

Genomic and biological characterization of chiltepín yellow mosaic virus, a new tymovirus infecting *Capsicum annuum* var. *aviculare* in Mexico

Israel Pagán · Mónica Betancourt · Jacinto de Miguel · Daniel Piñero · Aurora Fraile · Fernando García-Arenal

Abstract The characterization of viruses infecting wild plants is a key step towards understanding the ecology of plant viruses. In this work, the complete genomic nucleotide sequence of a new tymovirus species infecting chiltepín, the wild ancestor of *Capsicum annuum* pepper crops, in Mexico was determined, and its host range has been explored. The genome of 6,517 nucleotides has the three open reading frames described for tymoviruses, putatively encoding an RNA-dependent RNA polymerase, a movement protein and a coat protein. The 5' and 3' untranslated regions have structures with typical signatures of the tymoviruses. Phylogenetic analyses revealed that this new virus is closely related to the other tymoviruses isolated from solanaceous plants. Its host range is mainly limited to solanaceous species, which notably include cultivated *Capsicum* species. In the latter, infection resulted in a severe reduction of growth, indicating the potential of this

virus to be a significant crop pathogen. The name of chiltepín yellow mosaic virus (ChiYMV) is proposed for this new tymovirus.

Introduction

Bird pepper [*Capsicum annuum* var. *aviculare* (Dierbach) D'Arcy & Eshbaugh, syn. *C. annuum* var. *glabriusculum* (Dunal) Heiser & Pikckergill] is the closest wild relative and the ancestor of chilli pepper (*Capsicum annuum* L.) [1], with a distribution range extending from the south of the United States to the north of South America [2, 3]. This perennial shrub produces small red berries consumed by frugivorous birds and humans, which are the most important dispersal agents [4]. In Mexico, *C. annuum aviculare*, known as “chiltepín” or “chile piquín”, is harvested from natural and cultivated populations as a wild spice, and is also used as a traditional medicine [3]. Chiltepín production in Mexico has been estimated as 50 tons/year, having great importance for subsistence farmers of the central and northern regions of the country [3, 5, 6]. The agronomic interest of chiltepín exceeds its value as a local commodity, as it is genetically compatible with the domesticated varieties of *Capsicum annuum*, and wild *Capsicum* species are important sources of genetic diversity for breeding programs of cultivated pepper, particularly as sources of resistance to pathogens, including viruses in the taxa *Tobamovirus*, *Potyviridae* and *Geminiviridae* [7–9].

Viral diseases have a strong economic impact on agricultural populations of chiltepín, and chiltepín-infecting viruses can infect related plant species, being a potential threat for important crops cultivated in the same areas as chilli pepper [5, 10]. In addition, the presence of resistance factors in chiltepín [9] suggests that viruses may play a role

as ecological factors shaping chiltepin evolutionary dynamics in natural populations. Thus, the characterization of chiltepin-infecting viruses may have a double interest: from an agricultural point of view it may help to design strategies aimed at controlling existing viral diseases and to anticipate the emergence of new viruses, and from an ecological point of view it may contribute to understand the role of viruses in the population dynamics of wild host plants.

Here we report the biological and genomic characterization of a new tymovirus species found infecting wild chiltepin populations during surveys carried out in Mexico in 2001. Tymoviruses (family *Tymoviridae*) have isometric particles, which encapsidate a positive-sense single stranded RNA genome with three open reading frames (ORF). ORF 1, almost entirely overlapping with ORF 2, encodes a 69 kDa protein necessary for cell-to-cell movement and involved in suppression of RNA silencing (OP) [11, 12]. ORF 2 is the longest one, and encodes a protein of 206 kDa with various conserved domains found in RNA-dependent RNA-polymerases (RP). At the 3' end of this gene there is the so-called tymobox, a 16 nucleotides domain highly conserved among tymoviruses, which is thought to be the promoter for the transcription of the subgenomic mRNA for ORF 3 [13–15] that encodes the 20 kDa coat protein (CP). Tymoviral genomes have typical secondary structures at the 5' and 3' untranslated regions (UTR). The 5'-UTR can be folded in a series of hairpins, whose number varies depending on the species, and the 3'-UTR displays a tRNA-like structure [16]. Tymoviruses are horizontally transmitted by beetles in a semi-persistent manner, and vertical transmission through seed is relatively infrequent, having been reported only for some species in the genus. Species of the genus *Tymovirus* have been isolated from many dicotyledonous plants, including some of the agronomically most important botanical families, such as *Cucurbitaceae*, *Brassicaceae* and *Solanaceae*, and occasionally may be significant as crop pathogens [12].

The experimental host range and the nucleotide sequence of two isolates of a new tymovirus species, for which the name chiltepin yellow mosaic virus (ChiYMV) is proposed, have been determined. The experimental host range predominantly includes species of the *Solanaceae*, showing different virulence levels depending on the virus-host genotype \times genotype interactions. The nucleotide sequence of the genomic RNA of the two isolates of this new species, and subsequent data analysis indicate that the genome is similar in size and organization to the other species of the genus *Tymovirus*. In addition, phylogenetic analyses show that this new virus species is closely related to the other tymoviruses isolated from solanaceous species.

Materials and methods

Virus isolates

A virus isolate is defined here as the virus preparation obtained from one field infected plant. Chiltepin plants showing symptoms of a yellow mosaic (Fig. 1) from wild populations sampled in Tula (Tamaulipas, Mexico, 22°59'N, 99°38'W) were used to inoculate pepper plants of cv. Doux des Landes, which has no resistance to any known virus, and plants of *Nicotiana benthamiana*. To do so, sap extracts from infected field plants were inoculated in 0.01 M phosphate buffer pH 7, 0.2% sodium diethyldithiocarbamate, and symptoms were monitored. Pepper plants inoculated with field plants identified as 20.5, and 20.8 developed a systemic mosaic, and were the source for the isolation and characterisation of isolates Tula 20.5 and Tula 20.8. Tymovirus-like virions were purified as described by Hord et al. [17], and disrupted in 4% sodium dodecyl sulfate, 0.1 M Tris-HCl, pH 9.0, with bentonite at 1 mg/ml, and virion RNA extracted with phenol-chloroform. Virion RNA was used for the biological and genetic characterization of the two isolates.

Fig. 1 Yellow mosaic symptoms in leaves of wild chiltepin infected with virus isolate 20.5 collected from Tula, Tamaulipas, Mexico



Biological characterization of isolates Tula 20.5 and Tula 20.8

To determine the host range and the symptoms induced by Tula 20.5 and Tula 20.8 infection, different species of the botanical families *Amaranthaceae*, *Asteraceae*, *Brassicaceae*, *Chenopodiaceae*, *Cucurbitaceae*, *Fabaceae*, *Lamiaceae*, and *Solanaceae* were inoculated with each of the

Tula 20.5 and Tula 20.8 isolates (Table 1). For each isolate, five to seven plants per species (10–15 days old) were mechanically inoculated by applying purified virion RNA (100 ng/μl) in 0.1 M Na₂HPO₄ onto the first two completely expanded leaves dusted with carborundum. The inoculated plants were maintained in a greenhouse (20–25°C, and 16 h of light), and symptoms were recorded over a 4-week period. Viral infection was also analysed by

Table 1 Host range and symptoms of two Tula isolates

Plant species	Systemic infection ^a isolate		Systemic symptoms
	20.5	20.8	
<i>Chenopodiaceae</i>			
<i>Chenopodium amaranticolor</i>	0/7	0/7	–
<i>Chenopodium quinoa</i>	0/7	0/7	–
<i>Amaranthaceae</i>			
<i>Gomphrena globosa</i>	2/6	2/6	Asymptomatic
<i>Brassicaceae</i>			
<i>Brassica campestris</i>	0/7	0/7	–
<i>Asteraceae</i>			
<i>Lactuca sativa</i>	0/7	0/7	–
<i>Cucurbitaceae</i>			
<i>Cucumis melo</i>	0/6	0/6	–
<i>Cucumis sativus</i>	0/7	0/7	–
<i>Fabaceae</i>			
<i>Phaseolus vulgaris</i>	0/5	0/5	–
<i>Vicia faba</i>	0/6	0/6	–
<i>Vigna unguiculata</i>	0/7	0/7	–
<i>Lamiaceae</i>			
<i>Ocimum basilicum</i>	0/7	0/7	–
<i>Solanaceae</i>			
<i>Capsicum annuum</i>			
<i>annuum</i> (Doux des Landes)	3/7	7/7	Mosaic, leaf curl
<i>annuum</i> (Dulce italiano)	7/7	7/7	Mosaic, leaf curl
<i>aviculare</i> (San Luis)	7/7	7/7	Yellow mosaic
<i>aviculare</i> (Xilitla)	0/7	1/7	Yellow mosaic
<i>Capsicum chinense</i>	7/7	7/7	Mosaic, stunting
<i>Capsicum frutescens</i> (Tabasco)	7/7	7/7	Mosaic, stunting
<i>Datura stramonium</i>	5/5	5/5	Mosaic, leaf curl
<i>Nicotiana benthamiana</i>	7/7	7/7	Interveinal chlorosis, leaf curl
<i>Nicotiana clevelandii</i>	6/7	7/7	Leaf curl, stunting
<i>Nicotiana glutinosa</i>	1/7	3/7	Mosaic, stunting
<i>Nicotiana tabacum</i> (Samsun)	2/7	5/7	Mottle, stunting
<i>Nicotiana tabacum</i> (Xanthi-nc)	1/7	2/7	Mottle, stunting
<i>Nicotiana rustica</i>	7/7	7/7	Vein clearing, stunting
<i>Physalis floridana</i>	1/7	2/7	Mottle
<i>Solanum chilense</i>	0/7	0/7	–
<i>Solanum lycopersicum</i>	5/6	5/6	Mosaic, stunting
<i>Solanum peruvianum</i>	0/7	1/7	Mosaic

^a Values are number of infected over inoculated plants as determined by dot-blot hybridisation

dot-blot hybridization with a ^{32}P -labelled RNA probe specific for the Tula 20.5 and 20.8 isolates, representing a fragment of the genome including the tymobox and the region 415 nucleotides (nt) upstream of it. This region was amplified from isolate Tula 20.8 using the Superscript III one-step reverse transcription-polymerase chain reaction (RT-PCR) with Platinum Taq kit (Invitrogen, Carlsbad, CA, USA), and the degenerated primers TYMOBOX and TYMO8dir, corresponding to the tymobox and a conserved region located 415 nt upstream the tymobox, respectively (Supplementary Table S1; Fig. 2) and cloned as described below.

Virulence (V) of Tula 20.5 and Tula 20.8 was estimated as the effect of viral infection on host growth, determined as the reciprocal of the ratio of the dry weight of the above-ground parts of each infected plant in relation to the mock-inoculated controls [$V = 1 - (DW_i)/(DW_m)$, where DW_i is the dry weight of the inoculated plant, and DW_m is the average dry weight of five mock-inoculated plants of the same host] [18]. Dry weight was determined by maintaining the harvested plants at 65°C until constant weight.

Differences in host susceptibility and infectivity of isolates Tula 20.5 and Tula 20.8 were analysed by one-way analysis of variance (ANOVA), using either host species or virus isolate as factors, respectively. Differences in virulence according to isolate or to host species were determined by two-way ANOVA using *Capsicum* species and virus isolate as factors in a full factorial model. Differences in the virulence of isolates for each individual *Capsicum* species were assessed using one-way ANOVAs with virus isolate as factor. To determine if values of analysed traits were significantly different among classes within each factor least significant difference (LSD) analyses were employed in all cases. Host species and virus isolate were considered as random effect factors.

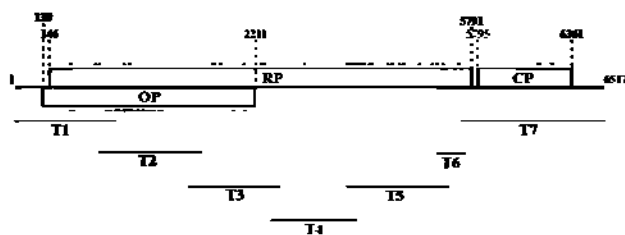


Fig. 2 Genome organization of Tula 20.5 and 20.8 isolates and position of overlapping cDNA clones T1–T7 used for sequence determination. Three major ORFs are indicated by grey boxes with nucleotide positions of the start and stop codons, respectively. Numbers correspond to nucleotide positions in the genome of the two isolates. Asterisks denote nucleotide differences between the two isolates resulting in amino acid changes

Genome sequencing and nucleotide analysis

The complete nucleotide sequence of the genomic RNA was determined using virion RNA as template for RT-PCR using different sets of primers. Primer pairs were designed to produce five internal and two terminal fragments in such a way that adjacent fragments overlapped by at least 135 nt. The five internal fragments were obtained using primers TYMO1dir, and TYMO3dir to TYMOBOX (see Supplementary Table S1), which largely covered the non-overlapping region of ORF 2. The 5' end fragment, corresponding to the 5'-UTR and the overlapping region of ORF 1 and ORF 2, was amplified using SMART PCR cDNA synthesis kit (Clontech, Mountain View, CA, USA), and the designed primer TYMO2rev. The 3' end, corresponding to ORF 3 and 3'-UTR, was obtained adding adenosine monophosphates to the 3' hydroxyl termini of RNA molecules with poly(A) polymerase (GE Healthcare, Chalfont St. Giles, UK) [19], and then was amplified using an oligo(dT) primer and the primer TYMO9dir. Primer sequence and their location in Tula 20.5 and Tula 20.8 genome are summarized in Supplementary Table S1 and Fig. 2. Amplified fragments were gel purified using QIAquick Gel Extraction kit (Qiagen, Valencia, CA, USA), and then cloned using the pCR II Kit (Invitrogen). *E. coli* JM109 (Invitrogen) were transformed with the ligated vector, and plasmid DNA was isolated using Miniprep Express Matrix (Q-BIOgene, Montreal, QC, Canada).

At least three clones per fragment were sequenced and aligned using Muscle 3.7 (<http://www.drive5.com/muscle/>, [20]), revealing 99–100% nucleotide identity in all cases. Similar nucleotide identity was found in overlapping regions of adjacent fragments. Hence, the complete nucleotide sequences of isolates Tula 20.5 and Tula 20.8 were determined by assembling the seven fragments using Se-AL Carbon *v2.0a11* (<http://tree.bio.ed.ac.uk/software/>, [21]) (Fig. 2). Sequences determined in this work can be found in the EMBL database under accession codes: FN563123 (Tula 20.5) and FN563124 (Tula 20.8).

Phylogenetic analyses

Phylogenetic relationships between Tula 20.5 and Tula 20.8, and the other members of the genus *Tymovirus* were analysed using the entire genomic sequences, and the three open reading frames separately. Only those species for which full genome sequences were available were used. Genomic sequences of *Anagris vein yellowing virus* (AVYV, acc. No.: NC_011559), *Chayote mosaic virus* (ChMV, acc. No.: NC_002588), *Diascia yellow mottle virus* (DiaYMV, acc. No.: NC_011086), *Dulcamara mottle virus* (DuMV, acc. No.: NC_007609), *Eggplant mosaic virus* (EMV, acc. No.: NC_001480), *Erysimum latent virus* (ErLV, acc. No.:

NC_001977), *Kennedya yellow mosaic virus* (KYMV, acc. No.: NC_001746), *Nemisia ring necrosis virus* (NeRNV, acc. No.: NC_011538), *Okra mosaic virus* (OkMV, acc. No.: NC_009532), *Physalis mottle virus* (PhyMV, acc. No.: NC_003634), *Plantago mottle virus* (PIMoV, acc. No.: NC_011539), *Ononis yellow mosaic virus* (OYMV, acc. No.: NC_001513), *Turnip yellow mosaic virus* (TYMV, acc. No.: NC_004063), and *Scrophularia mottle virus* (SrMV, acc. No.: NC_011537), were aligned with the sequences of Tula 20.5 and Tula 20.8 using Muscle 3.7. Best substitution model was selected using Akaike Information Criterion (AIC) implemented in Modeltest 3.7 [22], and phylogenetic trees were estimated by the neighbor-joining and maximum likelihood procedures using PAUP* version 4.10b (Sinauer Associates, Sunderland, MA, USA). Primary branching patterns (tree topologies) obtained using the two methods were similar. Analyses for each gene were done using the nucleotide and the amino acid sequences, resulting also in similar topologies. To simplify, here we only show the trees obtained by the neighbor-joining method using the nucleotide sequences.

Results

Isolation of a tymovirus from wild chiltepin plants

Pepper plants (cv. Doux des Landes) inoculated with sap from leaves of chiltepin plants showing a yellow mosaic in their native habitat at Tula (see “Materials and methods”) showed a systemic mosaic starting 14 dpi. These plants did not react in ELISA with antisera against *Cucumber mosaic virus* (CMV, Sediag, France) and *Tomato spotted wilt virus* (Loewe Phytodiagnostica, Germany), and against a monoclonal antibody specific for viruses of the genus *Potyvirus* (Agdia, USA), which commonly infect pepper. Plants were inoculated at an inoculum dilution that resulted in 0/10 *N. benthamiana* plants infected, and 1/10 pepper plants infected. In pepper plants inoculated with sap from field plants 20.5 and 20.8, dsRNA preparations [23] showed a single band migrating slightly ahead of dsRNA from plants infected with the tobamovirus *Tobacco mild green mosaic virus*, suggesting the presence of a virus with a ssRNA genome of 6,000–6,500 nt (not shown). Virions were purified from these plants following the procedure described for CMV by Lot et al. [24], and electron microscopy of virion preparations negatively stained with phosphotungstic acid [25] showed isometric particles of about 30 nm in diameter (not shown). These particles were built of a single CP species with an apparent molecular mass of 20 kDa according to its electrophoretic mobility in SDS polyacrylamide gels after virion disruption in 2% SDS at 100°C, no other protein band being observed (not

shown). Agarose gel electrophoresis of nucleic acid extracts from these particles showed a single band of ssRNA of about 6,000 nt which hybridised with a 5'-³²P-labelled oligonucleotide probe corresponding to the tymobox sequence (not shown). These results indicated that a virus isolate was transferred from each chiltepin plant to pepper, and that these isolates belonged to the genus *Tymovirus*. These isolates were named Tula 20.5 and Tula 20.8, and their virions were purified for further characterisation according to Hord et al. [17] procedure.

Host range and virulence of isolates Tula 20.5 and Tula 20.8

The host ranges of Tula 20.5 and Tula 20.8 were determined using 28 plant species and cultivars inoculated at similar ages, belonging to eight botanical families (see “Materials and methods”); the results are shown in Table 1. Both isolates were indistinguishable in host range and symptomatology (Table 1): with the exception of *Gomphrena globosa* (*Amaranthaceae*), all species that were systemically infected belonged to the family *Solanaceae*, and showed leaf mottle and mosaic, with occasional leaf lamina distortions, which in most species lead to a reduced growth and stunting. No local lesion host or hosts with infection restricted to the inoculated leaves were identified. Within the family *Solanaceae*, species varied in susceptibility ($F_{1,33} = 18.72$, $P = 1 \times 10^{-5}$), *Solanum chilense* and *S. peruvianum* being the more resistant ones, followed by *Nicotiana glutinosa* and *N. tabaccum*. *Capsicum* species were fully susceptible to infection, but it is noteworthy that plants derived from seeds of two populations of the original host, chiltepin (San Luis and Xilitla), varied largely in susceptibility. Differences in infectivity between isolates Tula 20.5 and Tula 20.8 were not significant ($F_{1,35} = 0.48$, $P = 0.493$), except for *C. annuum* Doux des Landes and for *N. tabaccum* Samsun ($F_{1,61} = 9.45$, $P = 0.016$).

A two-way ANOVA, using isolate and host species as factors, showed that on average virulence did not differ between Tula 20.5 and Tula 20.8 ($F_{1,61} = 1.11$, $P = 0.296$), but significantly differed among *Capsicum* species or cultivars ($F_{4,61} = 7.20$, $P = 1 \times 10^{-4}$), being higher for *C. chacoense* and *C. chinense* than for *C. annuum* Douce Italiano and *C. frutescens* Tabasco, with an intermediate level for *C. annuum* Doux des Landes. The interaction between both factors was also significant ($F_{4,61} = 4.26$, $P = 0.005$). Consequently, virulence of the two viral isolates was analysed on each plant species individually. One-way ANOVAs indicated that in *C. chacoense* and *C. frutescens* Tabasco, virulence was significantly higher for Tula 20.5 than for Tula 20.8, the converse occurred in *C. annuum* Doux des Landes, and no

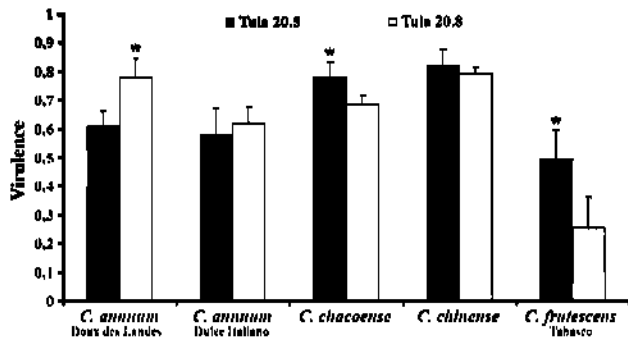


Fig. 3 Virulence of isolates Tula 20.5 and Tula 20.8 on five *Capsicum* cultivars or species. Virulence (V) is estimated as the effect of virus infection on plant growth, determined as the reciprocal of the ratio of the dry weight of the above-ground parts of each infected plant in relation to the mean value of mock-inoculated controls. Data are mean \pm standard errors of seven replicates for Tula 20.5 (black), and Tula 20.8 (grey). Asterisks denote statistically significant differences between isolates

differences were observed in *C. chinense* and *C. annuum* Dolce Italiano (Fig. 3). Hence, virulence level of Tula isolates in *Capsicum* species depends on virus genotype \times plant genotype interactions.

Organization of Tula 20.5 and Tula 20.8 genomic RNA

The complete nucleotide sequence of the genomic RNA (gRNA) was determined for isolates Tula 20.5 and Tula 20.8, and shown to consist of 6,517 nt for both isolates (Fig. 2). The gRNA of these isolates has all the features found in the other tymoviruses fully sequenced so far. The base composition has a small guanine content and high proportion of cytosine residues (24% A, 26% U, 35% C, 15% G, for both isolates). The genome contains three major ORFs, with similar size and arrangement as in other species of the genus, and the 5'- and 3'-UTRs have the characteristic secondary structures of tymoviruses.

The deduced amino acid (aa) sequences of the three main putative genes found in the positive strand of the gRNA correspond to the OP (ORF 1), RP (ORF 2) and CP (ORF 3) genes as described for other tymoviruses. The longest ORF starts at nucleotide 146 (first nucleotide of the ATG codon) and ends at position 5791 (last nucleotide of the TGA codon). This gene encodes for the RP protein, involved in viral replication, comprising 1,881 aa, and is the second largest RP among the classified species in the genus. The predicted translation product of the RP of Tula 20.5 and Tula 20.8 contains the motifs typically found in tymoviral polyproteins: the methyltransferase domains (aa 65–220), the papaine-like protease domain (aa 789–884), the helicase domain (aa 983–1216), and the RNA-dependent RNA polymerase domains (aa 1540–1766) [26, 27]. Finally, the 3' terminal conserved domain known as the

tymobox, 16 nt in length, is located between nucleotides 5762 and 5777. Consistent with other tymoviruses, seven nucleotides upstream of the RP start codon is located the ATG of the OP gene, which almost completely overlaps the RP gene. The OP gene starts at nucleotide 139 and ends at position 2211, the predicted translation product being 690 aa in length, one of the largest among the tymoviruses. Amino acids 239–242 form the tetrapeptide GILG, found in all known members of the genus [28]. Finally, three nucleotides downstream of the stop codon of the RP the CP gene starts, extending from nucleotides 5795–6361. The predicted protein has 188 aa, similar to most tymoviruses (Fig. 2), and in agreement with the apparent molecular mass of 20 kDa determined by SDS-PAGE.

The ability to form secondary structures typical of tymoviruses was found in the non-coding regions of both isolates. The 5'-UTR of the gRNA, 138 nt in length, has a high content of U and C, and can be folded into a series of characteristic hairpins containing C–C and C–U mismatches. Like the majority of the previously described tymoviruses, the 5'-UTR of Tula 20.5 and Tula 20.8 has three hairpins, although an extra hairpin has been described for some species [29]. In addition, the 3'-UTR, comprising the 157 terminal nucleotides of the gRNA, has a tRNA-like structure (TLS). The anticodon arm of the TLS contains the CAC anticodon for valine, as in most of the sequenced tymoviruses. In agreement with nucleotide and amino acid sequence identity (see Supplementary Table S2), the Tula 20.5 and Tula 20.8 isolates have an EMV-like TLS [30]. Three upstream stem-loops precede the TLS, the 3' terminal one showing high nucleotide sequence identity and secondary structure similarity with that described for EMV. The predicted secondary structures of the 5'- and 3'-UTRs of Tula 20.5 and Tula 20.8 are shown in Supplementary Fig. S1.

Phylogenetic relationships of Tula 20.5 and Tula 20.8 with other tymovirus species

The complete nucleotide sequence of the gRNA of Tula 20.5 and Tula 20.8, and the three predicted ORFs (RP, OP and CP), were aligned with the corresponding regions of other tymoviruses which had been completely sequenced.

The full-length genomes of the two isolates from Tula shared 96% of nucleotide identity. The most conserved gene between Tula 20.5 and Tula 20.8 is the CP (99%), while the less conserved gene is the RP (95%) (Supplementary Table S2), with most of the amino acid changes located in the C-terminal half of the protein (Fig. 2). However, amino acid identity was lowest in the OP (93%). The percentage of nucleotide identity of the genome of Tula 20.5 and Tula 20.8, and the other tymoviruses ranged from 57 to 65%. Overall, the RP showed the highest

nucleotide and amino acid identities (58–66% and 45–62%, respectively), which was much lower for the OP (39–46% and 7–40%, respectively), with the two exceptions of EMV and ChMV, which showed nucleotide identities in the range of the RP (54 and 60%, respectively). The nucleotide identity in the CP varied depending on the tymovirus species: while most interspecies comparisons showed identities between 47 and 68%, a group of four species showed identities not higher than 31% (ErLV, KYMV, OMY and TYMV), with similar results for the amino acid sequences (Supplementary Table S2).

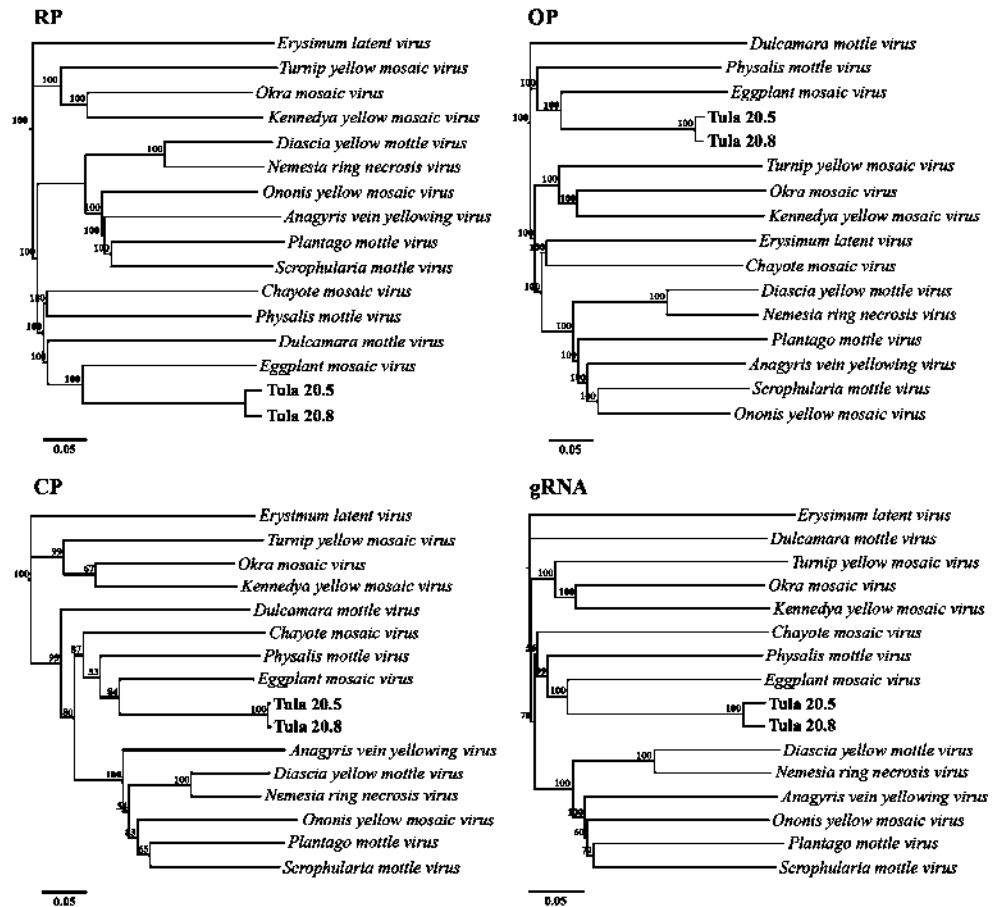
Regardless of differences in nucleotide and amino acid identities between tymoviral genes, phylogenetic trees inferred for each gene and for the full-length genome showed similar topologies (Fig. 4), and divided the species into three groups: one including ErLV, KYMV, OkMV and TYMV; one including all species infecting hosts in the *Solanaceae* and the *Cucurbitaceae* (Tula 20.5, Tula 20.8, ChMV, DuMV, EMV and PhyMV); and one including AVYV, DiaYMV, NeRV, OYMV, PIMoV and SrMV. This division was supported by the significance of the nodes in the phylogenetic trees, and only the inclusion of DuMV in any of the three groups was not well supported in the gRNA tree. The classification of isolates Tula 20.5 and

Tula 20.8 within the group comprising viruses infecting *Solanaceae* and *Cucurbitaceae* was in agreement with the overall higher percentages of nucleotide identity that ChMV, DuMV, EMV and PhyMV shared with these isolates in all the genes, and the lower identity values observed for species of the other groups (Supplementary Table S2; Fig. 4).

Discussion

Information on viruses infecting wild plants is scarce, which is a handicap for analyses of two important aspects of virus ecology: the role of virus infection in the composition and dynamics of plant ecosystems, and the role of wild plants as sources of emergent viruses, or as inoculum sources of epidemic viruses, for related crops [31–33]. Viruses are a major factor limiting the production of *Capsicum* crops, including chiltepin, in Mexico [5], where the domesticated pepper shares areas of distribution with its wild ancestor, *C. annuum aviculare* [3, 5]. Within this frame of long-term research, we present here the characterisation of a new virus infecting wild chiltepin in Northern-Central Mexico. This virus has the particle and

Fig. 4 Neighbor-joining phylogenetic trees based on nucleotide sequence alignments of each of three genes [replication-associated (RP), overlapping (OP), coat protein (CP)] and the genomic RNAs (gRNA) of isolates Tula 20.5, Tula 20.8 (both in *bold*) and 14 other tymoviruses. Numbers at nodes indicate bootstrap percentage based on 1,000 replicates. Scale bar represents 0.05 nucleotide substitutions per site related to branch length



genome structure characteristic of the genus *Tymovirus*, for which the species demarcation criteria include: (1) serological specificity, (2) differences in host range, (3) overall nucleotide sequence identity not higher than 80%, (4) capsid protein sequence identity not higher than 90%, (5) differences in the 3' terminal structure [34]. Although we have not analysed the serological specificity of ChiYMV, the reported characteristics of this virus fits with the other four demarcation criteria to be considered a new species in the *Tymovirus* genus, for which the name chiltepin yellow mosaic virus (ChiYMV) is proposed.

The experimental host range of ChiYMV does not correspond to those reported for any other known *Tymovirus* [35, 36]. As most tymoviruses, ChiYMV has a narrow experimental host range, with a higher susceptibility of species from the family of the natural host [36–38]. No local lesion host was identified. Interestingly, our results indicate that in some domesticated *Capsicum* spp. ChiYMV infection induces severe symptoms and causes a significant reduction in plant growth (Fig. 3), indicating that ChiYMV could be a threat to *Capsicum* crops. ChiYMV is seed-transmitted in the domestic pepper (our unpublished results), which adds to its potential as a crop pathogen and warnings of the risks of long-distance dispersal. We have not surveyed ChiYMV in crops, such as chilli pepper. The potential of ChiYMV to infect crops will be the subject of future work. An interesting finding of the host range assays was the very different susceptibility to ChiYMV of two accessions of chiltepin (compare data for San Luis and Xilitla in Table 1). This suggests that polymorphisms for resistance to ChiYMV infection occur in natural populations of chiltepin, which would be likely given a long-term interaction between the virus and the host plant, and a negative effect of ChiYMV infection on the fitness of chiltepin in its natural habitat.

The genomic organisation of ChiYMV is similar to that of the other species of the genus *Tymovirus*. Comparative analyses of tymoviral RP genes showed that the 5' and 3' thirds of this gene are highly conserved between species [28]. These regions of the RP gene contain the conserved methyltransferase and polymerase functional domains with signatures typical of the tymoviruses, which may explain their lower variability across species due to functional constraints [39]. In TYMV it has been demonstrated that the RP polyprotein is autocatalytically cleaved by the papain-like protease, yielding two replication-associated proteins of 140 kDa, containing the methyltransferase, the protease and the helicase domains, and 66 kDa, comprising the RNA-dependent RNA polymerase domains [40]. One cysteine and one histidine within the RP papain-like domain are the putative active site residues [40, 41]. These two residues were located also in the RP gene of ChiYMV, which strongly suggest that a similar mechanism of

polyprotein processing is operating in this virus. Moreover, these residues have been found in most tymoviruses [28], suggesting that this is a general mechanism for gene expression by the members of the genus. An additional cleavage site in the 140 kDa protein has been recently described in TYMV. This site is located between the protease and the helicase domains and involves residues Ser879 and Gln880 [42]. These residues were not found in the RP gene of ChiYMV. Although this suggests that the 140 kDa proteolytic processing is not a feature of this new tymovirus, the residues involved in the cleavage of this protein might differ between species. Future studies of the RP proteolytic pathways in ChiYMV are needed to clarify this point. The OP gene almost completely overlaps with the RP gene and is in the –1 reading frame relative to the RP, so that synonymous nucleotide changes in the third codon position of the RP result in amino acid changes in the OP. Nucleotide and amino acid identities among tymoviruses is much lower for the OP than for the RP genes, which suggests that the RP drives the evolution of the OP. For the 3' terminal CP gene, overall identity of ChiYMV with other tymovirus showed values always lower than 90% at both the nucleotide and the amino acid levels, and hence fitting another species demarcation criterion (Supplementary Table S2). The CP is expressed through a subgenomic RNA in all tymoviruses, and the 16 nucleotides highly conserved sequence known as tymobox, also found in ChiYMV, has been proposed to be its promoter [43].

Finally, analyses of the secondary structure of ChiYMV 5'- and 3'-UTRs showed high similarity with those of the other sequenced tymoviruses. The 5'-UTR of all the previously described species can be folded in a series of hairpins, three in most species, with protonable internal loops consisting in C–C and C–U mismatches [29, 44]. Important functional roles in encapsidation, translation and viral accumulation have been assigned to these hairpins [45]. This could explain the conservation of the 5'-UTR secondary structures among the species of the genus, albeit their critical role in RNA packaging has been recently challenged [46]. The tymoviral 3'-UTR also has a significant base pairing that can be folded in a tRNA-like structure common to most of them. Upstream of this structure, a variable number of stem loops has been described depending on the species [29, 37, 47, 48]. The same arrangement was found in the ChiYMV 3'-UTR, which showed the typical tRNA structure, but dissimilarities in the upstream stem loops, thus accounting for the last demarcation criterion required for a new species of the genus *Tymovirus*.

Phylogenetic analyses based on the complete genome and the OP, RP and CP nucleotide sequences of ChiYMV and 14 tymovirus species showed almost identical topologies,

indicating that ChiYMV has not evolved through recombination. Similar results have been previously reported for other species of this genus [29, 37, 49–51]. Thus, it is probable that these viruses, including ChiYMV, evolved by independent mutational changes from a common tymovirus ancestor [49, 51]. A notable exception could be the 3'-UTR, for which recombination has been proposed to result in a largely different structure in DuMV as compared with other species [52], which could explain the different clustering of DuMV when individual genes, or the complete genomic RNA, were considered (Fig. 4). In addition, the inferred phylogenies for the three major genes and for the complete genome divided the *Tymovirus* species into three different groups, with all the *Solanaceae*-infecting species, including ChiYMV, clustering together, which has been proposed to reflect their host specificity [29, 37, 53].

In summary, we have characterized a new virus infecting wild chiltepin populations in Mexico. In contrast with other genera of plant-infecting viruses, it seems to be a trait of the genus *Tymovirus* that a high fraction (about one-third) of the currently known species were isolated from wild plants. This suggests that tymoviruses are prevailing in wild plant ecosystems, as substantiated by a recent report on viruses infecting wild plants in the Tallgrass Prairie Preserve in Oklahoma [54]. The role of tymovirus infection in ecosystem webs could be complex, as indicated by one of the few reports of a benefit for a host plant derived from virus infection: the bright symptoms induced by KYMV in the wild legume *Kennedya rubicunda* was associated with a decreased herbivory [55]. The biological and genomic characterisation of ChiYMV reported here opens the door to future studies on the evolutionary ecology of this virus in its wild host and natural habitats.

Acknowledgments This work was in part funded by grants “Interacciones entre virus y sus plantas huésped en poblaciones silvestres de tabaco y pimiento” (Programa de Cooperación Científica con Iberoamérica 1999, Ministerio de Educación y Cultura, Spain) and “Impacto de los patógenos en la conservación de especies amenazadas: aplicación a las poblaciones silvestres de chiltepin en México” (BIOCON05/101, Fundación BBVA, Spain).

References

1. Nuéz F, Gil-Ortega R, Costa J (1996) El cultivo de pimientos, chiles y ajíes. Mundi Prensa, Madrid
2. Hernández-Verdugo S, Luna-Reyes R, Oyama K (2001) Genetic structure and differentiation of wild and domesticated populations of *Capsicum annuum* (*Solanaceae*) from Mexico. *Plant Syst Evol* 226:129–142
3. Votava EJ, Nabham GP, Bosland PW (2002) Genetic diversity and similarity revealed via molecular analysis among and within

- an in situ population and *ex situ* accessions of chiltepin (*Capsicum annuum* var. *glabriusculum*). *Conserv Genetics* 3:123–129
4. Tewksbury JJ, Nabhan GP, Norman D, Suzán H, Tuxill J, Donovan J (1999) In situ conservation of wild chiles and their biotic associates. *Conserv Biol* 13:98–107
5. Torres-Pacheco I, Garzón-Tiznado JA, Brown JK, Becerra-Flora A, Rivera-Bustamante RF (1996) Detection and distribution of geminiviruses in Mexico and the southern United States. *Phytopathology* 86:1186–1192
6. Almaza JG, Maiti RK, Foroughbakhch PR, Cárdenas ML, Núñez-González MA, Moreno-Limón S, Hernández-Piñero JL, Valades MC (2001) Bromatología del Chile piquín (*Capsicum annuum* L. var. *aviculare* (Dierb.) D. & E.). Resúmenes XV Congreso Mexicano de Botánica. Querétaro
7. Boukema RW (1980) Allelism of genes controlling resistance to TMV in *Capsicum*. *Euphytica* 29:433–439
8. Kyle MM, Palloix A (1997) Proposed revision of nomenclature for potyvirus resistance genes in *Capsicum*. *Euphytica* 97:183–188
9. Hernández-Verdugo S, Guevara-González RG, Rivera-Bustamante RF, Oyama K (2001) Screening wild plants of *Capsicum annuum* for resistance to pepper huasteco virus (PHV): presence of viral DNA and differentiation among populations. *Euphytica* 122:31–36
10. Méndez-Lozano J, Torres-Pacheco I, Fauquet CM, Rivera-Bustamante RF (2003) Interactions Between Geminiviruses in a naturally occurring mixture: *Pepper huasteco virus* and *Pepper golden mosaic virus*. *Phytopathology* 93:270–277
11. Bozarth CS, Weiland JJ, Dreher TW (1992) Expression of ORF-69 of *Turnip yellow mosaic virus* is necessary for viral spread in plants. *Virology* 187:124–130
12. Chen J, Li WX, Xie D, Peng JR, Ding SW (2004) Viral virulence protein suppresses RNA silencing-mediated defense but upregulates the role of microRNA in host gene expression. *Plant Cell* 16:1302–1313
13. Ding S, Howe J, Keese P, Mackenzie A, Meek D, Osorio-Keese M, Skotnicki M, Srifah P, Torronen M, Gibbs A (1990) The tymobox, a sequence shared by most tymoviruses: its use in molecular studies of tymoviruses. *Nucleic Acids Res* 18:1181–1187
14. Martelli GP, Sabanadzovic S, Abou-Ghanem Sabanadzovic N, Edwards MC, Dreher T (2002) The family *Tymoviridae*. *Arch Virol* 147:1837–1846
15. Haenni AL, Dreher TW (2008) Tymoviruses. In: Mahy BWJ, van Regenmortel MHV (eds) *Encyclopedia of virology*, vol 5. Academic Press, London, pp 199–207
16. Briand JP, Jonard G, Guilley H, Richards K, Hirth L (1977) Nucleotide sequence ($n = 159$) of the amino-acid-accepting 3'-OH extremity of turnip-yellow-mosaic-virus RNA and the last portion of its coat-protein cistron. *Eur J Biochem* 72:453–463
17. Hord M, Villalobos W, Macaya-Lizano AV, Rivera C (1997) Chayote mosaic, a new disease in *Sechium edule* caused by a tymovirus. *Plant Dis* 81:374–378
18. Sacristán S, Fraile A, Malpica JN, García-Arenal F (2005) An analysis of host adaptation and its relationship with virulence in *Cucumber mosaic virus*. *Phytopathology* 95:827–833
19. Martin G, Keller W (1998) Tailing and 3'-end labelling of RNA with yeast poly(A) polymerase and various nucleotides. *RNA* 4:226–230
20. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
21. Rambaut A (1996) Se-AI: sequence alignment editor. <http://evolve.zoo.ox.ac.uk/>
22. Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818

23. Valverde RA (1990) Analysis of double-stranded RNA for plant virus diagnosis. *Plant Dis* 74:255–258
24. Lot H, Marrou J, Quiot JB, Esvan C (1972) Contribution à l'étude du virus de la mosaïque du concombre (CMV): Méthode de purification rapide du virus. *Ann Phytopathol* 4:25–38
25. Hitchborn JH, Hills GJ (1965) The use of negative staining in the electron microscopy examination of plant viruses in crude extracts. *Virology* 27:528–540
26. Koonin EV, Dolja VV (1993) Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit Rev Biochem Mol Biol* 28:375–430
27. Bransom KL, Dreher TW (1994) Identification of the essential cysteine and histidine residues of the *Turnip yellow mosaic virus* protease. *Virology* 187:148–154
28. Ranjith-Kumar CT, Gopinath K, Jacob ANK, Svirividhya V, Elango P, Savithri SH (1998) Genomic sequence of *Physalis mottle virus* and its evolutionary relationship with other tymoviruses. *Arch Virol* 143:1489–1500
29. Koenig R, Pleij CWA, Lesemann D-E, Loss S, Vetten HJ (2005) Molecular characterization of isolates of Anagris vein yellowing virus, Plantago mottle virus and Scrophularia mottle virus—comparison of various approaches for tymovirus classification. *Arch Virol* 150:2325–2338
30. Dreher TW, Goodwin JB (1998) Transfer RNA mimicry among tymoviral genomic RNAs ranges from highly efficient to vestigial. *Nucleic Acids Res* 26:4356–4364
31. Jones RAC (2009) Plant virus emergence and evolution: origins, new encounter scenarios, factors driving emergence, effects of changing world conditions, and prospects for control. *Virus Res* 141:113–130
32. Malmstrom CM, Stoner CJ, Brandenburg S, Newton LA (2006) Virus infection and grazing exert counteracting influences on survivorship of native bunchgrass seedling competing with invasive exotics. *J Ecol* 94:264–275
33. Power AG, Mitchell CE (2004) Pathogen spillover in disease epidemics. *Am Naturalist* 164:S79–S89
34. Index of viruses—*Tymoviridae* (2009) In: ICTVdB—The Universal Virus Database. Columbia University, New York. <http://www.ictvonline.org>
35. Alexandre VA, Duarte LML, Rivas EB, Chagas CM, Barradas MM, Koenig R (2000) *Petunia vein banding virus*: characterization of a new tymovirus from *Petunia x hybrida*. *Plant Dis* 84:739–742
36. Guy PL, Dale JL, Adena MA, Gibbs AJ (1984) A taxonomic study of the host range of tymoviruses. *Plant Pathol* 33:337–346
37. Bernal JJ, Jimenez I, Moreno M, Hord M, Rivera C, Koenig R, Rodriguez-Cerezo E (2000) *Chayote mosaic virus*, a new tymovirus infecting *Cucurbitaceae*. *Phytopathology* 90:1098–1104
38. Stephan D, Siddiqua M, Hoang AT, Engelmann J, Winter S, Maiss E (2008) Complete nucleotide sequence and experimental host range of *Okra mosaic virus*. *Virus Genes* 36:231–240
39. Rozanov MN, Koonin EV, Gorbalenya AE (1992) Conservation of the putative methyl-transferase domain: a hallmark of the 'Sindbis-like' supergroup of positive-strand RNA viruses. *J Gen Virol* 73:2129–2134
40. Bransom KL, Wallace SE, Dreher DW (1996) Identification of the cleavage site recognised by the turnip yellow mosaic virus protease. *Virology* 217:404–406
41. Rozanov MN, Drugeon G, Haenni A-L (1995) Papain-like proteinase of *Turnip yellow mosaic virus*: a prototype of a new viral proteinase group. *Arch Virol* 140:273–288
42. Jakubiec A, Drugeon G, Camborde L, Jupin I (2007) Proteolytic processing of *Turnip yellow mosaic virus* replication proteins and functional impact on infectivity. *J Virol* 81:11402–11412
43. Schirawski J, Voyatzakis A, Zaccomer B, Bernardi F, Haenni A (2000) Identification and functional analysis of the *Turnip yellow mosaic virus* subgenomic promoter. *J Virol* 74:11073–11080
44. Hellendoorn K, Mihiels PJA, Biutenhuis R, Pleij CWA (1996) Protonable hairpins are conserved in the 5'-untranslated region of tymovirus RNAs. *Nucleic Acids Res* 24:4910–4917
45. Hellendoorn K, Verlaan PWG, Pleij CWA (1997) A functional role for the conserved protonable hairpins in the 5'-untranslated region of *Turnip yellow mosaic virus* RNA. *J Virol* 71:8774–8779
46. Shin HI, Tzanetakis IE, Dreher TW, Cho TJ (2009) The 5'-UTR of *Turnip yellow mosaic virus* does not include a critical encapsidation signal. *Virology* 387:427–435
47. Hellendoorn K, Mat AW, Gulyaev AP, Pleij CWA (1996) Secondary structure model of the coat protein gene of *Turnip yellow mosaic virus* RNA: long C-rich, single stranded regions. *Virology* 224:43–54
48. Koenig R, Barends S, Gulyaev AP, Lesemann D-E, Vetten HJ, Loss S, Pleij CWA (2005) Nemesia ring necrosis virus: a new tymovirus with a genomic RNA having a histidylatable tobamovirus-like 3' end. *J Gen Virol* 86:1827–1833
49. Gibbs A, Keese PK (1995) In search of the origins of viral genes. In: Gibbs A, Calisher CA, García-Arenal F (eds) *Molecular basis of virus evolution*. Cambridge University Press, Cambridge, pp 76–90
50. Segwagwe AT, Putnam ML, Druffel KL, Pappu HR, Eastwell KC (2008) Molecular characterization of a new tymovirus from *Diascia* ornamental plants. *Arch Virol* 153:1495–1503
51. Srifah P, Keese P, Weiller G, Gibbs A (1992) Comparisons of the genomic sequences of *Erysimum latent virus* and other tymoviruses: a search for the molecular basis of their host specificities. *J Gen Virol* 73:1437–1447
52. Tzanetakis IE, Tsai C-H, Martín RR, Dreher TW (2009) A tymovirus with an atypical 3'-UTR illuminates the possibilities for 3'-UTR evolution. *Virology* 392:238–245
53. Gibbs A (1999) Tymoviruses. In: Granoff A, Webster R (eds) *Encyclopedia of virology*, vol 3. Academic Press, London, pp 1850–1853
54. Muthukumar V, Melcher U, Pierce M, Wiley GB, Roe BA, Palmer MW, Thapa V, Ali A, Ding T (2009) Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions. *Virus Res* 141:169–173
55. Gibbs AJ (1980) A plant virus that partially protects its wild legume host against herbivores. *Intervirology* 13:42–47