Complete genome sequence of an isolate of leek yellow stripe virus from garlic in India

Article *in* JOURNAL OF PLANT PATHOLOGY · October 2017

DOI: 10.4454/jpp.v99i3.3945

4 authors:

Nitika Gupta
ICAR-Directorate of Floricultural Research, Pune
8 PUBLICATIONS **14** CITATIONS
SEE PROFILE

Sabina Islam
Indian Agricultural Research Institute
17 PUBLICATIONS **35** CITATIONS
SEE PROFILE

Susheel Kumar Sharma
ICAR Research Complex for NEH region, Manipur Centre, Imphal
86 PUBLICATIONS **187** CITATIONS
SEE PROFILE

V. K. Baranwal
Indian Agricultural Research Institute
142 PUBLICATIONS **911** CITATIONS
SEE PROFILE

Short Communication

COMPLETE GENOME SEQUENCE OF AN ISOLATE OF LEEK YELLOW STRIPE VIRUS FROM GARLIC IN INDIA*

N. Gupta[1], S. Islam[2], S.K. Sharma[3] and V.K. Baranwal[1]

[1]*Plant Virology Unit, Division of Plant Pathology, Indian Agricultural Research Institute, New Delhi 110012, India*
[2]*Division of Vegetable Science, Indian Agricultural Research Institute, New Delhi 110012, India*
[3]*ICAR Research Complex for NEH Region, Manipur Centre, Imphal, 795004, India*

SUMMARY

The complete genome sequence of leek yellow stripe virus (LYSV) isolate AC-50 from garlic in India was determined and the predicted amino acid (aa) sequence was analyzed. The LYSV RNA genome is 10,131 nucleotides (nt) long excluding the poly(A) tail (KP168261) and has a large open reading frame (ORF) encoding a putative polyprotein of 3152 aa with conserved motifs typical of members of the genus *Potyvirus* in the family *Potyviridae.* The pretty interesting *Potyviridae* ORF (PIPO) coding region was identified in the P3 coding region. LYSV isolate AC-50 shares maximum nt and aa sequence identity of 79.9% and 87.2% with a LYSV isolate from Australia (HQ258895) at the full genome and polyprotein level, respectively, and clusters with clade II isolates from China, Mexico, Australia, Brazil and Spain. The P1 coding region of isolate AC-50 was highly variable with an identity range of 48.9-70.9% with other LYSV isolates and a deletion of 204 nt compared to Japanese and Australian isolates. The ratio of nonsynonymous (*d*N) and synonymous (*d*S) polymorphic sites suggested that purifying selection dominates in the evolution of LYSV and the mean *d*N/*d*S ratio was highest for P1, confirming that this coding region is under less evolutionary constraints.

Keywords: Leek yellow stripe virus, Sequence analysis, PIPO, P1 gene, variability, purifying selection.

Garlic (*Allium sativum* L.), an economically important horticultural crop, is grown around the world for its medicinal and culinary properties (Keusgen, 2002). Garlic stocks worldwide are infected with a number of viruses often designated as 'garlic viral complex'. Fourteen viruses belonging to five different taxa are identified to be associated with garlic worldwide, including eight allexiviruses, two potyviruses, two carlaviruses, one *Fijivirus* and one *Tospovirus* (Nam *et al.,* 2015; Katis *et al.,* 2012; Gawande *et al.,* 2010; Conci *et al.,* 2003). In India seven viruses belonging to four different taxonomic groups have been reported in garlic, include garlic virus X (GarV-X) and garlic virus A (GarV-A) from the genus *Allexivirus*; onion yellow dwarf virus (OYDV) and leek yellow stripe virus (LYSV) from the genus *Potyvirus*; garlic common latent virus (GarCLV) and shallot latent virus (SLV) from the genus *Carlavirus*; and iris yellow spot virus (IYSV) from the genus *Tospovirus*, respectively (Baranwal *et al.,* 2011; Gawande *et al.,* 2015; Gupta *et al.,* 2013; Majumder and Baranwal, 2014; Gawande *et al.,* 2010). Among these viruses, LYSV is economically important and has flexuous and filamentous particles containing a single-stranded, positive sense RNA genome. LYSV is perpetuated by asexual propagation of garlic (Parrano *et al.,* 2012) and is transmitted by aphids. The garlic virus complex can reduce garlic bulb weight by nearly three quarters and LYSV alone can result in approximately one-quarter reduction in bulb weight (Lunello *et al.,* 2007).

LYSV was first reported in leek by Bos *et al.* (1978) and in garlic by Walkey *et al.* (1987). In India, it was reported for the first time in 2013 from garlic cultivar AC-50 (Gupta *et al.,* 2013). In the present study, we report on the complete genome sequence of LYSV isolate AC-50 from garlic in India and its molecular characteristics.

Light yellow striping of the leaves and stunting were observed in 21 garlic cultivars in experimental fields of ICAR-Indian Agricultural Research Institute (ICAR-IARI) in New Delhi. Leaf samples of each of the symptomatic garlic cultivars were collected and tested for LYSV in double antibody sandwich ELISA (DAS-ELISA) and for OYDV in direct antigen coating ELISA (DAC-ELISA) using virus specific antibodies (Bioreba, Switzerland). DAS-ELISA showed four (AC-50, PGS-14, G-282 and JG-03-263) out of the 21 garlic cultivars positive for LYSV and all of them positive for OYDV. Immunosorbent electron microscopy of DAS-ELISA-positive samples with antibody

Corresponding author: V.K. Baranwal
Fax: + 91.11.25843113
E-mail: vbaranwal2001@yahoo.com

*Supplementary materials can be downloaded at:
http://XXXXXXXXXXXXX.XXXXXXXXXXX

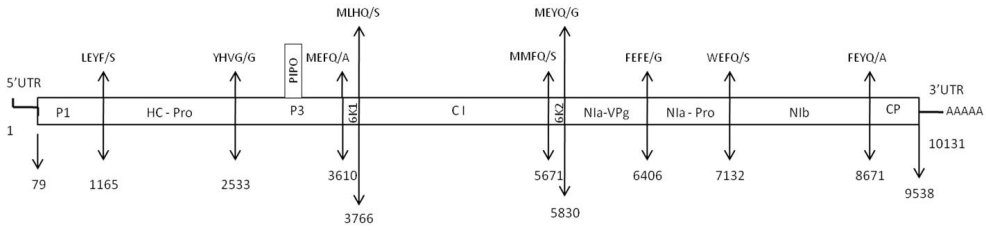


Fig. 1. Schematic representation of the genomic organization of LYSV isolate AC-50. The 5' and 3' nontranslated regions (NTR) are represented by lines and the large open reading frame (ORF) is depicted by an open box. The numbers below the diagram indicate the starting position (nt) predicted for each gene. The putative cleavage sites in the polyprotein are indicated above the genome. The putative protein PIPO is indicated by the small box above the main reading frame.

coating of fresh leaf samples was done to observe flexuous filamentous particles. A LYSV isolate originating from garlic accession AC-50 was selected for cloning and sequencing of the full-length viral genome using total RNA extracted from LYSV-infected leaves using an RNeasy Plant Mini kit (Qiagen, Germany) (Gupta *et al.,* 2013). The cDNA was synthesized using MuMLV reverse transcriptase (Takara, Japan) according to the manufacturer's protocol. The viral cDNA was amplified by polymerase chain reaction (PCR) with nine primer pairs designed throughout the LYSV genome sequence (Supplementary Table 1). PCR reactions were performed using upstream and downstream primers with a total of 5 µl of 10X PCR buffer including $MgCl_2$ (NEB), 1 µl of each primer (20 pmol/µl), 2.5 µl of 10 mM dNTPs (2.5 mM each), 0.5 µl Taq DNA polymerase (5 U/µl), and 5 µl of cDNA in a 50 µl final volume reaction. The amplification used the following thermocycling conditions: 35 cycles of 30 s at 94°C, 45 s at 58-60 °C (Supplementary Table 1), 1 min at 72°C followed by a final extension of 10 min at 72°C. The amplified fragments were cloned into pGEM-T and three clones of each segment were bi-directionally sequenced (Xcelris Genomics, India). The consensus sequences obtained for each fragment were assembled using CLUSTAL X. The genome sequence was determined from nine overlapping cDNA clones. 5' and 3' RACE was done to amplify the 5' and 3' untranslated region of the LYSV genome using ROCHE kit (Life Science Custom Biotech, United States) according to the manufacturer's protocol.

The complete genome of LYSV isolate AC-50 consisted of 10,131 nucleotides (nt), excluding the 3' terminal poly(A) tail (GenBank accession No. KP168261). The AUG codon located at nt position 79-81 is likely to be the translation initiation codon, since it was in a context (AATAUGGCT) that is similar to the context of the initiation codon of other isolates of LYSV. The genome consists of a single large open reading frame (ORF) of 9456 nt starting at position 79 and terminating with UGA at position 9535-9537, followed by a 3' untranslated region (UTR) of 594 nt. The ORF potentially encodes a polyprotein of 3152 aa with an estimated molecular weight of 357.64 kDa (http://web.expasy.org/compute_pi/). As predicted by Adams *et al.* (2005), nine putative protease cleavage sites were identified which give rise to the following ten putative mature proteins (Fig. 1): P1, HC-Pro, P3, 6K1, CI (cylindrical inclusion), 6K2, VPg (viral protein genome linked), NIa-Pro (nuclear inclusion a – proteinase), Nib (nuclear inclusion b) and CP (coat protein). Conserved motifs in each of these proteins were identified, including 271H-8X-D-33X-S314 and 335F-V-V-R-G339 in the P1 protein, 412K-I-T-C415 and 670P-T-K672 in HC-Pro (for aphid transmission), 1314G(X)2GXGKST1322 and 1403D-E-C-H1406 in the CI (for helicase function) protein, 2155HX34-DX67GXC-GX14H2276 in NIa, 1943G-D-D1945 (for replicase activity) and 2596CDADGS2601 in the Nib, and 2870DAG2872 in CP protein (for aphid transmission) (Shukla *et al*., 1994; Koonin, 1991; Kadare and Haenni, 1997; Atreya *et al*., 1995; Urcuqui-Inchima *et al.,* 2001; Pirone and Blanc, 1996). The recently described ORF coding the putative protein PIPO (Chung *et al.,* 2008) was identified within the P3 region encoded in a +2 reading frame starting from a G1A7 motif at nt 3022 and ending with a UGA termination codon at nts 3192-3194. The predicted protein is 56 aa long without the P3 N-fused region and shares a high degree of sequence identity with the PIPO protein of a LYSV isolate from Spain (JX429965). The highly conserved potybox 'b' (UCAAGCA) exists in the 5'-UTR but the potybox 'a' (AUAACAU) was not found in the genomic sequence of LYSV isolate AC-50.

Comparative sequence analyses revealed that LYSV isolate AC-50 shared 73.9-79.9% and 79.1-87.2% sequence identities at the nt and aa levels, respectively, with other LYSV isolates. The highest sequence identity of 79.9 and 87.2% at the nt and aa levels was with a LYSV isolate from Australia (HQ258895). Highest CP sequence identity at the nt and aa levels was 86.8% and 82.1% with isolates from Australia (HQ258895) and China (AJ307057), respectively. The P1 coding region of LYSV isolate AC-50 showed maximum 70.9% and 62.9% identity at nt and aa level, respectively, with an isolate from China (AJ307057). Like for other potyviruses, the 5'UTR and P1 regions of the LYSV genome are highly variable, and 22.5-74.4% and 42.7-62.9% sequence identities were observed at the nt and aa levels, respectively. The CI and NIa-Pro regions were the most conserved with 93.3-96.6% and 88-95.4% sequence identity at the aa level, respectively. The 3'UTR and other functional regions of the polyprotein of LYSV had 70-93

Table 1. Characteristics of the polyprotein of LYSV isolate AC-50 and its predicted cleavage products.

Untranslated regions and functional regions of polyprotein	Size (nt)	Amino acids	Predicted cleavage site	% sequence identity with isolates from 11 available LYSV sequence from NCBI (nt and aa)
5'UTR	1-78 (78)	--	LEYF/S	22.5-74.4 (-)
P1	79-1164 (1086)	362	YHVG/G	48.9-70.9 (42.7-62.9)
HC-Pro	1165-2532 (1368)	456	MEFQ/A	75.5-78.6 (82-84.6)
P3	2533-3609 (1077)	359	MLHQ/S	70.1-79.6 (64.3-83.5)
6K1	3610-3765 (156)	52	MMFQ/S	75.6-83.9 (90.3-100)
CI	3766-5670 (1905)	635	MEYQ/G	79.6-83.2 (93.3-96.6)
6K2	5671-5829 (159)	53	FEFE/G	71.6-83.6 (73.5-92.4)
VPg	5830-6405 (576)	192	WEFQ/S	74.3-80.5 (80.2-90.1)
NIa-Pro	6406-7131 (726)	242	FEYQ/A	76.8-82.2 (88-95.4)
NIB	7132-8670 (1539)	513	--	78.5-81.6 (87.5-92.5)
CP	8671-9534 (864)	288	--	78.0-82.1 (81.6-86.8)
3'UTR	9538-10131 (594)	--		77.3-90.1 (-)

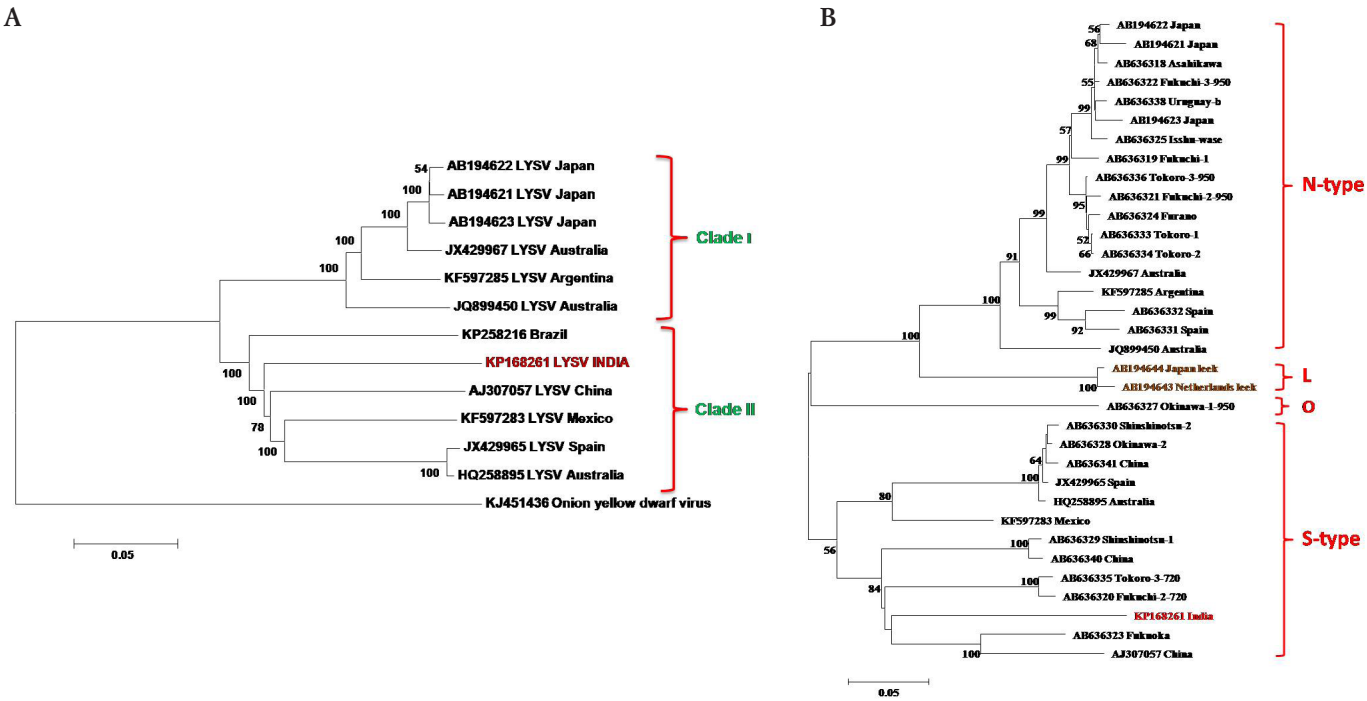


Fig. 2. Phylogenetic analysis of LYSV isolates in the full-length genome sequence (A) and in the 5'UTR-P1 region (B) using the Neighbour-Joining algorithm. The evolutionary distances were computed using p-distance method with 1000 bootstrap replicates. Only >50 % bootstrap values are shown here. The scale bar indicates the number of substitutions per site. Onion yellow dwarf (OYDV) was used as an outgroup.

% sequence identity at both the nt and aa levels when compared with other LYSV isolates (Table 1).

Alignment of the full-length sequence of 12 LYSV isolates including isolate AC-50 was performed using the Muscle programme included in the MEGA6 (http://www.megasoftware.net/) software package (Tamura *et al.,* 2013). The evolutionary distances were computed using the p-distance method with 1000 bootstrap replications. Phylogenetic analysis showed that LYSV isolate AC-50 groups with isolates from Australia (HQ258895), Spain (JX429965), Mexico (KF597283), Brazil (KP258216) and China (AJ307057), whereas two Australian isolates (JX429967 and JQ899450), one Argentinian isolate (KF507285) and three Japanese isolates (AB194621, AB194622 and AB194623) formed a separate clade (Fig. 2A). Phylogenetic analysis also confirmed that the LYSV 5'UTR-P1 nt sequences from garlic formed two main groups, the N-type (including isolates from Northern Japan and some from Uruguay and Australia) and the S-type (including isolate AC-50 from India and isolates from Southern Japan, China, Korea and other Australian isolates) (Bampi *et al.,* 2015; Chen *et al.,* 2001; Takaki *et al.,* 2005; Yoshida *et al.,* 2012) (Fig. 2B). In addition, phylogenetic analysis of the CP coding region of all LYSV isolates also divided the LYSV in two major groups (Supplementary Fig. 1).

Split decomposition analysis of LYSV sequences was also performed using SplitsTree v4.14.2 with default parameters (Huson and Bryant, 2006) to suggest the evolutionary history of LYSV isolates. Split divided the LYSV full genome sequences into two major broad groups with isolate

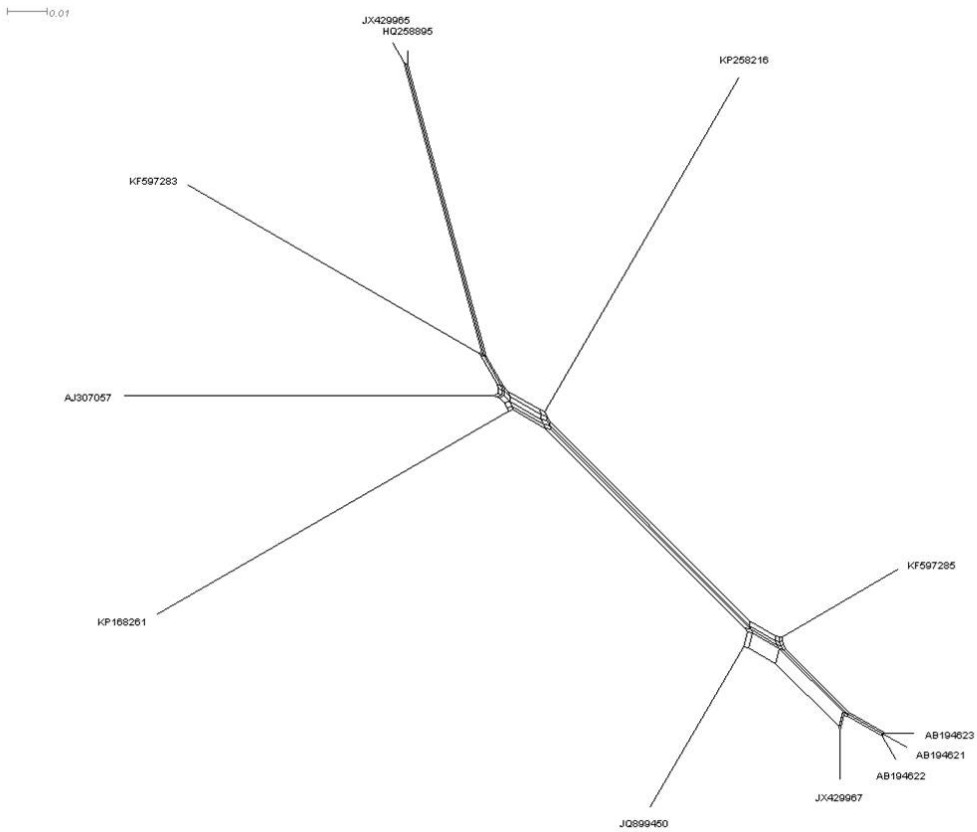


Fig. 3. Phylogenetic network of the full-length genome sequence of 12 LYSV isolates constructed by split-decomposition analysis.

AC-50 more closely related to Chinese isolate (AJ307057) (Fig. 3).

The rates of synonymous (*d*S) and non-synonymous (*d*N) substitutions per site were estimated from codon aligned gene wise nucleotide sequences in the recombination- aware algorithm for single likelihood ancestor counting (SLAC), using the HKY85 nucleotide substitution bias model at p value of 0.2 in Datamonkey webserver (http://www.datamonkey.org/) (Delport *et al.,* 2010). Gene-wise datasets of all 12 LYSV full genomes were analyzed for substitution rates. Mean values of *d*N/*d*S for all the genes were calculated for all the data sets separately and the tree constructed by Datamonkey software for each of the coding region was used to compute gene-specific mean *d*N/*d*S (Supplementary Table 2). The overall values of < 1 indicated that all the LYSV genes are under purifying selection in LYSV (Supplementary Table 2). The mean *d*N/*d*S ratios were higher for P1 (0.49), CP (0.19) and P3 (0.16) and lower in 6K1 (0.03) and CI (0.03) genes. The results indicated that the selection pressure on P1, CP and P3 was higher compared to other coding regions of the LYSV genome. The higher selection pressure on P1 might provide an evolutionary force for host adaptation (Valli *et al.,* 2008).

The study provided insights into the relatedness of LYSV isolate AC-50 from garlic in India with other LYSV isolates from other countries and on the evolutionary forces driving genetic diversity. It is anticipated that the information generated in this study will help design molecular detection techniques for the identification of infected garlic plants and their quick removal for the production of healthy garlic plants.

ACKNOWLEDGEMENTS

The funding from ICAR outreach project is gratefully acknowledged. First author is grateful to Department of Science and Technology for INSPIRE fellowship during Ph.D. Programme. Authors are also thankful to Head Division of Plant Pathology and Director ICAR-Indian Agricultural Research Institute for necessary lab facilities.

REFERENCES

Adams M.J., Antoniw J.F., Beaudoin F., 2005. Overview and analysis of the polyprotein cleavage sites in the family Potyviridae. *Molecular Plant Pathology* **6**: 471-487.

Atreya P.L., Lopez-Moya J.J., Chu M., Atreya C.D., Pirone T.P., 1995. Mutational analysis of the coat protein N-terminal amino acids involved in potyvirus transmission by aphids. *Journal of General Virology* **76**: 265-270.

Bampi D., Mituti T., Pavan M.A., Hammond J., Krause-Sakate R., 2015. *Leek yellow stripe virus* isolates from brazil form a distinct clade based on the p1 gene. *Journal of Plant Pathology* **97**: 457-463.

Baranwal V.K., Singh P., Jain R.K., Joshi S., 2011. First report of garlic virus X infecting garlic in India. *Plant Disease* **95**: 1197.

Bos L., Huijberts N., Huttinga H., Maat D.Z., 1978. *Leek yellow stripe virus* and its relationship to *Onion yellow dwarf virus*; characterization, ecology and possible control. *Netherlands Journal of Plant Pathology* **84**: 185-204.

Chen J., Chen J., Adams M.J., 2001. Molecular characterisation of a complex mixture of viruses in garlic with mosaic symptoms in China. *Archives of Virology* **146**: 1841–1853.

Chung B.Y.W., Miller W.A., Atkins J.F., Firth A.E., 2008. An overlapping essential gene in the Potyviridae. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 5897-5902.

Conci V.C., Canavelli A.E., Lunello P., Dirienzo J., Nome S.F., Zumelzu G., Italia R., 2003. Yield losses associated with virus-infected garlic plants during five successive years. *Plant Disease* **87**: 1411-1415.

Delport W., Poon A.F., Frost S.D.W., Pond S.L.K., 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**: 2455-2457.

Gawande S.J., Khar A., Lawande K.E., 2010. First report of *Iris yellow spot virus* on garlic in India. *Plant Disease* **94**: 1066.

Gawande S.J., Gurav V.S., Ingle A.A., Gopal J., 2015. First report of *Garlic virus A* in garlic from India. *Plant Disease* **99**: 1288.

Gupta N., Prabha K., Islam S., Baranwal V.K., 2013. First report of *Leek yellow stripe virus* in India from garlic. *Journal of Plant Pathology* **95**: 69-77.

Huson D. H., Bryant D., 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254-267.

Kadare G., Haenni A.L., 1997. Virus-encoded RNA helicases. *Journal of Virology* **71**: 2583-2590.

Katis N.I., Maliogka V.I., Dovas C.I., 2012. Viruses of the genus *Allium* in the Mediterranean region. *Advances in Virus Research* **84**: 163-208.

Keusgen M., 2002. Health and Alliums. In: Rabinowitch H.D., Currah L. (eds). Allium Crop Science: Recent Advances, pp. 357–378. CAB International, Collingwood, UK.

Koonin E.V., 1991. The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**: 2197-2206.

Lunello P., Rienzo J.D., Conci V.C., 2007. Yield loss in garlic caused by *Leek yellow stripe virus* Argentinean isolate. *Plant Disease* **91**: 153-158.

Majumder S., Baranwal V.K., 2014. Simultaneous detection of four garlic viruses by multiplex reverse transcription PCR and their distribution in Indian garlic accessions. *Journal of Virological Methods* **202**: 34-38.

Nam M., Lee Y.H., Park C.Y., Lee M.A., Bae Y.S., Lim S., Lee J.H., Moon J.S., Lee S.H., 2015. Development of multiplex RT-PCR for simultaneous detection of garlic viruses and the incidence of garlic viral disease in garlic genetic resources. *Plant Pathology Journal* **31**: 90-96.

Parrano L., Afunian M., Pagliaccia D., Douhan G., Vidalakis G., 2012. Characterization of viruses associated with garlic plants propagated from different reproductive tissues from Italy and other geographic regions. *Phytopathologia Mediterranea* **51**: 549-565.

Pirone T.P., Blanc S., 1996. Helper-dependent vector transmission of plant viruses. *Annual Review of Phytopathology* **34**: 227-247.

Shukla D.D., Ward C.R., Brunt A.A., 1994. The Potyviridae, pp. 516. CAB International, Wallingford, UK.

Takaki F., Sano T., Yamashita K., Fujita T., Ueda K., Kato T., 2005. Complete nucleotide sequences of attenuated and severe isolates of *Leek yellow stripe virus* from garlic in northern Japan: Identification of three distinct virus types in garlic and leek world-wide. *Archives of Virology* **150**: 1135-1149.

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.

Urcuqui-Inchima S., Haenni A.L., Bernardi F., 2001. Potyvirus proteins: a wealth of functions. *Virus Research* **74**: 157-1755.

Valli A., Dujovny G., García J.A., 2008. Protease activity, self-interaction and small interfering RNA binding of the silencing suppressor p1b from *Cucumber vein yellowing ipomo virus*. *Journal of Virology* **82**: 974-986.

Walkey D.G.A., Webb M.J.W., Bolland C.J., Miller A., 1987. Production of virus-free garlic (*Allium sativum)* and shallot (*A. ascalonicum*) by meristem tip-culture. *Journal of Horticultural Science* **62**: 211-220.

Yoshida N., Shimura H., Yamashita K., Suzuki M., Masuta C., 2012. Variability in the P1 gene helps to refine phylogenetic relationships among *Leek yellow stripe virus* isolates from garlic. *Archives of Virology* **157**: 147-153.

Received May 20, 2017
Accepted June 12, 2017

COMPLETE GENOME SEQUENCE OF AN ISOLATE OF LEEK YELLOW STRIPE VIRUS FROM GARLIC IN INDIA

N. Gupta, S. Islam, S.K. Sharma and V.K. Baranwal

Supplementary data

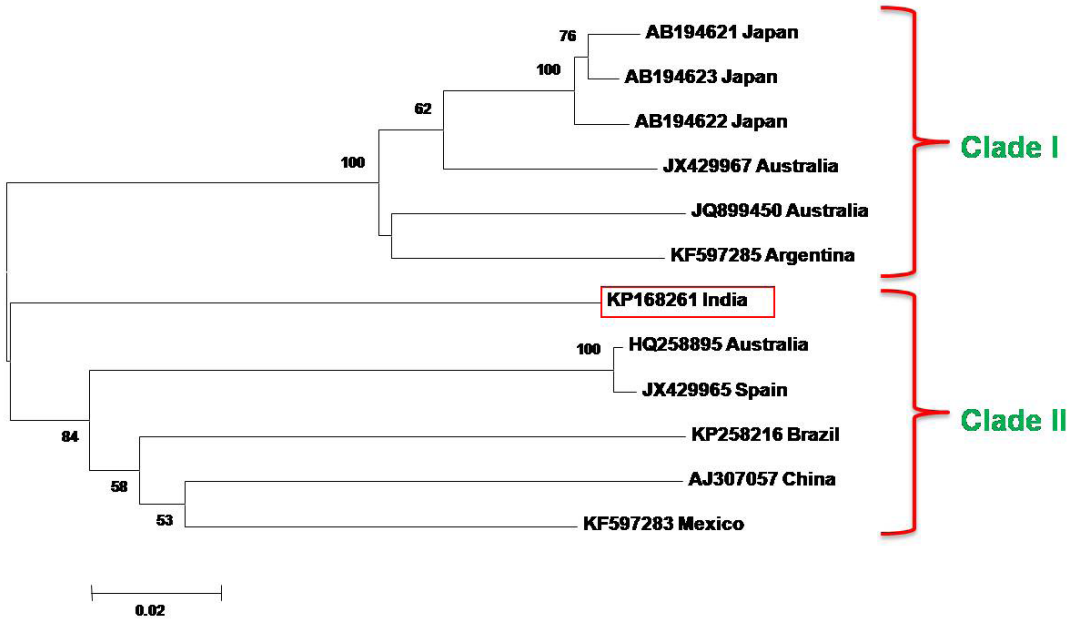
Supplementary Table 1. Primer sets used for PCR amplification of full-length genome of LYSV isolate AC-50.

Set No.	Primers and their sequences	Length	Annealing Temp.	Expected amplicon size
1.	LYSV CP FULL F-GCTGGTGAGGAGATTGATG LYSV CP FULL R-CTGCATATGCGCACCATC	19 18	58°C	~ 864 bp
2.	LYSV HCPRO F-ACCAGATCCGTCTTTGGTCA LYSV HCPRO R-CTTCGTATAGCGRACRACTTC	20 21	60°C	~1000 bp
3.	LYSV CI F-GCCTTTCGACAAGCTAAATGAC LYSV CI R-GGTGTYGCTGAAACYTTRAG	22 20	60°C	~2840 bp
4.	LYSV F-GTTTGATGAGTGTCATTGTTT LYSV R-CTTGCCAGTGCTATTCTTTTG	21 21	60°C	~1100 bp
5.	LYSV NIA F-AAATTGTACGAAAGTGAGTTG LYSV NIA R-CATCCAATCCTTGTTACCTG	21 20	60°C	~1700 bp
6.	LYSV NIB F-TGTTGGGATGGTATCACACT LYSV NIB R-CCCATGTATGCTAGTCTGTG	20 20	56°C	~1800 bp
7.	LYSV P1 F-ATCTCAACACAACTTATGCAA LYSV P1 R-GCACACAACGACCGTCCAATCT	21 22	60°C	~1000 bp
8.	LYSV 3' F INNER GCACACAGCACATGACGTG LYSV 3' F OUTER CTCACGCACAAATGAAAGCAG	19 21	A/C to manufacturer's protocol	~670 bp
9.	LYSV 5' R INNER CTGTTCTGAACCTTTCCATAGCC LYSV 5' R OUTER CCTCATCACAATCGTGCTCCTT	23 22	A/C to manufacturer's protocol	~260 bp

Supplementary Table 2. Estimates of selection pressure acting on the coding regions of LYSV.

Coding region	Position (codons)	Normalized *d*N/*d*S[a]		Selection pressure					
		Log (L)	Mean	Positive		Negative		Neutral	
				N	(%)	N	(%)	N	(%)
P1	430	–12588.7	0.49169	9	2.093	54	12.55	367	85.34
Hc-Pro	456	–6407.25	0.10492	0	0	253	55.48	203	44.51
P3	359	–5322.56	0.16904	1	0.27	152	42.33	206	57.38
6K1	52	–573.588	0.0312	0	0	29	55.76	23	44.23
CI	635	–7526.08	0.0342	0	0	407	64.09	228	35.90
6K2	53	–695.71	0.1154	0	0	23	43.39	30	56.60
Vpg	192	–2467.14	0.06633	0	0	110	57.29	82	42.70
Nia-Pro	242	–2954.45	0.04186	0	0	147	60.74	95	39.25
Nib	513	–6371.33	0.05376	0	0	312	60.81	201	39.18
CP	289	–5004.12	0.1957	0	0	86	29.75	203	70.24

[a]Mean *d*N/*d*S values < 1 indicate negative or purifying selection, *d*N/*d*S values = 1 indicate neutral selection, *d*N/*d*S values > 1 indicate positive selection for each coding region-specific dataset. The LYSV dataset include different LYSV genes including Indian isolates. Acronyms and accession numbers of sequences are mentioned in manuscript (*d*N = non synonymous and *d*S=synonymous polymorphic sites).



Supplementary Fig. 1. Phylogenetic relationship of 12 LYSV isolates in the coat protein coding region based using the Neighbour-Joining algorithm. The evolutionary distances were computed using the p-distance method with 1000 bootstrap replications. Only >50% bootstrap values are shown. The scale bar indicates the number of substitutions per site.