The Complete Genome Sequence of Clade B, Wheat Streak Mosaic Virus Isolate from Turkey

Havva LBAGI (hilbagi@nku.edu.tr)
Tekirdag Namık Kemal University
Rick E Masonbrink
Iowa State University
W Allen Miller
Iowa State University

Research Article

Keywords: Wheat, WSMV, HTS, Illumina sequencing

Posted Date: July 11th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3151096/v1

Abstract

Wheat streak mosaic virus is one of the most widespread viruses in cereal crops, causing severe losses, dramatically affecting worldwide wheat production. Currently, four distinct clades of WSMV have been grouped and named: A (Mexico), B (Europe, Asia, Russia), C (Iran), and D (United States, Argentina, Brazil, Australia, Canada, Turkey). Each of these groups is based on genome-wide variability, emphasizing the CP. Previously reported Turkish wheat isolates of WSMV clustered within both clades D and B. However, the placement of the Turkish WMSV into clade B is only based on a partial genome sequence. Here, we used high throughput sequencing to assemble the complete genome sequence of WSMV type B isolate collected from wheat found in the European part of Turkey. Excluding the poly(A) tail, the genome of isolate S34Edirne (Genbank no. MZ405098) consists of 9,360 nucleotides and contains a single large open reading frame encoding a polyprotein of 3,033 amino acids. The characteristic lack of a GAG (Gly_{2761}) codon within the CP of the polyprotein is typical for the clade B, WSMV-ΔE isolates, which are widely found throughout the European continent. However, two American isolates were recently placed in this group. Sequence comparisons revealed that WSMV Turkish wheat isolate is the most closely related to Czech isolate, with highly similar nucleotide and amino acid identities at 98.83-99.13%, respectively. The result of this study indicates that the WSMV full-length genome of S34Edirne isolate should be placed into clade B of the European WSMV-ΔE isolates.

Introduction

Wheat streak mosaic virus (WSMV) is a species of the Tritimovirus genus in the family Potyviridae causing importantly wheat yield losses. The virus is transmitted effectively by the eriophyid mite *Aceria tosichella* Keifer (Acarina: Eriophyidae) from wheat to wheat (Brakke 1958; Slykhuis 1955; Seifers et al. 1997). WSMV infects many plants in the Poaceae family, including all varieties of wheat (*Triticum aestivum* L.), oat (*Avena sativa* L.), barley (*Hordeum vulgare* L.), maize (*Zea mays* L.), rye (*Secale cereale* L.), pearl millet (*Pennisetum glaucum* L.), sorghum (*Sorghum bicolor* L. Moench), however, dicotyledonous plants are largely unaffected WSMV (McNeil et al. 1996; Seifers et al. 1996; Sigh and Kundu 2017; Hadi et al. 2011). The seed transmission rate has been reported as 0.5 to 1.5%, while experimentally infected wheat cultivars transmit at 0.4%, and very low levels in commercial wheat were with a maximum of 0.22% (Jones et al. 2005; Lanoiselet et al. 2008). Wheat streak disease was identified for the first time in the USA in 1929 as a yellow mosaic (McKinney 1937). Later, it was reported in various states of the USA and Canada (Hunger 2010) as well as in Argentina, Brazil, and Australia (Dwyer et al. 2007; Stenger and French, 2009; Truol et al. 2004). In Europe, the virus was first reported in 1961 in Romania and later in France, Hungary, Russia, Slovakia, Austria, the Czech Republic, Italy, Poland, and Germany (Pop 1961; Gaborjanyi and Nagy 1988; Kudela et al. 2008; Gadiou et al. 2009; Trzmiel and Szydło 2012; Leichtfried 2013; Glinushkin et al. 2013; Schubert et al. 2015). Additionally, WSMV was first identified in wheat, oats, and some grasses in the western part of Turkey by Bremer (1971). WSMV was subsequently reported in various regions of Turkey (İlbağı et al. 2003; Akbaş et al. 2005lıc et al. 2012). The genome size of the newly isolated WSMV is approximately 9.3–9.4 kb long and has a single open

reading frame which is translated into a large polyprotein. This polyprotein is comprised of 10 mature proteins: P1, HC-Pro, P3, 6K1, CI, 6K2, NIa-Pro, NIa-VPg, and CP. The 5' terminus has a VPg, and the 3' terminus has a poly (A) tail (Stenger et al. 1998; Choi et al. 2002). Previously phylogenetic analyses showed that WSMV isolates were grouped into four distinct clades (Rabenstein et al. 2022; Stenger et al. 2002). Accordingly, Clade A represents the El Batan 3 isolate from Mexico, which is the most divergent genome across all clades (Sánchez-Sánchez et al. 2002; Choi et al. 2001). Clade B includes WSMV isolates from Europe and Russia, called WSMV-ΔE isolates, characterized by an identical 3-nucleotide deletion resulting in the lack of Gly_{2761} in the CP (Gadio et al. 2009). Remarkably, some WSMV isolates from the USA were also reported to contain a deletion of GAG at nucleotide positions 8412 to 8414 as being in clade B (Redila et al. 2021). Clade C represents a WSMV isolate from Iran (Rabenstein et al. 2002). Clade D contains isolates from North and South America, Australia, and Turkey (Stenger et al. 2002; Rabenstein et al. 2002; Stenger and French 2009; Robinson and Murray 2013). The complete genome sequences of European isolates in clade B showed differences in the putative protein P1/HC-Pro cleavage site (Choi et al. 2002; Schubert et al. 2015). The P1/HCPro cleavage site for clade A isolate has an HGLRWY/GDS motif, clade B isolates contain the HGLRWY/ C(G) EP (S), and clade D isolates from America and Asia have an HGL(F)RWY/GDQ motif (Schubert et al. 2015). WSMV isolates from various countries worldwide showed that genetic diversity in European populations of WSMV has apparently received less interest (Rabenstein et al. 2002; Gardio et al. 2009). Nevertheless, phylogenetic analysis of the CP cistrons of American isolates are closely related; however, they are distinct from the El Batán 3 sequence (Choi et al. 2001). More polymorphic sites, parsimony informative sites, and increased diversity were observed by Robinson and Murray (2013) in European isolates (Clade I) and American, Australian, and American Pacific Northwest isolates (Clade II), suggesting the more recent establishment of the virus in the latter. Based on phylogenetic analysis of the nucleotide sequence of the CP region, American and Australian isolates share sequence similarities to those of the American Pacific Northwest (APNW) (Dwyer et al. 2007; Stenger and French 2009). On the other hand, a partial sequence from Turkey clustered into clade B (Gadiou et al., 2009); however, there is no information on the complete genome sequence of WSMV from Turkey into clade B.

This study aimed to investigate the viral pathogens in cereals, including wheat and grasses collected from Turkey, using high throughput sequencing. Here, we report the complete and partial nucleotide sequences into clade B associated with WSMV, clustering with other European, Russian, and Asian isolates and some of the newly included three American isolates.

Material and Methods

Plant sampling

A total of 32 plants, including wheat (16), grasses including *Phalaris aquatica* L. (8), *Echinochloa crus-galli* L. (2), *Sorghum halepense* L. (2), *Lolium perenne* L. (1), *Avena fatua* L. (2) and volunteer wheat (1) leaf samples were collected individually to each sample from the wheat fields and border of the cereal

fields in the Trakya region of Turkey. Wheat and grass samples exhibited yellowing, dwarfing, and streak mosaic symptoms; however, *L. perenne* and *E. crus-galli* grass samples were asymptomatic. Collected leaf samples were frozen in liquid nitrogen and stored at − 80 ◦C in a deep freezer until used.

RNA Extraction, Library Preparation and HTS

For HTS analysis, total RNA was extracted individually from symptomatic and asymptomatic each leaf sample using either a Direct-Zol Miniprep kit (Zymo Research, Irvine, CA, USA), and RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. RNA was treated with Turbo DNAse (Thermo Fisher Scientific, Waltham, MA, USA). After resuspending the extracted total RNA, the quality and quantity were checked with a Qubit 3.0 fluorometer and a NanoDrop 2000. Strand-specific cDNA libraries were prepared from ribosomal depleted RNA using the NEBNext Ultra II Directional RNA Library Prep kit (New England Biolabs, Ipswich, MA, USA). The quality of the cDNA library was also checked at 2100 Bioanalyzer System-High Sensitivity DNA Assay (Agilent, Santa Clara, CA, USA). Subsequently, 32 cDNA libraries were run on an Illumina HiSeq 3000 (Illumina Inc., San Diego, CA, USA) at the Iowa State University in DNA Facility.

Determination of the Full-Length Genome Sequence by Bioinformatic Analysis

The raw reads were trimmed for adapters and sorted by sample, and the quality of sequences was analyzed with Cutadapt. Reads were assembled with Spades (Bankevich et al. 2012) using default parameters with the RNA flag. Following, the reads were mapped back to the assemblies using Hisat2 (Kim et al. 2017), converted to bam with Samtools (Li et al. 2009), and read counts were processed using feature counts from the Subread package (Liao et al. 2009). Annotation of assembled contigs was performed using NCBI Blast 2.11.0+ (Altschul et al. 1990) to the current ViralDb (last updated February 2021) within a threshold of identity at 50% with an e-value of 10^{-5}. Each potential ORF was translated and input into BLASTp to NCBI nr to determine putative functions. Phylogenetic trees were constructed using 23 whole-genome sequences of WSMV isolates were retrieved from NCBI, as shown in Table 1. The phylogenetic analysis was inferred using the neighbor-joining method in Mega 7, applying the Tamura-nei model and 1000 bootstrap replicates (Kumar et al. 1993). Pairwise alignments of the nucleotide and deduced amino acid sequences were done using Bioedit 7.2 (version 7.2.5; https://bioedit.software.informer.com/7.2, Hall 1999).

RT-PCR test and Sanger Sequencing

Total RNA from the wheat leaf sample was isolated using a Qiagen plant mini kit (Qiagen, Hilden, Germany) in order to confirm HTS data. The concentrations of the isolated RNA were determined using Nanodrop (2000). cDNA was obtained using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Waltham, USA) and used to amplify viral sequence the primers designed from assembled contigs. RT-PCR test was performed using DreamTaq DNA polymerase Mastermix (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's recommendations. The thermocycling conditions were as follows: initial denaturation step of 95 ◦C for 3 min; 35 cycles of 95 ◦C

for 30 s, 52 ◦C for 60 s, and 72 ◦C for 30 s; and a final extension step of 72 ◦C for 5 min. After conducting RT-PCR, the amplification product was visualized by gel electrophoresis using 1.5% agarose gels under UV light. Sanger sequencing was performed at Refgen Biotechnology company, Ankara, Turkey.

RESULTS

A total RNA library was individually constructed from each samples and sequenced using an Illumina platform to identify the potential pathogenic agents in cereal and grass samples collected from cereal-growing fields in Trakya, Turkey. Sequencing yielded a total of 4.612 million paired-end reads at 150 nucleotides each. These reads were assembled into 55 contigs with Spades and showed a high identity to wheat streak mosaic virus with a range of 88.3–100% identity and extremely low e-values at or below $1e^{-54}$. Of the 4.6 million reads, 6,379 reads were aligned back to the genome, achieving coverage of 102x. The scaffolded genome of the S34Edirne wheat isolate (Genbank acc.no. MZ405098.1) is 9,360 nt long, excluding the polyA tail at the 3' end. A partial sequence from a wheat isolate (Genbank acc.no. ON364116.1) has been also obtained which is 3348 nt long. The amplicon of WSMV (434 bp) obtained by RT-PCR test was identical to the sequence assembled from HTS. Thus, the complete genome of S34Edirne wheat isolate encodes a single polyprotein of 3,033 amino acid residues and has a typical genome organization of the Potyviridae family. This single polyprotein produces ten mature proteins containing P1 (352 aa, 40.19 kDa), HC-Pro (384 aa, 44.29 kDa), P3 (278 aa, 32.12 kDa), 6K1 (51 aa, 6 kDa), CI (644 aa, 72.48 kDa), 6K2 (51aa, 5.88 kDa), NIa-VPg (197aa, 23.16 kDa), NIa-Pro (229aa, 25.73 kDa), NIb (499aa, 56.95 kDa), CP (348aa, 36.52 kDa). The 5' UTR of S34Edirne is 130 nt, while the 3' UTR is 130 nt long. The poly-A tail contains in the S34Edirne genome a long tail at 20 nt. The protein sequence contains a conserved 'GYCYM' pentapeptide motif beginning at amino acid position 623 within the carboxy-terminal third of HC-Pro as occurred in WSMV polyprotein sequences, as well as a tyrosine residue at position 1882 of the peptide sequence results in the motif ‘YGFDP’. The genome of S34Edirne includes a codon deletion (GAG) in the CP region of the polyprotein as other European/Asian isolates and two American isolates recently included into clade B showed having three nucleotide deletions at positions 8,412 to 8,414. This gap corresponding to one amino acid deletion within the CP region of polyprotein is identical in other European isolates. The CP gene of the S34Edirne isolate contains the motif HGLRWY/CEP in the putative protein P1/helper-component proteinase (HCPro) protease cleavage site as other European isolates from Germany, Ukraine, Poland, the Czech Republic, and two American isolates; however, one isolate Marmagne from France differs in one amino acid as shown HGLRWY/CES. In contrast, the American and Asian isolates contain the motif HGL(F)RWY/GDQ, though the El-Batan isolate from Mexico has a significantly different motif from WSMV sequences (Fig. 1).

The complete and partial nucleotide sequences of WSMV Turkish wheat isolates S34Edirne and S34-Edirne were analyzed and compared with 27 complete nucleotide sequences of WSMV isolates available from the EMBL-EBI and GenBank databases, as exhibited in Table 1. The phylogenetic tree constructed from these sequences reveals three clustered clades A, B, and D (Fig. 2). The Turkish wheat isolates clustered with European, Asian, and three American isolates within clade B and placed the most closely

with DSMZ PV-0356 isolate from Ukraine in the same branch. The isolates from the USA, Australia, Argentina, Turkey, and Iran were placed into clade D. Moreover, the El-Batan isolate clustered in clade A as it is quite different from the other isolates of WSMV. However, none of the WSMV isolates clustered into clade C based on the complete and partial nucleotide sequences of WSMV (Fig. 2).

Sequence comparison with S34Edirne revealed that the highest level of nucleotide identity was 98.84% for the Czech isolate, while the lowest nucleotide identity was 79.14% for El Batan 3 isolate from Mexico. Amino acid multiple sequence alignments of WSMV isolates revealed that the highest level of identity was 99.13% with isolate Czech isolate from the Czech Republic. In comparison, the lowest level of identity was 88.81%, with El Batan 3 isolate from Mexico.

Using all available WSMV complete genome sequences, we were able to associate the two Turkish wheat isolates within clade B WSMV genotypes. The S34Edirne and S34_Edirne isolates are from the Edirne province, located at the border of Greece and Bulgaria. Moreover, the Turkei isolate was taken from Tekirdağ, located in Trakya, the European part of Turkey. Thus, two isolates from Trakya were placed into clade B with other European WSMV isolates. While it is possible that seed or wheat curl mites could transmit these isolates. Because Edirne's cereal fields are fairly close to cereal fields at the border of two countries; thus, the virus might be spreading both via wheat curl mites by the wind, or by contaminated seeds during seed import between countries even if WSMV transmit at the lowel level by seed. Thus, this could happen either by natural distribution through its viruliferous vector wheat curl mites or by disseminating the virus-infected seed materials. On the other hand, Turkey1 and Turkey 2 isolates are included in clade D and are from central Turkey. These two WSMV isolates are well close to American WSMV isolates. It might be possible to materialize during the exchange of the past century germplasm between countries (Fig. 3). Obtained these results point out that two different WSMV genotypes are coexisting in Turkey. Further research on WSMV isolates may enable researchers to better define the genetic diversity across divergent WSMV genotypes.

DISCUSSION

The complete nucleotide sequence of the S34Edirne isolate and a partial sequence of the S34_Edirne isolate were assembled and annotated in the present study. The local WSMV strain in Trakya was placed into clade B using a partial sequence (isolate Turkei). In a previous study, it was revealed that the complete genome for isolate Turkey1 and a partial CP sequence of Turkey2 isolate clustered into clade D by having a high sequence identity with WSMV American isolates and lacked indels relative to Sidney 81 and Type isolates (Rabenstein et al. 2002; Gadiou et al. 2009). This study shows the first complete genome of WSMV took place in a European isolate in clade B between Turkish isolates so far identified. Turkish strains WSMV were first identified in 1971 in wheat, oats, and some grasses in western Turkey (Bremer 1971). Subsequent studies showed that the presence of WSMV in wheat-growing fields in opposite regions of Turkey (İlbağı et al. 2003; Akbaş et al. 2005lıc et al. 2012). Gadiou et al. (2009) suggested that the Turkei isolate, derived from the Trakya region of Turkey, had clustered within clade B. Subsequently, this might be indicates two distinct WSMV genotypes may coexist in Turkey. However, our

results reveal that the WSMV sequences of two isolates resides in clade B with European and Asian isolates. Nevertheless, Turkey 1 and Turkey 2 isolates remain in clade D with American isolates as pointed out in previous study. All the results associated with WSMV so far show that WSMV could be transmitted through seed into the country, or by wheat curl mites to the neighboring cereal fields at the border of the country. Moreover, it might be also possible another reason spreading by exchange breeding materials. All means of the WSMV spreading globally should be investigated in detailed further studies. Thus, the genome variability could cause problem in resistance breeding in order to develop the resistance cultivars to virus. The genome of S34Edirne contains a three-nucleotide deletion (GAG) resulting in one amino acid deletion (Gly_{2761}) within CP gene region of the polyprotein, a characteristic common to all European WSMV isolates (Gadiou et al. 2009). Recently, three isolates (NE01_19, KM19 and R020) from the USA were determined to contain three nucleotide deletions in the CP gene, which is typical in the European isolates (Redila et al. 2021). Altogether, this evidence indicates that transmission of these strains is likely via human-assisted movement from Europe to the USA, despite the fact that WSMV is transmitted by seed at very low levels (0.5–1.5%) (Jones et al. 2007; Lanoiselet et al. 2008). However, WSMV is also transmitted by the eriophyid mite *Aceria tosichella* Keifer (Acarina: Eriophyidae) between wheat plants at short and medium distances (Brakke 1958; Seifers et al. 1997). Thus, the mites could spread to short distances by wind but to long distances spread could be performing by assisted by flying insects (Slykhuis 1955; Gibson and Painter 1957). As a matter of fact, Rabenstein et al. (2002) reported that the hard red winter wheat (*Triticum aestivum* L.) in South Dakota, Nebraska, and Kansas was due to the introduction of 'Turkey wheat', which had been extensively exchanged between the Great Plains and Black Sea regions. Heun et al. (1997) indicated the earliest cultivated wheat forms were domesticated feral grasses, with genetic relationships placing their origination in the Karacadağ Mountain in the south-eastern part of Turkey. Thus, further studies could identify other strains of WSMV in the Turkey, as it is the origin of domesticated wheat. Apart from this, Asian isolates from Iran also contain the divergent strains of WSMV in clades C and D (Stenger and French 2009; Robinson and Murray 2013; Rabenstein et al. 2002; Gadiou et al. 2009; Schubert et al. 2015; Sigh et al. 2018; Redila et al. 2021). Thus far, studies of WSMV show that isolates are separated into four distinct clades (A-D) by nucleotide sequence comparisons. Clade A is from Mexico which contains El Batan isolate; clade B is from Europe and Russia; clade C is from Iran; and clade D, which contains some the North American isolates and two Turkish isolates (Stenger et al. 2002; Rabenstein et al. 2002). The Mexican isolate El-Batan3 clustered in clade A is the most divergent isolate from the North American strains and has 79% nucleotide sequence identity to Sidney 81 and Type (Sánchez-Sánchez et al. 2002; Choi et al. 2001). Recently Sigh and Kundu (2017) identified and clustered subclade B1 in some grass isolates into clade B. Similarly, four subclades: D1, D2, D3, D4 within clade D isolates from American origin were clustered by French and Stenger (2002). The genome of S34Edirne has 9360 nt excluding at poly-A tail at 3´terminus; however, the poly-A tail in the S34Edirne genome contains a long tail at 20 nt differentially than other WSMV isolates. Nevertheless, WSMV RNA encodes a single polyprotein of 3,033 amino acid residues and has a genome organization typical for a member of the family Potyviridae (Stenger et al. 1998; Choi et al. 2000; Sigh et al. 2018). Stenger et al. (1998) indicated that a conserved $GYCYM$ pentapeptide motif and $YGFDP$ motif at amino acid 1882 in WSMV polyproteins. In addition, our study presents that the

genome of S34Edirne isolate contains two motifs as other WSMV isolates. The phylogenetic tree indicates that S34Edirne and S34_Edirne isolates cluster to clade B, while only three clades were identified here, lacking a representative of clade C as reported by Schubert et al. (2015).). Sequence comparisons showed that the isolates of S34Edirne had homology at the nucleotide and amino acid identity level of 98.84–99.13% with Czech isolate within clade B. Similarly, Gadiou et al. (2009) stated that WSMV-ΔE isolates share 97.5–99.0% nucleotide identity with the completely sequenced Czech WSM isolate and was identical at the amino acid sequence level. Mishchenko et al. (2019) also reported the isolate Ukraine-Mal-18 has a high level of the sequence identity (93.5%-95.9% nt and 93.6–95.0% aa) with the clade B isolates. As reported, type B isolates shared a high similarity of ~ 92% (nt) and 94% (aa) to type B1 isolates and of ~ 90% (nt) and 95% (aa) to type D isolates.

Because of the recent effects of global climate change, human-assisted movement of WMSV worldwide or exchange of breeding materials could be a cause for the new WSMV genotypes, though genetic divergence may have occurred locally. Thus, all studies of WSMV show that much of the divergence among WSMV strains could be explained by genetic drift (Choi et al. 2001). Our results confirm the coexistence of two distinct WSMV genotypes in Turkey. The isolates S34Edirne and S34_Edirne of this study clustered to clade B with European and Asian isolates. In contrast, in the former study, two isolates (Turkey 1 and Turkey 2) from Turkey clustered to clade D, and have high nucleotide sequence identities with American isolates. Further studies will be needed to determine the existence other WSMV genotypes in Turkey and the world. In addition, the prevailing control strategies should be modified to better identify how WSMV is transmitted, especially focused on seed transmission to prevent spreading globally.

Declarations

The authors declare no conflict of interest

Funding Information

This study was funded by the Scientific and Technological Research Council of Turkey, 512 International Postdoctoral Research Scholarship Program (TUBITAK-BIDEB), and Tekirdağ 513 Namık Kemal University, The Scientific Research Projects Coordination Unit (NKU-BAP, Project 514 No: NKUBAP.03.GA.21.289) to H.I., and the Iowa State University Plant Sciences Institute and 515 the DARPA Insect Allies Program funding to WAM. This paper of the Iowa Agriculture and Home 516 Economics Experiment Station, Ames, IA, Project No. 4308 was supported in part by Hatch Act 517 and State of Iowa funds.

Author contribution

HI contributed to conceptualization, investigation, writing-original draft, writing-review; RM contributed to analyze the data, writing-review and editing; WAM contributed writing-review and editing.

Data availability statement

The complete genome and partial genome sequences of the WSMV are in GenBank under accession number MZ405098 and ON364116, and SRA accession number SRS5023133.

References

1. Akbaş B, İlhan D, Güner Ü (2005). Seed-borne viruses in wheat seeds used by growers in the Central Anatolia region. Plant protection bulletin 45(1):55–60
2. Albrecht T, White S, Layton M, Steiglein M, Haley S, Nachappa P (2022). Occurrence of wheat curl mite and mite-vectored viruses of wheat in Colorado and insights into the wheat virome. Plant Disease 106(10):2678–2688
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol. https://doi.org/10.1016/S0022-2836(05)80360-2
4. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. (2012) Spades: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5):455–477
5. Brakke MK (1958) Properties, assay and purification of wheat streak mosaic virus. Phytopathology 48:439–445
6. Bremer K (1971) Wheat streak mosaic virus in Turkey. Phytopathologica Mediterranea 10: 280–282
7. Choi IR, Hall JS, Henry M, Zhang L, Hein GL, French R, Stenger DC (2001) Contributions of genetic drift and negative selection on the evolution of three strains of wheat streak mosaic Tritimovirus. Archives of Virology 146: 619–628
8. Choi IR, Horken K, Stenger DC, French R (2002) Mapping of the P1 proteinase cleavage site in the polyprotein of wheat streak mosaic virus (genus Tritimovirus). J Gen Virol 83:443–450
9. Denizhan E, Szydlo W, Skorack A (2013) Eriophyoid studies in Turkey: review and perspectives. Biological Lett 50(1): 45–54
10. Dwyer GI, Gibbs MJ, Gibbs AJ, Jones RAC (2007) Wheat streak mosaic virus in Australia: relationship to isolates from the Pacific Northwest of the USA and its dispersion via seed transmission. Plant Disease 91:164–170
11. Fellers JP, Webb C, Fellers MC, Shoup Rupp J, De Wolf E (2019) Wheat virus identification within infected tissue using nanopore sequencing technology. Plant Disease 103 2199–2203
12. French R, Stenger DC (2002) Wheat streak mosaic virus. CMI/AAB descriptions of plant viruses No. 398. Wellesbourne, Warwickshire: Association of Applied Biologists
13. Gaborjanyi R, Nagy PD (1988) Wheat streak mosaic virus in Hungary. Novenytermeles 37:391–395
14. Gadiou S., Ku´dela O., Ripl J., Rabenstein F., Kundu JK., Glasa M. 2009. An amino acid deletion in wheat streak mosaic virus capsid protein distinguishes a homogeneous group of European isolates and facilitates their specific detection. Plant Disease 93:1209–1213.
15. Gibson WW, Painter RH (1957) Transportation by aphids of the wheat curl mite, Aceria tulipae (K.), a vector of the wheat streak mosaic virus. Journal of Kansas Entomological Society 30: 147–153

16. Glinushkin AP, Beloshapkina OO, Vinogradov SV, Nikolaev NA (2013) Monitoring of virus symptoms in winter wheat variety sample from the collection of All Russian Institute of Plant Industry named after N.I. Vavilov. Vestnik Orel Gau 2:11–16
17. Hadi BAR, Langham MAC, Osborne L, Tilmon KJ (2011) Wheat streak mosaic virus on wheat: biology and management. J Integr Pest Manag 2:1–5
18. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symposium Series 41:95–98
19. Heun M, Scha¨fer-Pregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F (1997) Site of einkorn wheat domestication identified by DNA fingerprinting. Science 278:1312–1314.
20. Hunger RL (2010) Wheat streak mosaic virus. In: Bockus W, Bowden RL, Hunger RL, Morril WL, Murray TD, Smiley RW (eds) Compendium of wheat diseases and pests, 3rd ed. APS Press, St. Paul, pp 115–117.
21. Ilbagı H, Yorgancı Ü, Cıtır A, Rabenstein F (2003) Occurrence of wheat streak mosaic virus (WSMV) on cereals in Trakya region in Turkey. J Turk Phytopathol 32:81–89
22. Jones RAC, Coutts BA, Mackie AE, Dwyer GI (2005) Seed transmission of wheat streak mosaic virus shown unequivocally in wheat. Plant Disease 89:1048–1050
23. Jones RAC, Vazquez-Iglesias I, Hajizadeh M, McGreig S, Fox A, Gibbs AJ (2022) Phylogenetics and evolution of wheat streak mosaic virus: Its global origin and the source of the Australian epidemic. Plant pathology 71:1660–1673
24. Kılıç Çulal H, Altındal D, Yardımcı N, Akgün I (2012) Detection of wheat streak mosaic virus and barley stripe mosaic virus by DAS-ELISA methods in seeds of different wheat cultivars. U Ü Agricultural Journal 26:17–25
25. Kim D, Paggi J, Salzberg SL (2017) HISAT-genotype: Next generation genomic analysis platform on a personal computer. BioRxiv. https://dx.doi.org/10.1101/266197
26. Kudela O, Kudelova M, Novakova S, Glasa M (2008) First report of wheat streak mosaic in Slovakia. Plant Disease 92:365
27. Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis. Pennsylvania State University, University Park, PA.
28. Kumssa TT, Rupp JS, Fellers MC, Fellers JP, Zhang G (2019) An isolate of wheat streak mosaic *virus* from foxtail overcomes Wsm2 resistance in wheat. Plant Pathology 68:783–789
29. Lanoiselet VM, Hind-Lanoiselet TL, Murray GM (2008) Studies on the seed transmission of wheat streak mosaic virus. Australasian Plant Pathology 37:584–588
30. Liao Y, Smyth GK, Shi W (2014) Feature counts: an efficient general-purpose program for assigning sequence reads to genomic features. Bioinformatics 30 (7): 923–930
31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome project data processing subgroup (2009) The sequence alignment/map format and sAMtools. Bioinformatics Applications Note 25 (16):2078–2079

32. Liao CS, Lu K, Baym M, Singh R, Berger B (2009) IsoRankN: spectral methods for global alignment of multiple protein networks. Bioinformatics 25: i253–i258
33. McKinney HH (1937) Mosaic diseases of wheat and related cereals. Circular No 442. US Department of Agriculture, Washington DC, pp 1–23
34. McNeil JE, French R, Hein GL, Baenziger PS, Eskridge KM (1996) Characterization of genetic variability among natural populations of wheat streak mosaic virus. Phytopathology 86:1222–122
35. McKinney HH (1937) Mosaic Diseases of Wheat and Related Cereals. US Department of Agriculture Circular No. 442:1–23
36. Mishchenko LT, Dunich AA, Ya-Skrypkina I, Kozub NO (2019) Phylogenetic analysis of two Ukrainian isolates of Wheat streak mosaic virus. Viruses and Cell 35(1):64–77
37. Pop I (1961) Die Strichelvirose des Weizens in der Ruma¨nischen Volksrepublik. J Phytopathol 43:325–336
38. Rabenstein F, Seifers DL, Schubert J, French R, Stenger DC (2002) Phylogenetic relationships, strain diversity and biogeography of tritimoviruses. J Gen Virol 83:895–906
39. Redila CD, Phipps S, Nouri S (2021) Full genome evolutionary studies of wheat streak mosaic-associated viruses using high throughput sequencing. Frontiers in Microbiology 12.
40. Robinson MD, Murray TD (2013) Genetic variation of Wheat streak mosaic virus in the United States Pacific Northwest. Phytopathology 103:98–104
41. Sanchez-Sanchez H, Henry M, Ca!rdenas-Soriano E, Alviso Villasana H (2001) Identification of Wheat streak mosaic virus and its vector *Aceria tosichella* Keifer in Mexico. Plant Disease 85:13–17
42. Schubert J, Ziegler A, Rabenstein F (2015) First detection of Wheat streak mosaic virus in Germany: molecular and biological characteristics. Arch Virol 160:1761–1766
43. Seifers DL, Harvey TL, Kofoid KD, Stegmeier WD (1996) Natural infection of pearl millet and sorghum by wheat streak mosaic virus in Kansas. Plant Disease 80:179–185
44. Seifers DL, Harvey TL, Martin TJ, Jensen SG (1997) Identification of the wheat curl mite as the vector of the high plains virus of corn and wheat. Plant Disease 81:1161–1166
45. Singh K, Kund JK (2017) Variations in Wheat streak mosaic virus coat protein sequence among crop and non-crop hosts. Crop and Pasture Sci 68:328–336
46. Singh K, Wegulo SN, Skoracka A, Kundu JK (2018) Wheat streak mosaic virus: a century-old virus with rising importance worldwide: wheat streak mosaic virus. Mol Plant Pathol 19:2193–2206
47. Slykhuis JT (1955) Aceria tulipae Keifer (Acrina: Eriphidae) in relation to the spread of wheat streak mosaic. Phytopathology 45:116–128
48. Stenger DC, Hall JS, Choi IR, French R (1998) Phylogenetic relationships within the family Potyviridae: wheat streak mosaic virus and brome streak mosaic virus are not members of the genus Rymovirus. Phytopathology 88: 782–787
49. Stenger DC, Seifers DL, French R (20029 Patterns of polymorphism in Wheat streak mosaic virus: sequence space explored by a clade of closely related viral genotypes rivals that between the most

divergent strains. Virology 302:58–70

50. Stenger DC, French R (20099 Wheat streak mosaic virus genotypes introduced to Argentina are closely related to isolates from the American Pacific Northwest and Australia. Arch Virol 154:331–336
51. Truol G, French R, Sagadin M, Ameodo J (2004) First report of wheat streak mosaic virus infecting wheat in Argentina. Australasian Plant Pathology 33:137–138
52. Trzmiel K, Szydło W (2012) Occurrence and characteristic of wheat streak mosaic virus (WSMV) in maize in Poland. Post-Epy Wochronıe Ros´Lin 52:707–713
53. Trzmiel K, Szydło W, Hasiów-Jaroszewska B (2021) Biological and molecular characterization of the two polish wheat streak mosaic virus isolates and their transmission by wheat curl mites. Plant Protection Science 57 (3):171–178

Table

Table 1. The list of the complete nucleotide sequences of wheat streak mosaic virus isolates has been analyzed in this study

Isolate name	Country origin	GenBank No.	References
S34Edirne	Turkey	MZ405098	in this study
S34-Edirne	Turkey	ON364116	in this study
Turkey1	Turkey	AF454455	(Rabenstein et al.,2002)
Sidney81	Nebraska,USA	AF057533	(Stenger et al.,1998)
COPhil	Colorado,USA	MT762109	(Albrecht et al.,2022)
KSH294	Kansas,USA	MF459661	(Kumssa et al.,2018)
WA94	Washington,USA	FJ348358	(Schubert et al.,2015)
KSWa12017	Kansas,USA	MK318281	(Fellers et al.,2017)
Arg2	Argentina	FJ348359	(Stenger et al.,2002)
El Batan 3	Mexico	AF285170	(Choi et al.,2001)
Czech	Czech Republic	AF454454	(Rabenstein et al.,2002)
Saadat-Shahr	Iran	EU914918	(Masumi et al.,2008)
Naghadeh	Iran	EU914917	(Masumi et al.,2008)
Marmagne	France	HG810953	(Schubert et al.,2015)
Hoym	Germany	HG810954	(Schubert et al.,2015)
Austria	Austria	LN624217	(Schubert et al.,2015)
Sze	Poland	MH939145	(Trzmiel and Szydlo,2021)
DSMZ PV-0356	Ukraine	MZ202336	(Mishchenko et al.,2019)
NE01-19	Kansas,USA	MW990167	(Redila et al.,2021)
KM-19	Kansas,USA	MW990168	(Redila et al.,2021)
Sosn	Poland	MH939146	(Trzmiel and Szydlo,2021)
RO20	Kansas,USA	MW990169	(Redila et al.,2021)
RA02-19c	Kansas,USA	MW990175	(Redila et al.,2021)
Type-PV57	Nebraska,USA	AF285169	(Choi et al.,2001)
Mon96	Nebraska,USA	AF511630	(Stenger et al.,2002)
BCWS5	Australia	OK181891	(Jones et al.,2022)
BCHPV1	Australia	OK181888	(Jones et al.,2022)

Figures

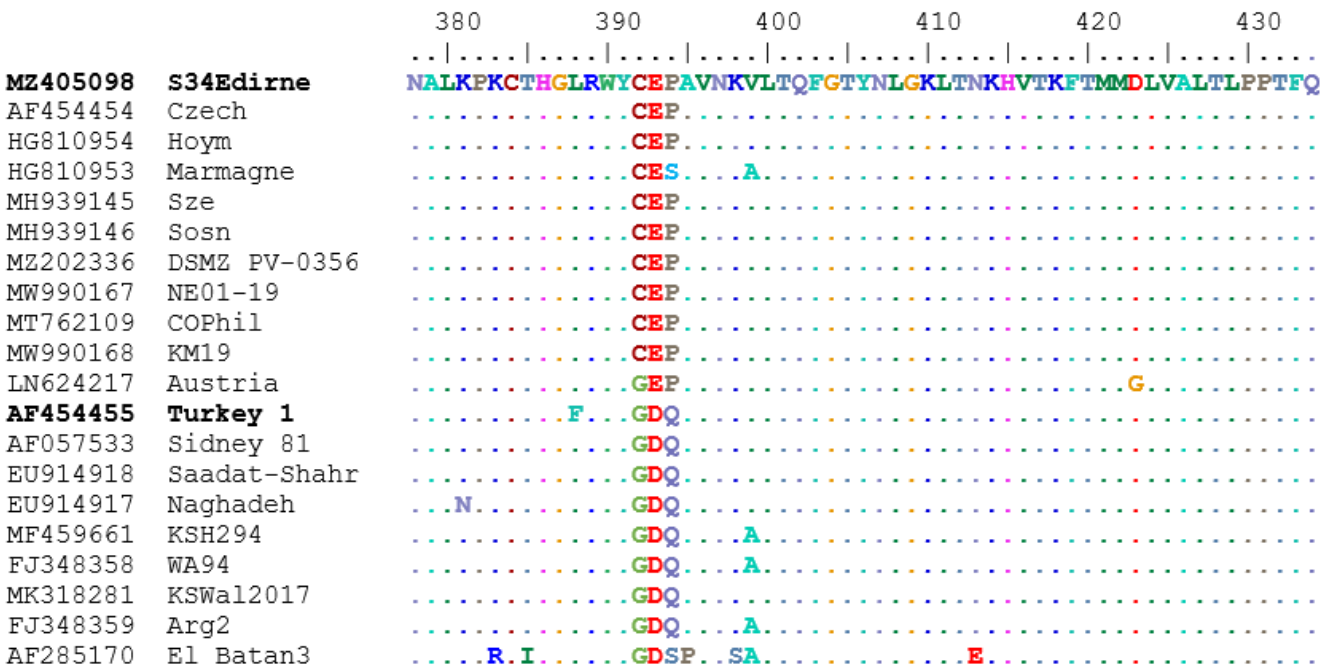

Figure 1

Putative protease cleavage sites in WSMV isolates within clades A, B, and D. The P1/HCPro cleavage site of European isolates differ from other American/Asian WSMV isolates, and Mexican isolate El-Batan3 is the most divergent isolate among the WSMV isolates

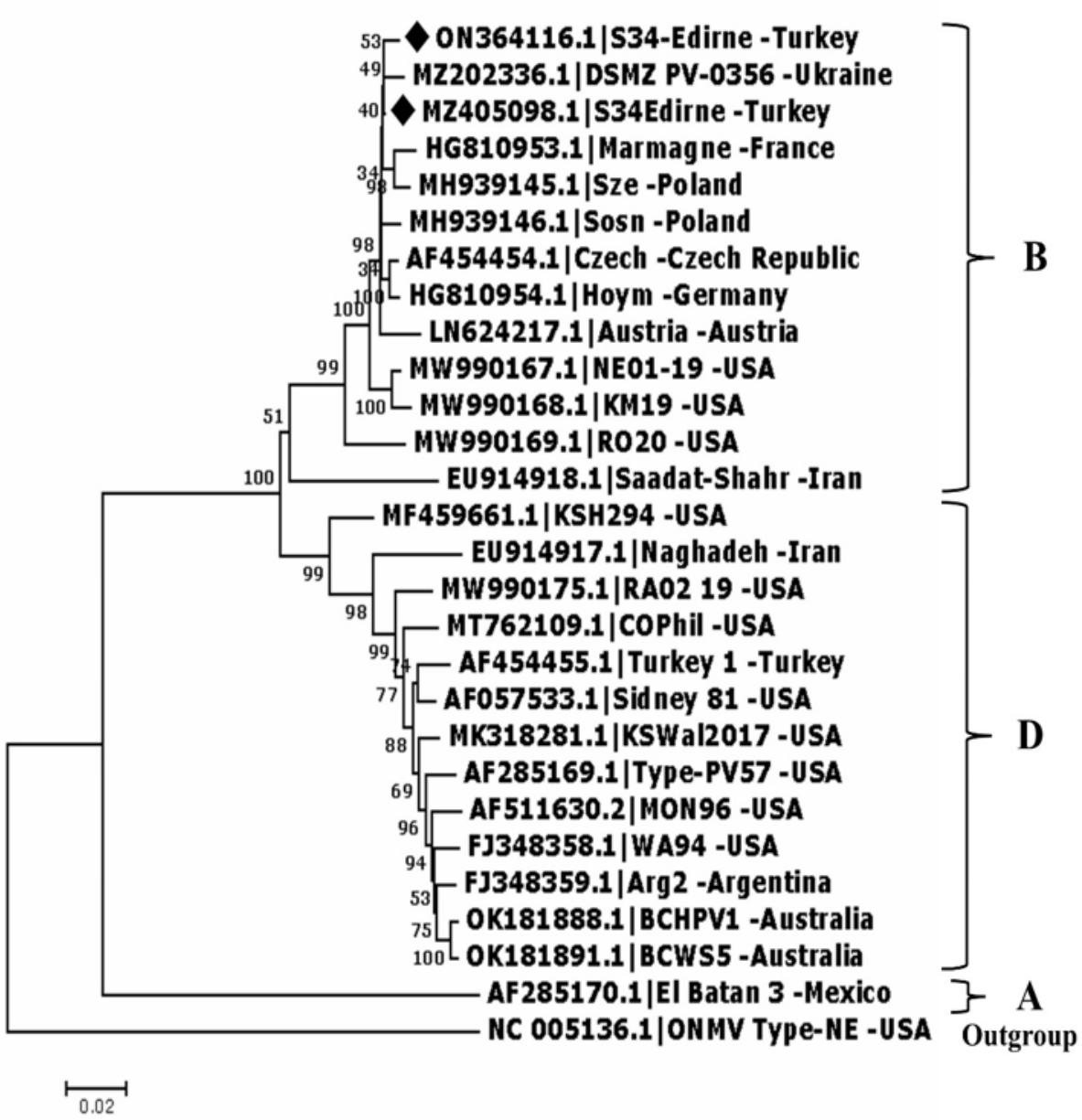

Figure 2

Phylogenetic tree constructed using the neighbor-joining method with nucleotide sequences of 27 wheat streak mosaic virus (WSMV) isolates. 1000 bootstrap replicates were performed and rooted to the genome of oat necrotic mottle virus (ONMV)

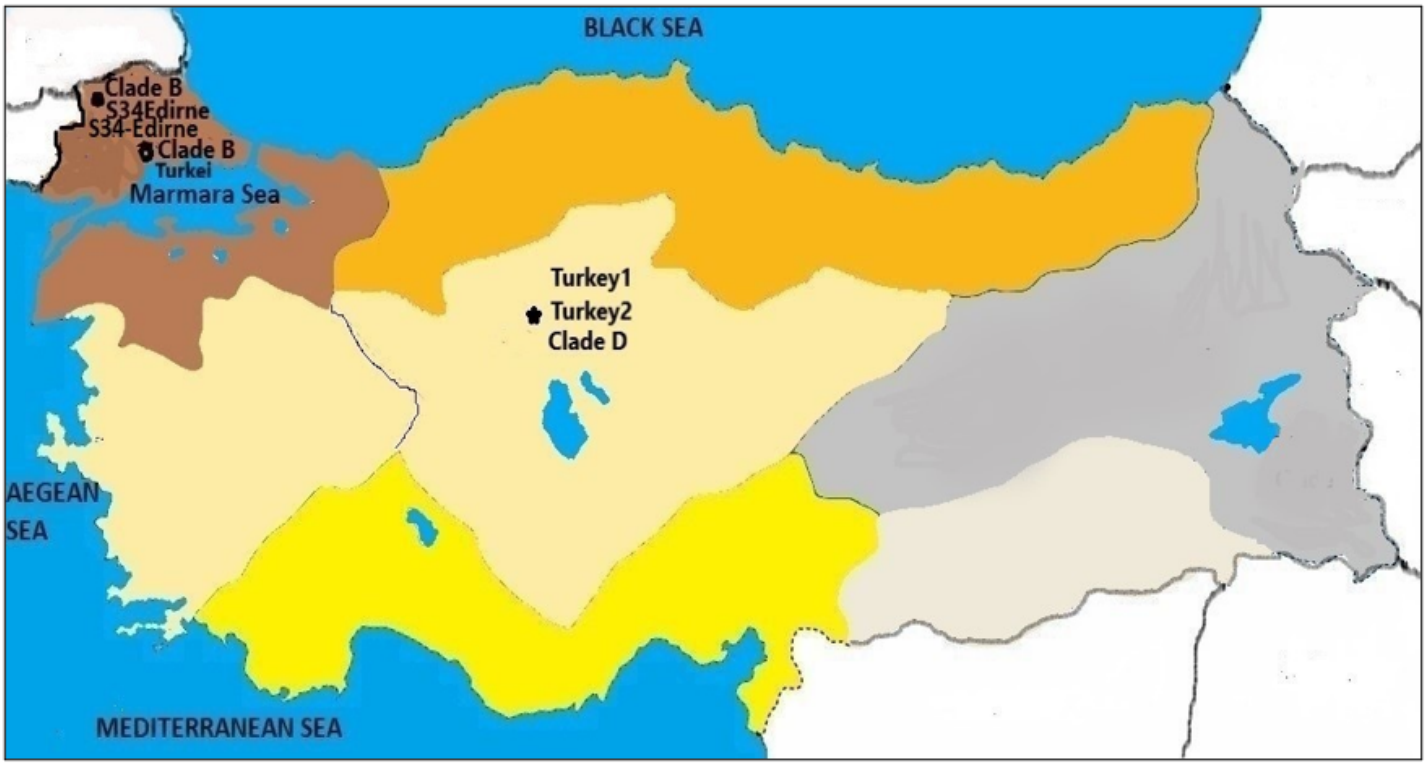

Figure 3

Four provinces in Turkey where WSMV isolates were procured (S34Edirne, S34-Edirne, Turkei, Turkey 1, Turkey 2 isolates)