GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Assessment of the genetic diversity of tomato yellow leaf curl virus

H.J. Wan[1], W. Yuan[2], R.Q. Wang[1], Q.J. Ye[1], M.Y. Ruan[1], Z.M. Li[1], G.Z. Zhou[1], Z.P. Yao[1] and Y.J. Yang[1]

[1]Institute of Vegetables, Zhejiang Academy of Agricultural Sciences, Hangzhou, Zhejiang, China
[2]Jiaxing Academy of Agricultural Sciences, Jiaxing, Zhejiang, China

Correspondence author: H.J. Wan
E-mail: wanhongjian@sina.com

Genet. Mol. Res. 14 (1): 529-537 (2015)
Received March 7, 2014
Accepted October 2, 2014
Published January 26, 2015
DOI http://dx.doi.org/10.4238/2015.January.26.7

ABSTRACT. The objective of the present study was to analyze the genetic diversity of tomato yellow leaf curl virus (TYLCV). Representative TYLCV sequences were searched in the National Center for Biotechnology Information database. Comprehensive analysis of TYLCV was performed using bioinformatics by examining gene structure, sequence alignments, phylogeny, GC content, and homology. Forty-eight representative TYLCV sequences were selected from 48 regions in 29 countries. The results showed that all TYLCV sequences were 2752-2794 nucleotides in length, which encoded 6 open reading frames (AV1, AV2, AC1, AC2, AC3, and AC4). GC content ranged from 0.41-0.42. Sequence alignment showed a number of insertions and deletions within these TYLCV sequences. Phylogenetic tree results revealed that the sequences were divided into 10 classes; homology of the sequences ranged from 72.8 to 98.6%. All 48 sequences contained the typical structure of TYLCV, including open reading frames and intergenic regions. These results provide a theoretical basis for the identification and evolution of the virus in the future.

Key words: Cluster analysis; Structural characteristics; Homology analysis; Tomato yellow leaf curl virus

INTRODUCTION

Tomato yellow leaf curl virus (TYLCV) was first reported in Israel in the early 1960s (Cohen and Harpaz, 1964). The virus is widely distributed and belongs to the *Begomovirus* genus of the Geminiviridae family and is transmitted by *Bemisia tabaci* (Xie and Zhang, 2002). TYLCV contains single-stranded DNA with a twinned particle morphology. Its genome is composed of 2 components, DNA-A and DNA-B, and each component is 2.5-2.8 kb in size (Fauquet et al., 2005). Currently, most reported TYLCVs are composed of DNA-A and have a genome size of approximately 2781 nucleotides. The virus can be classified into TYLCV-Israel (TYLCV-IL) and TYLCV-Mild (TYLCV-Mld) (Navot et al., 1991; Antignus and Cohen, 1994). Researchers have also identified 3 novel TYLCV strains, including Gezira (TYLCV-Gez), Iran (TYLCV-IR), and Oman (TYLCV-OM), which may have arisen through inter-species recombination events (Bananej et al., 2004; Idris and Brown, 2005; Khan et al., 2008).

Recent studies have been conducted for the molecular identification of TYLCVs. Zhang et al. (2008) identified and analyzed virus isolates in Shanghai, and showed that their homology to one another was lower than that to those from other domestic regions. Furthermore, the virus isolates shared high homology and the closest genetic relationships with isolates from the USA, indicating that they belonged to the same viral strain (Zhang et al., 2008).

Tomato (*Solanum lycopersicum* Mill.), which belongs to Solanaceae, is a worldwide vegetable crop (Moriones and Navas-Castillo, 2000). With the increase in tomato cultivation areas in recent years, TYLCV disease has become one of the most serious diseases limiting tomato production (Czosnek and Laterrot, 1997). Large areas of TYLCV outbreaks have occurred worldwide, which are thought to be related to global climate change. According to preliminary statistics, tomato products in at least 40 countries have suffered damage from this virus (Accotto et al., 2000; Boulton, 2003). Recently, most tomato production areas in China have been infected with the virus, with the number of infections increasing each year (Xu et al., 2007). Therefore, analyzing the genetic diversity of TYLCV will not only reveal insight into TYLCV pathogenesis but also provide a theoretical foundation for the spread, variation, and evolutionary analysis of the virus.

In the present study, TYLCVs from various countries were searched in the National Center for Biotechnology Information (NCBI) database. Structural characteristics, sequence homology, and phylogenetic relationships were analyzed to explore the degree of genetic variability among strains.

MATERIAL AND METHODS

Sequence database search and basic structure of TYLCV

Reported TYLCV isolates in the NCBI database (http://www.ncbi.nlm.nih.gov/) were searched using the key word "TYLCV". These virus isolates were from China (Zhejiang Province, Jiangsu Province, Anhui Province, Shanghai City, Shandong Province, Tianjin City, Beijing City, Yunnan Province, Hebei Province, Henan Province, Shanxi Province, and Xinjiang), Japan, Korea, the USA, Australia, Cuba, Mexico, Egypt, Iran, Morocco, Holland, Sudan, Spain, Portugal, France, Tunisia, Mauritius, Turkey, Lebanon, Puerto Rico, Oman, Dogna, Kuwait, Guatemala, Italy, and Ethiopia. Structural characteristics of these viruses were analyzed.

GC content in TYLCV

The GC content in TYLCV was calculated using the Geecee software (http://weblab.cbi.pku.edu.cn/program.inputForm.do?Program=geecee). GC content in TYLCVs was analyzed using the Cpgplot software (http://weblab.cbi.pku.edu.cn/program.inputForm.do?Program=cpgplot) by selecting TYLCP sequences with different GC contents.

Multiple-sequence alignment and phylogenetic tree

To analyze the conservation of the TYLCV sequence, multiple alignment of nucleotide sequences was conducted using the ClustalX program of the BioEdit software. A TYLCV phylogenetic tree was constructed using the neighbor-joining model in the MEGA 4.0 software (Tamura et al., 2007). Bootstrapping (1000 replicates) was used to evaluate evolutionary trees.

Homology analysis

Homology comparison of the candidate TYLCV sequences was conducted using the DNAMAN (Version 7) software, and genetic diversity was analyzed.

RESULTS

Identification of TYLCV isolates

Candidate TYLCV sequences were searched in the NCBI database using "TYLCV" as the key word. The selected representative TYLCV sequences were from 48 areas in 29 countries (Table 1). TYLCV sequences from 12 different provinces in China were included.

Structural characteristics of TYLCV

The structural characteristics of TYLCV were analyzed. The entire length of the TYLCV gene was between 2752-2794 bp, encoding 6 open reading frames (ORFs) (Table 1), specifically the *AV2* gene (encodes proteins related to protein suppression of host RNA silencing), *AV1* gene (encodes a coat protein), *AC3* gene located in the complementary chain (encodes a replication enhancer protein), *AC2* gene (encodes transcription activator protein), *AC1* gene (encodes a replication-related protein), and *AC4* gene (encodes proteins mainly accumulated in the chloroplast and mitochondrion of the cytoplasm and strongly suppresses RNA silencing). The locations of the 6 ORFs in the virus and its complementary chain are shown in Figure 1. In addition, there was a long non-coding region between *AC1* and *AV2*, known as the intergenic region, which contained the required sequences for virus replication, transcription initiation, and packaging. This region also contained conserved 9-nucleotide sequences of TAATATTAC and TATA boxes as well as the repeat sequences CAATCGGG and GGGTCG with stem loop structure, which are typical structural characteristics of the *Begomovirus* genome (Xie et al., 2002).

Table 1. Characteristics of TYLCV.

Virus strains	Gene accession	Location	GC content	Full length	Open reading frame	
					Virus strand	Complementary strand
TYLCV-ZJ3	AM698117	Zhejiang, China	0.41	2781	2	4
TYLCV-XH2	GU111505	Jiangsu, China	0.41	2781	2	4
TYLCV-AH-HB2	FN650808	Anhui, China	0.41	2781	2	4
TYLCV-SH2	AM282874	Shanghai, China	0.41	2781	2	4
TYLCV-SDHZ	HM627882	Shandong, China	0.41	2781	2	4
TYLCV-Tianjin	GU563330	Tianjin, China	0.41	2781	2	4
TYLCV-Beijing3	GU983859	Beijing, China	0.41	2781	2	4
TYLCV-Y10	AJ319675	Yunnan, China	0.41	2737	2	4
TYLCV-SJZ1	JF727878	Hebei, China	0.41	2782	2	4
TYLCV-HNZZ158	JQ004028	Henan, China	0.41	2781	2	4
TYLCV-SX-8	JN412854	Shanxi, China	0.41	2781	2	4
TYLCV-KS2-5	JQ807735	Xinjiang, China	0.41	2781	2	4
TYLCV- Japan:Haruno	AB192966	Spring, Japan	0.41	2781	2	4
TYLCV-Hwas	GU126513	Korea	0.41	2775	2	4
TYLCV-USA	EF539831	California, USA	0.41	2781	2	4
TYLCV- Australia:Bundaberg2	GU178819	Australia	0.41	2781	2	4
TYLCV-Cuban	AJ223505	Cuba	0.41	2781	2	4
TYLCV-Sinaloa	EF523478	Sinaloa, Mexico	0.41	2781	2	4
TYLCV-Egypt	AY594174	Egypt	0.41	2781	2	4
TYLCV-Abadeh	FJ355946	Abadeh, Iran	0.41	2782	2	4
TYLCV-Moroccan	EF060196	Morocco	0.41	2781	2	4
TYLCV-Netherlands	FJ439569	Holland	0.41	2781	2	4
TYLCV-Gezira	AY044138	Jezira, Sudan	0.42	2780	2	4
TYLCV-Jiroft: Iran	GU076452	Jīroft, Iran	0.42	2770	2	4
TYLCV- Mild[Japan: Osuka]	AB116636	Ōsuka, Japan	0.41	2787	2	4
TYLCV-Mild[Spain7297]	AF071228	Spain	0.41	2791	2	4
TYLCV-Mild[Portugal]	AF105975	Portugual	0.41	2793	2	4
TYLCV-Mild[Jordan]	EF054894	Jordan	0.41	2791	2	4
TYLCV- Mild[Reunion]	AJ865337	France	0.41	2791	2	4
TYLCV-Tunisia	EF101929	Tunisia	0.41	2781	2	4
TYLCV-Jordan	EF054893	Jordan	0.41	2781	2	4
TYLCV-Dominican Republic	AF024715	Dominican	0.41	2781	2	4
TYLCV-Mauritius	HM448447	Mauritius	0.41	2757	2	4
TYLCV-Mersin 1	AJ812277	Mersin, Turkey	0.41	2781	2	4
TYLCV-Poamoho	GU322423	Hawaii, USA	0.41	2781	2	4
TYLCV-Spain: Almeria	NC_004005	Almería, Spain	0.41	2781	2	4
TYLCV- Huasteca	JN680353	Huasteca, Mexico	0.41	2794	2	4
TYLCV-Shiraz:Iran	GU076446	Shiraz, Iran	0.41	2781	2	4
TYLCV-Ra3	EF051116	Lebanon	0.41	2781	2	4
TYLCV-Puerto Rico	AY134494	Puerto Rico	0.41	2781	2	4
TYLCV- DT2	JN604488	Oman	0.41	2767	2	4
TYLCV-New Caledonia	HE603245	New Caledonia Dogna	0.41	2780	2	4
TYLCV-Grenada: Hermitage	FR851297	Grenada	0.41	2752	2	4
TYLCV- KISR	JF451352	Kuwait	0.42	2776	2	4
TYLCV-Guatemala	GU355941	Guatemala	0.41	2781	2	4
TYLCV-8-4/2004	DQ144621	Italy	0.41	2781	2	4
TYLCV-Ethiopia	DQ358913	Ethiopia	0.42	2785	2	4
TYLCV-RE4	AM409201	France	0.41	2781	2	4

GC content of TYLCV

The GC content in TYLCV was found to be 0.42 in the TYLCV sequences in Sudan, Iran, Kuwait, and Ethiopia, while the GC content was 0.41 in all other countries, with an average of 0.4108. Two TYLCVs (isolated from Jiroft, Iran and from Zhejiang, China) with different GC contents were selected and compared. With changes in nucleotide length, changes in GC content were relatively consistent, and slight differences were observed at base pair positions 2300 and 2600 (Figure 2).

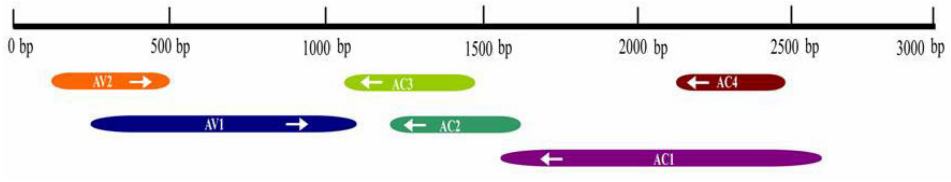

Figure 1. Basic structure of TYLCV.

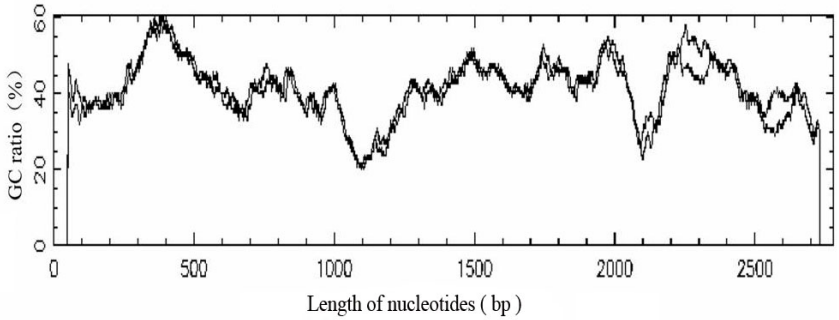

Figure 2. GC content between Iran (Jiroft) and China (Zhejiang).

Conservative analysis of TYLCV

Analysis of conservation in the 48 TYLCV members is shown in Figure 3; in this figure, the shaded region represents the number of bases that were identical. A larger shaded region indicates an increased number of shared bases, showing higher conservation. These results demonstrated that the conservation of TYLCV was high, and only a few blank regions with different sizes existed, indicating the insertion and deletion of different numbers of bases. Differences at the beginning of the sequence were large and conservation in this region was poor.

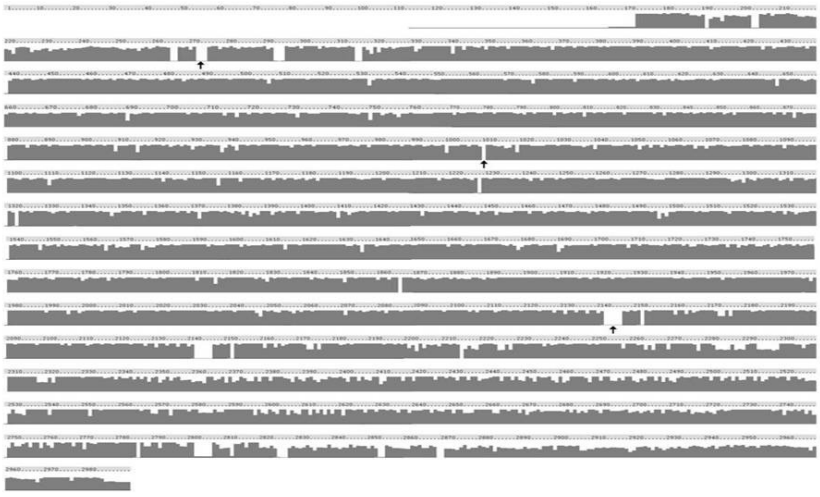

Figure 3. Comparative analysis between nucleotide sequences of TYLCV.

Phylogenetic relationships

To reveal phylogenetic relationships, a phylogenetic tree of TYLCVs from the 48 different countries and areas was constructed (Figure 4). The resulting tree had a large number of branches, and the 48 sequences were classified into 10 categories. The number of members in each class varied widely, with only 1 member each in Classes III, VI, and X, but 15 members in Class I alone. Currently, based on strain criteria, TYLCVs are classified into the TYLCV-Iran, TYLCV-Mld, TYLCV-Israel, and TYLCV-Gezira strains (Bananej et al., 2004; Idris and Brown, 2005; Khan et al., 2008). In the present study, however, the collected TYLCV isolates did not accurately cluster together in the phylogenetic tree according to strain criteria, showing large differences. Class VIII TYLCV from France (wild-type strain), Spain, Portugal, Jordan (wild-type strain), and Yokosuka, Japan, belonged to the TYLCV-Mld strain, the TYLCV from Sultan Gezira belonged to the TYLCV-Gezira strain, the Class IX TYLCV from Jiroft in Iran and Class VII TYLCV from Shiraz in Iran belonged to the TYLCV-Iran strain, and the Class I TYLCV from Grenada and the Class V TYLCV from Kochi in Japan belonged to the TYLCV-Iran strain. No detailed description on TYLCV from other areas was performed. TYLCVs in the 12 provinces of China were classified into 3 classes: Class X, including TYLCVs from Yunnan Province; Class IV, including TYLCVs from Beijing, Jiangsu, and Xinjiang of China and those of North America; and Class V, including TYLCVs from 8 provinces of China, as well as Australia, South Korea, and Japan. The TYLCVs exhibited no obvious classification in regards to geography, in which the Class I TYLCVs originated from different areas of Asia, America, Europe, Africa, and Oceania.

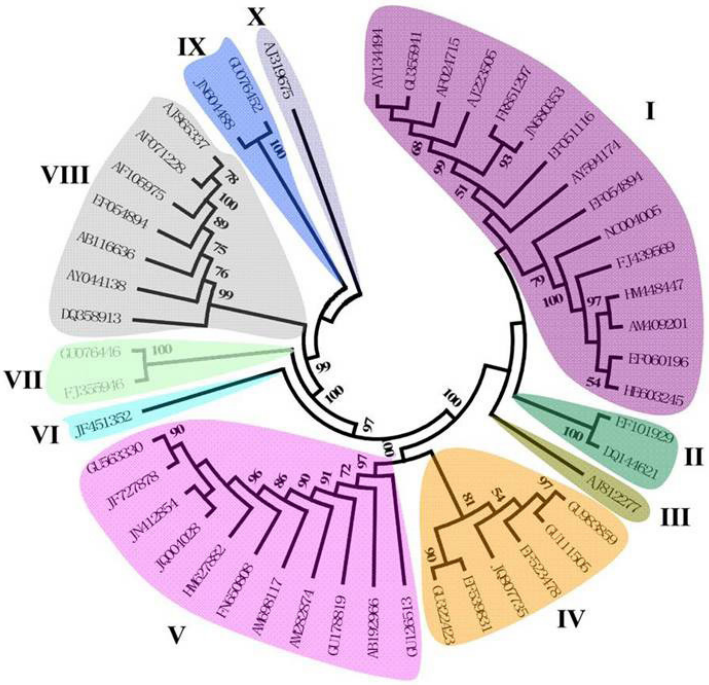

Figure 4. Phylogenetic tree of TYLCV nucleotide sequences from 48 regions.

Homology analysis

Based on the results of phylogenetic tree classification, 1 TYLCV member in each class was selected for genetic diversity analysis, and the nucleotide sequence of the 10 TYLCV members was subjected to homology alignment. The degree of homologous variation among the 10 members was large. The highest homology between the sequences was 98.6% (between TYLCV in Tianjin and TYLCV in Beijing), while the lowest homology was 72.8% (between TYLCV in Yunnan of China and the wild-type TYLCV strain of France). Homologies between TYLCV in Yunnan and TYLCV in other areas were all below 75.9% (Table 2).

Table 2. Homology analysis of 10 groups of TYLCV.

Category	Virus location	Puerto Rico	Tunisia	Mersin, Turkey	Beijing, China	Tianjin, China	Kuwait	Shiraz, Iran	France	Oman	Yunnan, China
1	Puerto Rico	100.0%									
2	Tunisia	97.8%	100.0%								
3	Mersin, Turkey	98.0%	98.2%	100.0%							
4	Beijing, China	97.5%	97.8%	98.2%	100.0%						
5	Tianjin, China	97.1%	97.5%	97.7%	98.6%	100.0%					
6	Kuwait	94.6%	94.9%	95.2%	94.9%	94.3%	100.0%				
7	Shiraz, Iran	94.6%	94.7%	95.0%	94.9%	94.6%	92.9%	100.0%			
8	France	92.0%	91.8%	93.3%	92.4%	92.2%	89.6%	91.2%	100.0%		
9	Oman	87.8%	87.9%	89.2%	88.5%	87.9%	86.7%	89.2%	86.4%	100.0%	
10	Yunman, China	75.2%	75.7%	75.7%	75.9%	75.6%	74.4%	74.6%	72.8%	76.0%	100.0%

DISCUSSION

Structural characteristics of TYLCV

Bioinformatics is a method of evaluating biological data using computer science, and plays an important role in accelerating research on plant functional genomics. In this study, we identified 48 TYLCV sequences, which encoded 2752-2794 bp and 6 ORFs. Previous studies have shown that the *AV1* gene in the Geminiviruses genome was most conserved and showed the highest homology among TYLCV isolates from the same area (Hong and Harrison, 1995; Padidam et al., 1995). These structures were similar to the pathogenic structures identified by Yu et al. (2009) and Xiong et al. (2011), which were typical characteristics of the *Begomovirus* genome. In this study, there were no large differences in GC content in different TYLCVs. Similarly, no large variations in GC content were observed in different virus species of the tomato spotted wilt virus (*Tospovirus*) in previous studies (Liu et al., 2009). In addition, sequence comparisons revealed that most bases were the same, with the exceptions of individual base insertions and deletions, suggesting that the sequence of TYLCV is relatively conserved.

Genetic diversity of TYLCVs

The phylogenetic tree constructed in this study indicated that the TYLCVs were classified into 10 groups. The number of virus members in each class ranged from 1 to 15. The viruses of the 4 strains were not clustered together; the viruses from the 12 areas in China also did not cluster together. This indicates that the viruses cannot be grouped by geographical

location or strain classification. The pathogen of TYLCV in Sichuan was analyzed by Xiong et al. (2011), who found that the pathogen may be derived from virus isolates in Yunnan Province. Zhang et al. (2008) hypothesized that the pathogen in Shanghai areas was derived from America and Africa. There have been no studies regarding the occurrence, transmission, and spread of TYLCV. In this study, we found that the genetic relationship of TYLCV is not only related to geography and strains but also other factors, such as transmission by wind and rivers, human trade, and evolution of the *Bemisia tabaci* mediator.

Based on the phylogenetic tree, 1 member in each class was selected for homology analysis. Differences between sequences were large, and the lowest and highest homologies were 72.8 and 98.6%, respectively. Sequence homologies between TYLCV in Yunnan and other members were lower than 75.9%. The International Virus Committee has stipulated that if the homology of the entire nucleotide sequence in Geminiviridae viruses is less than 89%, they should be considered different viruses, but if homology is greater than 89%, they belong to different strains of the same virus (Padidam et al., 1995; Fauquet et al., 2003). Although the virus isolates in Yunnan were referred to as TYLCV in the NCBI database, our results suggest that they might be distinct from viruses from other areas. Virus isolates from Yunnan were not TYLCV, and the viruses in other areas may be different strains of TYLCV.

ACKNOWLEDGMENTS

Research supported by the National HighTechnology Research and Development Program ("863"Program) of China (#2012AA100103006); Key Projects in the National Science & Technology Pillar Program (#2012BAD02B02); the National Natural Science Foundation of China (#31071800 and #31301774); Breeding of Vegetable Varieties in Zhejiang Province (#2009C02006-1), Zhejiang Province Agricultural Science and Technology Major Project-Breeding of New Vegetable Varieties (2012C12903) and National Staple Vegetable Industry Technology System (#CARS-25-G-16).

REFERENCES

Accotto GP, Navas-Castillo J, Noris E, Moriones E, et al. (2000). Typing of tomato yellow leaf curl viruses in Europe. *Eur. J. Plant Pathol.* 106: 179-186.

Antignus Y and Cohen S (1994). Complete nucleotide-sequence of an infectious clone of a mild isolate of tomato yellow leaf curl virus (TYLCV). *Phytopathology* 84: 707-712.

Bananej K, Kheyr-Pour A, Salekdeh GH and Ahoonmanesh A (2004). Complete nucleotide sequence of Iranian tomato yellow leaf curl virus isolate: further evidence for natural recombination amongst begomoviruses. *Arch. Virol.* 149: 1435-1443.

Boulton MI (2003). Geminiviruses: major threats to world agriculture. *Ann. Appl. Biol.* 142: 143.

Cohen S and Harpaz I (1964). Periodic, rather than continual acquisition of a new tomato virus by its vector, the tobacco whitefly (*Bemisia tabaci* Gennadius). *Entomol. Exp. Appl.* 7: 155-166.

Czosnek H and Laterrot H (1997). A worldwide survey of tomato yellow leaf curl viruses. *Arch. Virol.* 142: 1391-1406.

Fauquet C, Bisaro D, Briddon R, Brown J, et al. (2003). Virology division news: revision of taxonomic criteria for species demarcation in the family Geminiviridae, and an updated list of begomovirus species. *Arch. Virol.* 148: 405-421.

Fauquet CM, Mayo MA, Maniloff J, Desselberger U, et al. (2005). Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses. Academic Press, Waltham, 285-297.

Hong Y and Harrison B (1995). Nucleotide sequences from tomato leaf curl viruses from different countries: evidence for three geographically separate branches in evolution of the coat protein of whitefly-transmitted geminiviruses. *J. Gen. Virol.* 76: 2043-2049.

Idris AM and Brown JK (2005). Evidence for interspecific-recombination for three monopartite begomoviral genomes

associated with the tomato leaf curl disease from central Sudan. *Arch. Virol.* 150: 1003-1012.

Khan AJ, Idris AM, Al-Saady NA, Al-Mahruki MS, et al. (2008). A divergent isolate of tomato yellow leaf curl virus from Oman with an associated DNA beta satellite: an evolutionary link between Asian and the Middle Eastern virus-satellite complexes. *Virus Genes* 36: 169-176.

Liu YT, Li X, Li YZ, Li ZY, et al. (2009). Sequence alignment and analysis of the ssRNA-M of Tospovirus. *J. Southwest Agric. Univ.* 12: 1-7.

Moriones E and Navas-Castillo J (2000). Tomato yellow leaf curl virus, an emerging virus complex causing epidemics worldwide. *Virus Res.* 71: 123-134.

Navot N, Pichersky E, Zeidan M, Zamir D, et al. (1991). Tomato yellow leaf curl virus: a whitefly-transmitted geminivirus with a single genomic component. *Virology* 185: 151-161.

Padidam M, Beachy RN and Fauquet CM (1995). Classification and identification of geminiviruses using sequence comparisons. *J. Gen. Virol.* 76: 249-263.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

Xie Y and Zhang ZK (2002). Rapid detection of whitefly-transmitted geminiviruses by TAS-ELISA and PCR. *Plant Pathol.* 32: 182-186.

Xie Y, Zhou X, Zhang Z and Qi Y (2002). Tobacco curly shoot virus isolated in Yunnan is a distinct species of Begomovirus. *Chin. Sci. Bull.* 47: 199-201.

Xiong Y, Yang S, Qing L, Zhou CY, et al. (2011). Molecular identification and variation analysis of the pathogen causing tomato yellow leaf curl disease in Sichuan province. *Sci. Agricul. Sin.* 44: 477-484.

Xu Y, Cai X and Zhou X (2007). Tomato leaf curl Guangxi virus is a distinct monopartite begomovirus species. *Eur. J. Plant Pathol.* 118: 287-294.

Yu WG, Zhao TM, Yang ML, Zhao LP, et al. (2009). PCR detection and sequence analysis of whitefly-transmitted Geminivirus in tomato from Anhui and Shandong provinces. *Jiangsu J. Agricul. Sci.* 25: 747-751.

Zhang YP, Zhu WM, Cui HM, Qiu Y, et al. (2008). Molecular identification and the complete nucleotide sequence of TYLCV isolate from Shanghai of China. *Virus Genes* 36: 547-551.