New isolate of tomato yellow leaf curl virus in Iraq

Ali Alabde, Osamah Alisawi*, Fadhal Al Fadhal

Department of Plant Protection, Faculty of Agriculture, Kufa University, Najaf, Iraq

*Corresponding Author

Osamah Alisawi
Department of Plant Protection
University of Kufa
Najaf-Iraq
Tel: 009647816353396

E-mail: osama.alisawi@uokufa.edu.iq

Keywords: Tomato yellow leaf curl virus; next generation sequencing; bioinformatics; tomato genome.

Abstract

Tomato yield and quality in Iraq have been threatened by a variable range of infections caused by tomato yellow leaf curl virus. In previous studies, the TYLCV isolates were characterized using molecular tests for small fragments not the entire length of the virus. As a part of long term study, sample of slightly TYLCV-infected tomato has applied in this study to figure out complete sequence of TYLCV isolate. Three sets of primers that belong to three well-known isolates in Iraq were used in a PCR technique and interestingly the results were negative. A new Iraqi isolate has been characterized as a first novel Iraqi isolate detected ever using next generation (NGS) and bioinformatics techniques. The NGS platform has produced about 78,232,062 paired reads of the TYLCV-infected tomato var. Oula F1. The complete raw reads of the infected variety have been analyzed using RepeatExplorer pipeline and Map to reference. The full sequence of TYLCV was reconstructed and extracted to consist of 2770 nt and then

deposited in Genbank under accession number MT583814. Copy numbers and genome proportion of this sequence have been calculated that were 3523 and 0,086% respectively. Phylogenetic analyses of full nucleotide sequences confirmed close relationship to Iranian isolate (TYLCV-Kahnooj) than other published viruses. Additionally, a and β DNA satellites have not discovered in the TYLCV-infected sample.

Introduction

The family Geminiviridae is recently divided into seven registered genera: *Becurtovirus*, *Begomovirus*, *Curtovirus*, *Eragrovirus*, *Mastrevirus*, *Topocuvirus* and *Turncurtovirus.* Host range, insect vector, genome organization and genome-wide pairwise sequence identities have been applied to classify this family (Varsani *et al*., 2017). Tomato yellow leaf curl virus (TYLCV) belongs to genus *Begomovirus* and has a monopartite single-stranded DNA (ssDNA) encapsidated in geminate particles. TYLCV is transmitted by the whitefly, *Bemisia tabaci* (Genn.) that effectively spread virus particles through persistent manner (Panno *et al*., 2016). The disease caused by this virus is the most important disease on tomato in Iraq and worldwide (Mansoor *et al*., 2003; Salati *et al*., 2002; Al-Kuwaiti *et al*., 2013). Infected plants become stunted with chlorotic leaves, curled-up margins, and abscised flowers. Since the first infection was recorded in Israel 1939-1940, multiple regions have been registered to have isolates from this virus over the world and classified to different viral species (Antignus and Cohen 1994; Delatte *et al*., 2004; Pico *et al*., 1996; Polston and Anderson 1997). Tomato yellow leaf curl Sardinia virus TYLCSV infected tomatoes in Israel and Sardinia and was the first sequenced isolate in 1991(Kheyr-Pour *et al*., 1991; Navot *et al*., 1991), and since that date many isolates have been sequenced from various locations (Moriones and Navas-Castillo 2000). In Iraq, the disease was detected based on symptoms assay and serological tests. Recently, most popular molecular works on this virus have been achieved by PCR based on amplified short fragments of TYLCV genome such as coat protein domain (Al-Ani *et al*., 2010; Reddy *et al*., 2005; Maruthi *et al*., 2007). Further, many tomato genotypes have been screened against TYLCV isolates to characterize similarities and differences through these genotypes (Al Tamimi, 2020). TYLCV has been isolated not only from tomato but also from various plants such as eggplant, melon, cucumber, sesame, cowpea in addition to weeds as a primary source of

infection like *Malva parviflora* , *Chenopodium murale*, *Chenopodium quinoa* (Al- Jubouri *et al* ., 2014). Al-Kuwaiti *et al* (2013) have reported complete sequence of TYLCV that meet 99% with Spain isolate confirming its existence in Iraq. In this study, we utilized next generation sequencing and bioinformatics techniques for the first time in Iraq to find out new isolate with complete sequence of TYLCV. Furthermore, phylogenetic and sequence alignment analyses have been performed to understand how these isolates dispersed and related to each other.

Materials and methods

Plant material

From open fields of the main region of tomato production in Iraq, fresh and young leaves of TYLCV-infected tomato variety Oula F1 (Seminis Seeds Company, USA) were harvested. This variety is considered the most popular and TYLCV-sensitive commercial variety of tomato in the middle of Iraq.

DNA isolation, PCR and NGS sequencing

To extract the DNA, the young leaves (about 2 gm) of the infected tomato were applied using cetyl-trimethylammonium bromide (CTAB) method with minor modifications (Doyle and Doyle 1990). Three sets of primers that belong to isolates of TYLCV-IS, TYLCV-ES and TYLCV-mld were applied in a PCR technique to check their infection within examined sample (Al-Jubouri *et al*., 2014). The extracted sample was sent to sequencing at DNA-link Company, Republic of Korea. The genome was sequenced by using the Novaseq6000 2x150bp reads technique and application WGS (PCR Free550) based on the manufacturer's procedure.

RepeatExplorer program

To collect virus sequences from the whole data (NGS) of infected tomato, RepeatExplorer pipeline was applied to recognize these kinds of sequences in forms of clusters within repetitive DNA blocks (Novák *et al*. 2013). The targeted clusters of Repeat Explorer were extracted, and then Repbase dataset (Jurka *et al*. 2005) and Basic Local Alignment Search Tool (Altschul *et al*. 1990) were applied to identify contigs of each cluster. Furthermore, the extracted sequences were mapped to well-known plant virus sequences from DPVweb (Adams and Antoniw 2005).

Map to reference

Using Geneious version 11 (Kearse *et al*. 2012), the raw read was assembled against TYLCV Iraq as a reference sequence that diagnosed by Al-Kuwaiti *et al.* (2013) (Accession number JQ354991.1) and then consensus sequence was collected. Mapping to reference was repeated against consensus sequence (TYLCV) to confirm and obtain the actual number of assembled reads. Also, the reads were applied to assemble against a and β DNA satellites. The outcomes showed in a report that has a number of assembled reads and total used reads in addition to the frequent overlapped reads. This data used to calculate copy numbers (Number of assembled reads x read length/reference sequence length) and genome proportions (Number of assembled reads / numbers of total NGS reads x 100) (Mustafa *et al*. 2018). Finally, the two consensus sequences from RepeatExplorer and Map to reference were aligned using pairwise alignment in Geneious program to collect the right length of TYLCV isolate.

Phylogenetic analysis

The maximum likelihood (ML) method was applied to find phylogeny model using MEGA7 (Tamura *et al*. 2013). The Geneious program V. 11 (Kearse *et al*. 2012) was used for alignment and optimized manually. The entire nucleotide sequences of 13 main TYLCV strains including Najaf isolate, Kahnooj and JQ354991.1 were applied to phylogeny study. ClustalW alignment was used for extracting sequences about 2770 nt for each. The phylogenetic tree reconstructed using General Time Reversible (GTR). Bayesian phylogeny inference was used for analysis with Bayesian inference of phylogeny (MrBayes 3.2.6) (Huelsenbeck and Ronquist 2001).

Results

The extracted DNA concentration and quality were 418 ng\ml and 1.92 which was in a high quality level for sequencing approach (Fig. 1. A). PCR has surprisingly failed to get amplified sequences for the three TYLCV isolates indicating that sample has a new isolate. High quality raw reads have been obtained from NGS Illumina platform with about 78,232,062 paired reads; 150bp read length and 36.9% GC content (Fig. 1. B).

RepeatExplorer has been clustered the whole uploaded raw reads that showed tandemly repeated sequences, TYLCV sequences involved in cluster number 76 as a putative satellite with a genome proportion 0,084% and the consensus sequence was 2763 nt (Fig. 1. C). On the other side, the report of Map to reference revealed that 65,303 reads have been assembled to count genome proportion and copy numbers of the isolate that were 0,086% and 3523 respectively whereas the consensus sequence was 2770 nt. In order to obtain complete sequence, the aligned consensus sequences from both RepeatExplorer and Map to reference showed 100% similarity and the longest sequence was extracted. Lastly, we determined 2770 nt of TYLCV and named as Najaf isolate and then deposited in Genbank under accession number MT583814.1.

Phylogenetic analyses showed what we have proven earlier by pairwise alignment between Najaf isolate and each of Kahnooj and JQ354991.1 isolates. Kahnooj was most likely similar to Najaf isolate than JQ354991.1. TYLCV nucleotide and amino acid sequence alignments showed that Najaf isolate is closely related to Kahnooj isolate the Iranian TYLCV in 93.8% pairwise identity while it was 90.3% with JQ354991.1 isolate (Fig. 2). However, polymorphic regions in intergenic region (IR) (75%), V2 (91.9%), C1 (Rep) (93.1%), and C4 (88.3%) are higher than other coding regions V1 (97.3%), C3 (95.9%) and C2 (96.5%).

Our research has been extended trying to find DNA-a and β satellites using Map to reference tool but we did not get any positive hits.

Discussion

The deep sequencing of whole genome sequencing technique has revolutionized genomic research collecting a massive data of the host genome with complete sequences of cellular components. Additionally, this method is considered as unbiased approach to detect plant viruses with no previous knowledge about the suspected viruses. The NGS technique does not use specific primers, antibodies, and virus-specific reagents to diagnose pathogenic viruses (Adams *et al*., 2009). In this method, full image of plant viruses and complex infection could be investigated in the host genome unlikely to biased methods that focus on almost one expected virus (Jones *et al*., 2017). In Iraq, most of molecular tests concentrated on partial sequences of TYLCV such as coat protein domain to detect the virus and apply these amplified parts in

multiple alignments and phylogeny analyses. The only full-length sequenced of this virus from Iraq was JQ354991.1 that achieved by Al-Kuwaiti *et al* (2013) and named as TYLCV Iraq but it shared 99% similarity to Spain strain (Acc. AJ489258), Reunion Islands (Acc. AM409201) and Mauritius (Acc. HM448447). Additionally, Sadeq Al-Kuwaiti (2013) has applied rolling circle amplification (RCA) and non-degenerate primers to catch the full length of virus. Therefore, our new isolate is considered the first full sequenced Iraqi isolate of TYLCV with totally new data yet. We found Kahnooj isolate from Iran is closer than other isolates even from Iraq with 93.8 % similarity. It is highly expected that Najaf and Kahnooj isolates are sharing the same ancestor and most likely the virus was imported into Iraq by international trading from Iran (Mali *et al*., 2003; Gibbs *et al*., 2008; Al-Kuwaiti 2013).

Mapping reads to the reference has revealed the estimated copy numbers of this virus was 3523 within infected cell. Copy number value has been always considered as a good sign for variation and variability of endogenous components within host genome especially transposable elements including viruses (Mustafa *et al*. 2018; Alisawi 2019). Interestingly, this is one of NGS and bioinformatics benefits to calculate copy numbers of the components of interest and this approach has been applied in many studies to predict viral copies (Yang *et al*., 2017; Alisawi 2019). So far, the number of copies indicates virus particles indeed infected the host genome. In fact, this little amount of copies strongly explains the early stage of symptoms emerging in the infected sample.

We could not reveal the existence of DNA satellites in our infected sample probably due to the early stage of infection that induced very light symptoms. A and β DNA satellites are often associated with begomoviruses in their replication, movement within infected plant and horizontal transmission between plants, as well as involved in induction of virus symptoms and suppression of transcriptional gene silencing. This result is consistent with most published outcomes that reported the importance of DNA satellites for disease severity and inducing symptoms. However, we do expect to find these satellites in severely TYLCV-infected plants as no more researches have been covered DNA satellites in Iraq (Laufs *et al*., 1995; Saunders *et al*., 2004; Cui *et al*., 2005; Yang *et al*., 2011; Amin *et al*., 2011; Zhou, 2013).

Conclusion

This work concluded that whole genome sequencing technique alongside bioinformatics is highly recommended to detect complete sequence of TYLCV isolate within infected tomato genome raw reads. Consequently, a clear insight of virus pathway and identity has been revealed by phylogenetic analysis. More investigations need to be done for detecting full sequences of highly pathogenic plant viruses in Iraq.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

There are no funding sources to report.

Statement of Ethics

The authors have no ethical conflicts to disclose.

Author Contributions

A.A. carried out the experiment. O.A. analyzed the NGS data and wrote the manuscript with support from F.A. F.A. helped supervise the study.

References

Adams, I. P., Glover, R. H., Monger, W. A., Mumford, R., Jackeviciene, E., Navalinskiene, M., ... & Boonham, N., 2009. Next generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular plant pathology*, *10*(4), 537-545.

Adams M.J. & Antoniw J.F, 2005. DPVweb: An open access internet resource on plant viruses and virus diseases. *Outlooks on Pest Management* 16, 268-270.

Al-Ani, R. A., Diwan, S. N. H., & Adhab, M. A, 2010. Efficiency of *Thuja orientalis* and *Artimisia campestris* extracts to control of *Potato leaf roll virus* (PLRV) in potato plants. *Agriculture and Biology Journal of North America*, *1*(4), 579-583.

Alisawi, O. N, 2019. Virus integration and tandem repeats in the genomes of Petunia (Doctoral dissertation, University of Leicester).

Al-Jubouri, A. A., Al-Fadhal, F. A., & Samaka, H. M, 2014. Detection of *tomato yellow leaf curl virus* (TYLCV) and its strains on tomato plants and host rang determine by Polymerase Chain Reaction. *Kufa Journal for Agricultural Sciences*, *6*(4), 99-117.

Al-Kuwaiti, N.; B. Otto; C. Collins; S. Seal and Maruthi, M, 2013. Molecular characterization and first complete genome sequence of *tomato yellow leaf curl virus* (TYLCV) infecting tomato in Iraq. *New Disease Reports*, 27: 17.

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J, 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215, 403-410.

Al Tamimi, S. Q, 2020. Molecular identification of some isolates of *tomato yellow leaf curl virus* and test response some tomato genotypes and their control(Doctoral dissertation, University of Kufa)(In Arabic).

Amin, I., Hussain, K., Akbergenov, R., Yadav, J. S., Qazi, J., Mansoor, S., ... & Briddon, R. W, 2011. Suppressors of RNA silencing encoded by the components of the cotton leaf curl begomovirus-betasatellite complex. *Molecular plant-microbe interactions*, *24*(8), 973-983.

Antignus, Y., & Cohen, S, 1994. Complete nucleotide sequence of an infectious clone of a mild isolate of *tomato yellow leaf curl virus* (TYLCV). *Phytopathology*, *84*(7), 707-712.

Cui, X., Li, G., Wang, D., Hu, D., & Zhou, X., 2005. A begomovirus DNAβ-encoded protein binds DNA, functions as a suppressor of RNA silencing, and targets the cell nucleus. *Journal of Virology*, *79*(16), 10764-10775.

Delatte H, Naze F, Granier M, Reynaud B, Peterschmitt M, Lett M., 2004. Tomato-infecting begomoviruses in the South Western islands of the Indian Ocean. *Presented at the 4th International Geminivirus Symposium*, W3-4.

Doyle J. & Doyle J., 1990. Isolation of DNA from small amounts of plant tissues. *BRL Focus* 12, V15.

Gibbs, A. J., Ohshima, K., Phillips, M. J., & Gibbs, M. J., 2008. The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PloS one*, *3*(6), e2523.

Huelsenbeck J.P. & Ronquist F., 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754-755.

Jones, S., Baizan-Edge, A., MacFarlane, S., & Torrance, L., 2017. Viral diagnostics in plants using next generation sequencing: computational analysis in practice. *Frontiers in Plant Science, 8*, 1770.

Jurka J., Kapitonov V.V., Pavlicek A., Klonowski P., Kohany O. & Walichiewicz J., 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110, 462-467.

Jones, S., Baizan-Edge, A., MacFarlane, S., & Torrance, L., 2017. Viral diagnostics in plants using next generation sequencing: computational analysis in practice. *Frontiers in Plant Science, 8*, 1770.

Kheyr-Pour A, Bendahmane M, Matzeit V, Accotto GP, Crespi S, Gronenborn B., 1991. *Tomato yellow leaf curl virus* from Sardinia is a whitefly-transmitted monopartite geminivirus. *Nucleic Acids Res* 19:6763–6769.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S. & Duran C., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647-1649.

Laufs J., Traut W., Heyraud F., Matzeit V., Rogers S. G., Schell J., Gronenborn B., 1995. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of *tomato yellow leaf curl virus*. *Proc Natl Acad Sci U S A* 92:3879–3883.

Mali, V. R., Subr, Z., & Kudela, O., 2003. Seed transmission of como and potyviruses in faba bean and vetch cultivars introduced into Slovakia. *Acta phytopathologica et entomologica hungarica*, *38*(1-2), 87-97.

Mansoor, S., Briddon, R. W., Zafar, Y., & Stanley, J., 2003. Geminivirus disease complexes: an emerging threat. *Trends in plant science*, *8*(3), 128-134.

Maruthi M, Rekha A, Mirza S, Alam S, Colvin J., 2007. PCR-based detection and partial genome sequencing indicate high genetic diversity in Bangladeshi begomoviruses and their whitefly vector, *Bemisia tabaci*. *Virus Genes* 34, 373–385.

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

Mustafa S.I., Schwarzacher T. & Heslop-Harrison J., 2018. Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors. *Mitochondrial DNA Part A*, 1-14.

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

Novák P., Neumann P., Pech J., Steinhaisl J. & Macas J., 2013. RepeatExplorer: a Galaxy based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29, 792-793.

Panno, S., Iacono, G., Davino, M., Marchione, S., Zappardo, V., Bella, P., ... & Davino, S., 2016. First report of *tomato leaf curl New Delhi virus* affecting zucchini squash in an important horticultural area of southern Italy. *New Dis. Rep*, *33*(6), 2044-0588.

Pico B, Diez MJ, Nuez F., 1996. Viral diseases causing the greatest economic losses to the tomato crop. II. The *tomato yellow leaf curl virus*. *Sci Hort* 67:151–196.

Polston JE, Anderson PK., 1997. The emergence of white fly transmitted geminiviruses in tomato in the Western Hemisphere. *Plant Dis* 81:1358–1369.

Reddy, R. C., Colvin, J., Muniyappa, V., & Seal, S., 2005. Diversity and distribution of begomoviruses infecting tomato in India. *Archives of virology*, *150*(5), 845-867.

Sadeq Al-Kuwaiti, N. A., 2013. Molecular characterization of viruses infecting potato and vegetables in Iraq (Doctoral dissertation, University of Greenwich).

Salati, R., Nahkla, M. K., Rojas, M. R., Guzman, P., Jaquez, J., Maxwell, D. P., & Gilbertson, R. L., 2002. *Tomato yellow leaf curl virus* in the Dominican Republic: characterization of an infectious clone, virus monitoring in whiteflies, and identification of reservoir hosts. *Phytopathology*, *92*(5), 487-496.

Saunders, K., Norman, A., Gucciardo, S., & Stanley, J., 2004. The DNA β satellite component associated with ageratum yellow vein disease encodes an essential pathogenicity protein (βC1). *Virology*, *324*(1), 37-47.

Tamura K., Stecher G., Peterson D., Filipski A. & Kumar S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30, 2725- 2729.

Varsani, A., Roumagnac, P., Fuchs, M., Navas-Castillo, J., Moriones, E., Idris, A., ... & Martin, D. P., 2017. *Capulavirus* and *Grablovirus*: two new genera in the family Geminiviridae. *Archives of Virology*, *162*(6), 1819-1831.

Yang, Z., Mammel, M., Papafragkou, E., Hida, K., Elkins, C. A., & Kulka, M., 2017. Application of next generation sequencing toward sensitive detection of enteric viruses isolated from celery samples as an example of produce. *International journal of food microbiology*, *261*, 73-81.

Yang, X., Xie, Y., Raja, P., Li, S., Wolf, J. N., Shen, Q., *et al*., 2011. Suppression of methylation-mediated transcriptional gene silencing by βC1-SAHH protein interaction during geminivirus-betasatellite infection. *PLoS Pathog.* 7:e1002329. doi: 10.1371/journal.ppat.1002329.

Zhou, X., 2013. Advances in understanding begomovirus satellites. *Annual review of phytopathology*, *51*.

Figures

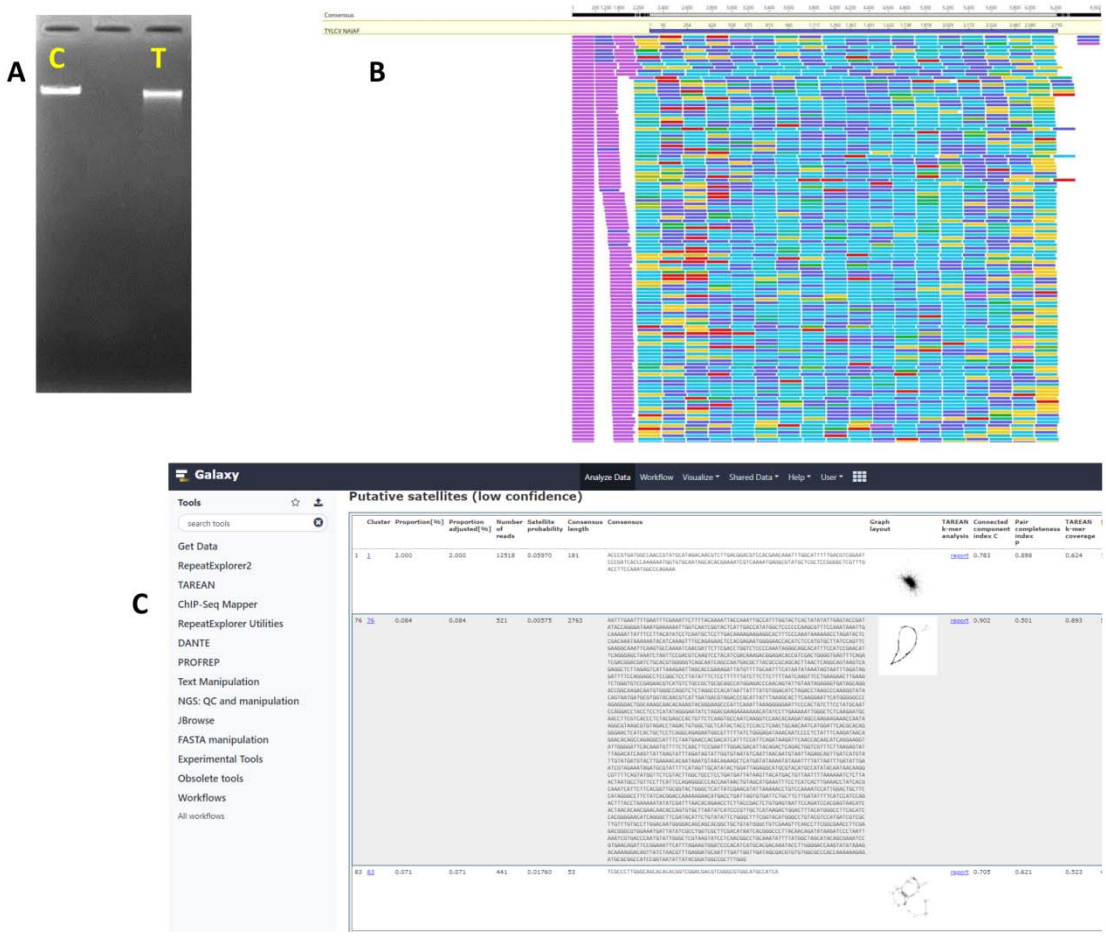

Fig. 1. A: Gel electrophoresis of genomic DNA of TYLCV-infected tomato var. Oula F1 (T), in

comparison with a control genomic DNA (C). B: Raw reads mapped to complete sequence of TYLCV-Najaf isolate (2770 nt), the assembled reads were 65,303 that used to calculate genome proportion and copy numbers. C: RepeatExplorer report showed cluster number 76 for TYLCV sequence (2763 nt) that picked up as a putative satellite due to tandemly arranged feature within raw reads.

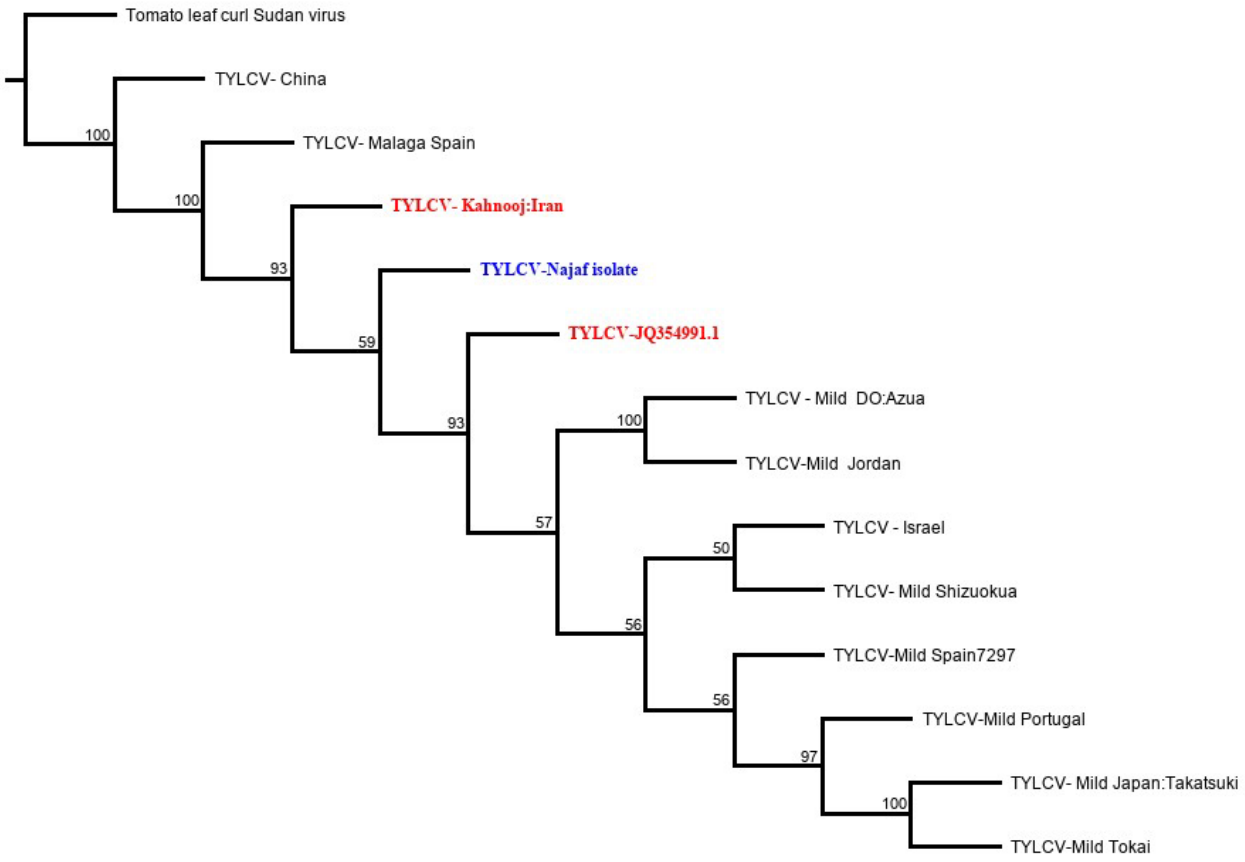

Fig. 2. Phylogenetic tree of 13 full lengths TYLCV isolates showing close relationship between Najaf and Kahnooj isolates. The tree was generated using Geneious tree builder with bootstrape analyses and 50% support, the out group member was *tomato leaf curl Sudan virus*.

Words count: 3407