species and geographic regions were analysed in terms of the phylogenetic relationships between particular isolates, as well as the evolutionary dynamics of the viral population. In our data set, the sequence identity ranged from 76.9% to

**TABLE 3** The groups of coevolving sites detected for different partitions of coat protein (*CP*) with Spidermonkey-BGM approach (www.datamonkey.org)

| Region | Site 1 (S1) | Site 2 (S2) | P(S1→S2) | P(S1→S2) | P(S1→S2) |
|---|---|---|---|---|---|
| P1.1 (nucleotides 1–477) | 5 | 127 | 0.37 | 0.54 | 0.91 |
| | 59 | 68 | 0.63 | 0.35 | 0.99 |
| P2.2 (nucleotides 478–1,533) | — | — | — | — | — |
| Full-length *CP* | — | — | — | — | — |

99.8% and 86.4% to 100% at nucleotide and amino acid levels, respectively. Interestingly, the host plants of the analysed isolates include both annual plants (i.e., lettuce, zucchini), where the life cycle forces the virus to change rapidly, and perennial trees (black locust) and shrubs (elderberry), which can constitute long-term reservoirs for virus evolution, as the long life cycle abolishes the short-term necessity for constantly switching hosts. Virus adaptation to multiple hosts has important implications and it is also thought to be a key mechanism for the maintenance of genetic diversity both in host and pathogen species (Bedhomme et al., 2012). No strict host family- or region-driven structure was found on the constructed trees. However, extended Bayesian MCMC analyses allowed us to identify specific branches on MCC trees on which the distribution of the region/host has changed during the divergence process. Such a well-supported changepoint was indicated on the branch grouping of six Polish isolates (TBRV-AG, TBRV-AY, TBRV-CK, TBRV-K8, TBRV-K5, and TBRV-S1) from annual plants, whereas no statistically supported switch on branches clustering isolates from a perennial plant was found. This finding agrees with the previous ones, assuming that in perennial plants, infections commonly persist for many years and the viral population can potentially be analysed after a much longer within-host evolution than in annual plants (Jridi et al., 2006). The existence of a well-supported switch in the geographic distribution of Lithuanian isolates was also found. The change of the parameter distribution contributes to the MCC2 tree (constructed based on the P.2 partition of *CP*). This partition shows that Lithuanian isolates are more related to the Polish isolates, unlike the relationships resulting from the first of the analysed partitions (P.1 partition, MCC1 tree). This situation may suggest that for the Lithuanian population, some significant geographic change correlated with the occurrence of a recombination event and subsequent spread of the recombinant strain. Oliver et al. (2010) suggested that the dispersion of genetically similar isolates of grapevine fanleaf nepovirus in distant geographic areas is a result of long-distance dissemination, probably through the extensive exchange of propagation material. This theory in part seems to fit our isolates originating from crop species, which are the object of increased turnover of seed or propagation material. Nevertheless, the need for further studies, including a wider range of TBRV isolates, is apparent, and more detailed analysis of the population structure of the virus in terms of its geographical origin and host plant will not be possible until more data are available.

Our research has revealed that recombination seems to be a pervasive phenomenon shaping TBRV population diversity. The role of recombination in genetic structure was confirmed by two different methods. The star radiation-shape shown in the parsimony network, together with reticulated connection in the case of some TBRV isolates, suggests that many of the analysed isolates have a common origin, but conflicting signals in the phylogeny may testify to both multiple evolutionary paths during the diversification of some isolates and the presence of recombinants in the analysed population. The occurrence of a recombination breakpoint between nucleotides 475 and 476 of *CP* results in different clustering of TBRV isolates on the phylogenetic trees. Previous research indicated that recombination potentially increases viral fitness by creating advantageous genotypes and removing deleterious mutations (Elena et al., 2011). In the case of TBRV, it plays an important role in increasing genetic diversity and adaptability, as well as contributing to the appearance of defective RNAs, as deleterious mutants of the parental virus genome (Budzyńska et al., 2020; Hasiów-Jaroszewska et al., 2012, 2018; Rymelska et al., 2013).

Finally, the role of selection in TBRV evolution was considered. First, we suggest that the evolution of TBRV *CP* is driven both by a balance between purifying and positive selection. We found that different methods predicted different sets of codons to be selected; however, for positively selected sites, only one codon was pinpointed by the three methods. In contrast, among those under negative selection, up to 179 were identified by all methods. Moreover, 100 of them were assigned with at least two methods. Secondly, several groups of coevolving amino acids were detected in our strains. Interestingly, some of the coevolving sites were under positive selection. For example, coevolving site at position 360, identified by CAPS2, was previously predicted to be under strong ($p$ = .003) positive selection (Budzyńska et al., 2020). Coevolution at the molecular level is an established phenomenon exhibited by proteins to optimize their performance and serve as an effective determinant of fitness (Jothi et al., 2006). The coevolution, or concerted mutations of amino acid residues allow proteins to maintain their overall structural-functional integrity, and therefore the knowledge about intramolecular coevolution is extremely important in order to know the role of particular regions, and is the basis for designing protein engineering experiments (Chakrabarti & Panchenko, 2010).

In summary, our study revealed that recombination, selection, and potential host switching may have played significant roles in TBRV evolution and may be important factors in the emergence of new viral variants. With this data, we were able to shed some light on the epidemiological and evolutionary dynamics of TBRV.

## ACKNOWLEDGEMENTS

This research was funded by National Science Centre in Poland grant number 2015/17/B/NZ8/02407 (B.H.-J.) and by Spain Agencia Estatal de Investigación-FEDER grant number PID2019-103998GB-I00 (S.F.E.).

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author.

## ORCID

*Daria Budzyńska* https://orcid.org/0000-0002-3923-0829
*Beata Hasiów-Jaroszewska* https://orcid.org/0000-0002-8267-023X
*Santiago F. Elena* https://orcid.org/0000-0001-8249-5593

## REFERENCES

Ansari, M.A. & Didelot, X. (2016) Bayesian inference of the evolution of a phenotype distribution on a phylogenetic tree. *Genetics*, *204*, 89–98.
Bedhomme, S., Lafforgue, G. & Elena, S.F. (2012) Multihost experimental evolution of a plant RNA virus reveals local adaptation and host-specific mutations. *Molecular Biology and Evolution*, *29*, 1481–1492.
Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A. et al (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*, e1006650.
Budzyńska, D., Minicka, J., Hasiów-Jaroszewska, B. & Elena, S.F. (2020) Molecular evolution of tomato black ring virus and de novo generation of a new type of defective RNAs during long-term passaging in different hosts. *Plant Pathology*, *69*, 1767–1776.
Chakrabarti, S. & Panchenko, A.R. (2010) Structural and functional roles of coevolved sites in proteins. *PLoS One*, *5*, e8591.
Digiaro, M., Yahyaoui, E., Martelli, G.P. & Elbeaino, T. (2015) The sequencing of the complete genome of a tomato black ring virus (TBRV) and of the RNA2 of three grapevine chrome mosaic virus (GCMV) isolates from grapevine reveals the possible recombinant origin of GCMV. *Virus Genes*, *50*, 165–171.
Elena, S.F., Bedhomme, S., Carrasco, P., Cuevas, J.M., de la Iglesia, F. et al (2011) The evolutionary genetics of emerging plant RNA viruses. *Molecular Plant-Microbe Interactions*, *24*, 287–293.
Fares, M.A. & McNally, D. (2006) CAPS: Coevolution analysis using protein sequences. *Bioinformatics*, *22*, 2821–2822.
Fuchs, M., Schmitt-Keichinger, C. & Sanfaçon, H. (2017) A renaissance in nepovirus research provides new insights into their molecular interface with hosts and vectors. *Advances in Virus Research*, *97*, 61–105.
Hall, T.A. (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.
Harrison, B.D., Mowat, W.P. & Taylor, C.E. (1961) Transmission of a strain of tomato black ring virus by *Longidorus elongatus* (Nematoda). *Virology*, *14*, 480–485.
Hasiów-Jaroszewska, B., Borodynko, N., Figlerowicz, M. & Pospieszny, H. (2012) Two types of defective RNAs arising from the tomato black ring virus genome. *Archives of Virology*, *157*, 569–572.
Hasiów-Jaroszewska, B., Budzyńska, D., Borodynko, N. & Pospieszny, H. (2015) Rapid detection of genetically diverse tomato black ring virus isolates using reverse transcription loop-mediated isothermal amplification. *Archives of Virology*, *160*, 3075–3078.
Hasiów-Jaroszewska, B., Minicka, J., Zarzyńska-Nowak, A., Budzyńska, D. & Elena, S.F. (2018) Defective RNA particles derived from tomato black ring virus genome interfere with the replication of parental virus. *Virus Research*, *250*, 87–94.
Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*, 254–267.
Jothi, R., Cherukuri, P.F., Tasneem, A. & Przytycka, T.M. (2006) Co-evolutionary analysis of domains in interacting proteins reveals insights into domain–domain interactions mediating protein–protein interactions. *Journal of Molecular Biology*, *362*, 861–875.
Jridi, C., Martin, J.F., Marie-Jeanne, V., Labonne, G. & Blanc, S. (2006) Distinct viral populations differentiate and evolve independently in a single perennial host plant. *Journal of Virology*, *80*, 2349–2357.
Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*, 1547–1549.
Lister, R.M. & Murant, A.F. (1967) Seed-transmission of nematode-borne viruses. *Annals of Applied Biology*, *59*, 49–62.
Miller, M., Pfeiffer, W. & Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: *2010 Gateway computing environments workshop (GCE)*. New Orleans, LA: IEEE, pp. 1–8. https://doi.org/10.1109/GCE.2010.5676129
Oliver, J.E., Vigne, E. & Fuchs, M. (2010) Genetic structure and molecular variability of grapevine fanleaf virus populations. *Virus Research*, *152*, 30–40.
Poon, A.F.Y., Lewis, F.I., Frost, S.D.W. & Kosakovsky Pond, S.L. (2008) Spidermonkey: rapid detection of co-evolving sites using Bayesian graphical models. *Bioinformatics*, *24*, 1949–1950.
Posada, D. (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution*, *25*, 1253–1256.
Pospieszny, H., Borodynko-Filas, N., Hasiów-Jaroszewska, B. & Elena, S.F. (2020) Effect of defective interfering RNAs on the vertical transmission of Tomato black ring virus. *Plant Protection Science*, *56*, 261–267.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*, 901–904.
Rambaut, A., Lam, T.T., Carvalho, L.M. & Pybus, O.G. (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution*, *2*, vew007.
Reynolds, C.R., Islam, S.A. & Sternberg, M.J.E. (2018) EzMol: A web server wizard for the rapid visualization and image production of protein and nucleic acid structures. *Journal of Molecular Biology*, *430*, 2244–2248.
Rymelska, N., Borodynko, N., Pospieszny, H. & Hasiów-Jaroszewska, B. (2013) Analysis of the biological and molecular variability of the Polish isolates of tomato black ring virus (TBRV). *Virus Genes*, *47*, 338–346.
Smith, K.M. (1946) Tomato black-ring: A new virus disease. *Parasitology*, *37*, 126–130.
Šneideris, D. & Staniulis, J. (2014) Phylogenetic analysis of Lithuanian tomato black ring virus isolates. *Zemdirbyste-Agriculture*, *101*, 193–198.
Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V. & Kosakovsky Pond, S.L. (2018) Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*, 773–777.
Wu, B., Blanchard-Letort, A., Liu, Y., Zhou, G., Wang, X. & Elena, S.F. (2011) Dynamics of molecular evolution and phylogeography of barley yellow dwarf virus-PAV. *PLoS One*, *6*, e16896.
Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S. & Baker, D. (2020) Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 1496–1503.
Zarzyńska-Nowak, A., Hasiów-Jaroszewska, B., Budzyńska, D. & Trzmiel, K. (2020) Genetic variability of Polish tomato black ring virus isolates and their satellite RNAs. *Plant Pathology*, *69*, 1034–1041.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Budzyńska D, Hasiów-Jaroszewska B, Elena SF. Genetic variability and evolutionary dynamics of tomato black ring virus population. *Plant Pathol*. 2021;70:1521–1531. https://doi.org/10.1111/ppa.13382