

Phylogenetic and diversity analyses revealed that leek yellow stripe virus population consists of two types: S and N

Adyatma Irawan Santosa (✉ adyatma_santosa@yahoo.com)

Ankara University Faculty of Agriculture: Ankara Universitesi Ziraat Fakultesi <https://orcid.org/0000-0002-2826-5444>

Filiz Randa Zelyüt

Ankara University Faculty of Agriculture: Ankara Universitesi Ziraat Fakultesi

Filiz Ertunç

Ankara University Faculty of Agriculture: Ankara Universitesi Ziraat Fakultesi

Research Article

Keywords: diversity analysis, evolutionary analysis, leek yellow stripe virus, phylogenetic analysis

Posted Date: April 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-276578/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Virus Genes on November 8th, 2022. See the published version at <https://doi.org/10.1007/s11262-022-01956-y>.

Abstract

Leek yellow stripe virus (LYSV, *Potyvirus*), a pathogen affecting *Allium* spp. worldwide, has been suspected to consist of two types: S and N, based on genetic and host species differences. In this study, phylogenetic and evolutionary analyses were performed to P1 and CP regions of genome of global LYSV isolates to clarify the variation among members of S-type and N-type. Constructed phylogenetic trees clearly divided isolates into S-type and N-type, with N-type was further split into L and N groups, according to hosts. Significant nucleotide (nt) and amino acids (aa) sequence variation were observed on full ORF, P1, HC-Pro, P3, VPg coding regions. The dN/dS values of P1 and CP confirmed that both genes are under strong negative selection pressure. Neutrality tests on Eastern Asian isolates argued that ancestor of current LYSV isolates may had evolved with garlic while they were in Asia before spread to other world regions through garlic propagative materials. The genetic differentiation and gene flow analysis showed that there was very frequent gene flow from S-type to L and N groups and these phylogroups differentiated from each other over time. Host differences, substantial nt and aa variation, inconsistent serological test results, and phylogenetic and diversity analyses results highly suggested that LYSV can be separated into two types: S and N.

Main Text

Among viruses affecting *Allium* spp. cultivation, leek yellow stripe virus (LYSV, *Potyvirus*) is of particular interest since it has been identified worldwide as an important virus on leek (*Allium ampeloprasum* L.), garlic (*Allium sativum* L.), and onion (*Allium cepa* L.) [1-4]. The virus genome sequence is *ca.* 10,131 nucleotide (nt) long which, typical to *Potyvirus*, contain a single open reading frame (ORF) of *ca.* 9,456 nt encoding a polyprotein of *ca.* 3,152 amino acid (aa) [5]. P1 protein as the most variable among potyvirus-encoding proteins, is important in evolutionary study of *Potyvirus* species [6]. CP is regarded as the most conserved region in *Potyvirus* genome and commonly used in phylogenetic studies of its species [7].

A small evolutionary study involving 12 fully sequenced isolates found that selection pressures on P1, CP and P3 were higher than other coding regions of LYSV genome [5]. Still, very little is known about the virus genetic diversity. Therefore, older and latest addition of nt sequences from diverse regions of world and different *Allium* spp. into NCBI GenBank were analyzed in this paper to get the first detailed insight into global genomic diversity, evolution, and population structure of LYSV. The analyses on P1 and CP genes largely followed [8] study which clustered isolates into two major lineages (S-type and N-type). Similarity percentages on other genomic regions were also investigated here to determine wheather S-type and N-type isolates could also be differentiated based on genetic and host species variation.

Nucleotide sequence of P1 gene is 1182 nt long (encoding a protein of 362 aa), covering position 91-1086, while CP gene is 864 nt long (encoding a protein of 288 aa), covering position 8683-9546 in LYSV genome; reference isolate: accession no. AJ307057. However, large variation in nt and aa lengths exist in some tested isolates.

Forty-six isolates with complete and 40 isolates with partial nt sequences of P1 (for a total of 86 isolates) and 121 isolates with complete nt sequences of CP were obtained from NCBI GenBank (Online Resource 1, Table 1). Each of P1 and CP sequences was aligned together using ClustalW with default parameters in MEGA X v.10.2.4 (megasoftware.net). Recombination signal on each of P1 and CP sequences alignment was scanned by Recombination Detection Program (RDP v.4.56). Recombination event should be supported by at least five of RDP, GENECONV, BootScan, MaxChi, ChiMaera, SiScan, and 3Seq algorithms with Bonferroni-corrected *P* value of < 0.05 to be considered credible [9]. Only results of recombination analysis on CP gene sequences were reported since analysis on the highly variable P1 gene sequences did not give reliable result.

Construction of phylogenetic trees based on each of P1 and CP nt sequences were carried out in MEGA X software using Maximum Likelihood statistical method, with uniform rates among sites and use all sites treatment of gaps/missing data [23]. Kimura-2 parameter was applied for each data set, with 1000 bootstrap replicates to test statistically significance of isolate clusters [24]. Estimations of nucleotide (nt) and amino acid (aa) similarity percentages among phylogroups in different genome regions and full ORF were done by Sequence Demarcation Tool (SDT) v1.2 software [25].

The mean evolutionary distance of LYSV of P1 and CP genes within and between phylogroups were calculated using Kimura-2 parameter [24] in MEGA X software, estimates of standard error (SE) obtained by a bootstrap procedure (1000 replicates).

Differences in each of P1 and CP gene aa sequence were examined using BioEdit v.7.2.5 [26]. Forty-six isolates which have complete sequence were aligned together for P1 gene analysis. In CP gene alignment, up to three isolates were selected as representative of their respective group or subgroup.

Genetic diversity and polymorphism analysis were done using DnaSP v.6.12.03 [27]. A gene is under positive (diversifying), neutral, and negative (purifying) selection when dN/dS (ω ratio) is > 1, = 1 and < 1, respectively [27]. The three 'Neutrality tests' (Fu and Li's F^* , Fu and Li's D^* , and Tajima's D) were carried out by DnaSP v.6.12.03 (default window length of 100 sites and step size of 25 sites; without out-group) [27]. K_S^* , K_{ST}^* , Z^* , Snn and F_{ST} values [28-29] which determined genetic differentiation among each of P1 and CP populations were calculated by DnaSP v.6.12.03 [27]. K_{ST}^* will be near zero if there is no genetic differentiation (null hypothesis) [30]. The smaller Z^* means smaller genetic differentiation among population [28]. The value of Snn describes a range of exact same population (value of 0.5) (null hypothesis) to distinctly differentiated population (value of 1) [29]. The null hypothesis in K_S^* , K_{ST}^* , Z^* , and Snn is rejected by a significant *P* value [28-30]. F_{ST} ranges between exact same population (value of 0) to fully distinct populations (value of 1) [28, 30]. $F_{ST} > 0.25$, in most cases, is indicating a large gene flow and big genetic differentiation in the tested populations [31].

Studies on LYSV genetic variation mainly were limited on construction of phylogenetic trees [4, 10, 17, 19]. Moreover, hosts difference and large genetic variances were observed among its phylogroups [2, 8].

Therefore, recombinant-free isolates were analyzed in this paper to learn more about molecular diversities and evolution of LYSV while also presenting evidences that its population consist of two distinct types.

Recombinant signals in CP gene of six isolates were detected by at least five statistical methods which supported by Bonferroni-corrected P value of < 0.05 . Recombinant events in isolate no. JX429967 occurred in two different regions (Online Resource 1, Table 2). These results presented the first detection of recombinant events on LYSV genome. Isolates with multiple parents were omitted from further analyses, to give more accurate phylogenetic analysis.

The both constructed phylogenetic trees clustered all isolates into three major groups (S, L, and N). Members of S group belong to S-type whereas members of L and N groups belong to N-type (Figure 1). P1 gene based phylogenetic trees by [8] also clustered isolates into two major lineages (S-type and N-type). S-type was further split into S and O groups, while N-type into N, U, Sp and L groups. However, this phylogrouping was not used widely since P1 is not commonly targeted in molecular detection of LYSV, and the high variations in P1 would potentially generate new groups every time new isolates were characterized. So, since P1 gene based phylogenetic tree of our study clustered isolates in similar manners, a new phylogrouping was proposed in which it simply clustered isolates into S-type and N-types. N-type was further split according to host differences into N (garlic isolates) and L (leek isolates) groups. Since this new P1 phylogrouping is also applicable in CP (each of tested isolates was in identical group in both P1 and CP trees), future LYSV isolates phylogrouping should be based on CP, which is more conserved and much more frequently sequenced in LYSV characterization works.

S-type of CP phylogenetic tree was further divided into 11 subgroups (S1-S11) based on nt changes and deletions in compared sequences, but these subgroups cannot be applied in S-type of P1 since members of different CP subgroups formed same branch in the P1 tree (Figure 1). S1 isolates experienced deletion in position 121-123 of their nt sequences, resulted in a single residue deletion in position 41 of their aa sequences. S7 and S9 isolates in position 124-126 of their nt sequences, resulted in deletion in position 42 of their aa sequences. S11 isolates in position 100-108 of their nt sequences, resulted in three residue deletions in position 44-46 of their aa sequences. DAG motif that is important in potyviruses transmission by aphids was conserved in CP of all tested isolates except no. AJ409307, which changed into DSG (Online Resource 1, Figure 1). Major aa changes in P1 gene happened throughout its coding sequence. Some isolates experienced large nt deletions which resulted in high aa length variations among compared isolates (Online Resource 1, Figure 2).

Identity analysis revealed very large variation in nt and aa sequences of P1 gene even among S-type isolates. In CP comparison, nt and aa identity percentages among S-type isolates were shown to be lower than that of among N-type isolates. S-type isolates also had high nt and aa differences to N-type isolates. In 115 complete CP sequences comparison, the lowest nt similarity percentage among S-type and N-type isolates (76.6%) was very close to, and the lowest aa similarity percentages (75.4%) was well below than that of proposed for *Potyvirus* species demarcation: 76% and 80% for nt and aa, respectively. The further comparisons on 46 complete ORF sequences determined that nt similarity percentages among S-type

and N-type isolates in P1, HC-Pro, P3, VPg coding regions, and full ORF were also lower than limits for *Potyvirus* species demarcation [7] (Table 1).

Besides that, Immunocapture-Reverse Transcriptase-Polymerase Chain Reaction (IC-RT-PCR) and standard RT-PCR has been shown to be 10^2 - 10^4 times more sensitive than Double-Antibody Sandwich-Enzyme-Linked Immunosorbent Assay (DAS-ELISA) in LYSV detection [32]. [19] reported that DAS-ELISA results for detection of isolates no. GQ475411-13 (N-type) were erratic and unreliable, while DAS-ELISA results for onion yellow dwarf virus (OYDV, *Potyvirus*) and garlic common latent virus (GarCLV, *Carlavirus*) in the same study were consistent and reliable. DAS-ELISA was also suggested to be less sensitive than RT-PCR in detection of isolates no. MN070130-31, MN864794-95 (N-type) [4]. These phenomena could be caused by the antibodies for those DAS-ELISAs were developed from CP of S-type isolates which have large sequence variations to N-type isolates.

The overall evolutionary distance value of P1 was 0.468 ± 0.022 ; and the main within lineages values were 0.411 ± 0.020 (S-type) and 0.148 ± 0.010 (N-type). Also, the value between phylogroups was 0.604 ± 0.032 (S-type/N-type). The overall mean evolutionary distance value of CP was 0.203 ± 0.011 ; and the values within main lineages were 0.190 ± 0.010 (S-type) and 0.114 ± 0.008 (N-type). In addition, the evolutionary distance between the main phylogroups was 0.237 ± 0.013 (S-type/N-type).

Within all isolates, the mean evolutionary distance value of P1 was significantly higher than that of CP, which may indicate that P1 gene region plays active roles in the formation of new types/variants through host selection and environmental adaptation in evolutionary processes. In addition, the evolutionary distance values obtained for both gene regions are particularly high within S-type lineage. The distance value of S-type from N-type was also high, suggesting large genetic variation between the two types. The high evolutionary distance values of S-type from N-type may be related to host plant selection or geographic adaptations, and these results may confirm the emergence of phylogroups over time.

In general, π values were higher among phylogroups and continental populations in CP gene region compared to the values obtained from P1 gene region. In P1, the highest π were 0.29896 (S-type) and 0.36447 (Africa) among phylogroups and continents populations, respectively. The lowest π were 0.02358 (L group) and 0.24684 (America). However, the low π of L population may be due to this phylogroup having only two isolates. Hd values which provide important information for predicting evolutionary processes, reached 1.000 for all phylogroups and continental populations. S, k, and h values were all obtained in descending order for S-type, N-type, N group, and L group. Meanwhile, the values obtained for those parameters, except k, among different continents were East Asia, America, Africa, Europe, and Oceania populations in descending order (Online Resource 1, Table 3). These results may indicate that there are high genetic diversities among global LYSV populations.

In analysis for CP gene, π ranged from 0.05075 (N group) to 0.17521 (Oceania). Although the number of isolates in N group was large, its π was very low. The highest π for phylogroup and continental populations were 0.16286 (S-type) and 0.17521 (Oceania), respectively. For Hd metric, high values

varying between 0.0972 and 1.000 were obtained for all phylogroups and continental populations (Online Resource 1, Table 3).

The obtained ω values were all < 1 , showing that P1 and CP genes were both under negative (purifying) selection pressures. The dN/dS ratios between phylogenetic groups and continental populations ranged from 0.28332 to 0.49867 for P1 gene, and between 0.09605 and 0.16526 for CP gene (Online Resource 1, Table 3). The evolutionary rates of P1 and CP genes were estimated to be 0.42 and 0.15, respectively, which were similar to that of obtained by [5]: 0.49 and 0.19, respectively. The results confirmed that both genes are under strong negative (purifying) selection pressure, with CP gene is under stronger pressure compared to P1 gene.

Fu and Li's F^* & D^* assigned statistically non-significant negative or positive values in P1 and CP of all populations. Thus, the results demonstrated that global LYSV population is undergoing balancing selection or bottlenecks. Positive values with statistical significance were only predicted by Tajima's D test for overall populations, S-type, and East Asia populations in both P1 and CP (Online Resource 1, Table 4). These positive with statically significant results indicated that they emerged from an excess of intermediate frequency alleles, and can evolve out of population balancing selection, bottlenecks, or structures.

S-type which consists largely of isolates from East Asian countries, and if it represents strained isolates, may have been endemic in East Asia and infected different hosts at different times in their life cycle. Although S-type isolates showed host selection, so far, onion was not found to be a host for them, as they were detected mostly in garlic and some (no. DQ925452-53) in leeks (Online Resource 1, Table 1). It is assumed that the genetic center of garlic was in Central Asia, and then spread to two major destinations: China and India, and Europe and Russia [33]. In our study, neutrality tests performed on Eastern Asian populations (China, Japan, South Korea, Taiwan, and Vietnam) gave positive and statistically significant results. Ancestor of current LYSV isolates may have evolved with garlic while they were in Asia, and then spread to different continents of the world through garlic propagation materials, as also observed in a study by [8]. The extensive adaptation of S-type isolates to garlic as a host was also confirmed by phylogenetic analysis of our study.

Gene flow and genetic differentiation analysis of LYSV populations revealed that all phylogroups (S-type, L group, N group, and N-type) in both P1 and CP genes had statistically significant K_S^* , K_{ST}^* and Z^* values. K_{ST}^* results indicated statistically significant genetic differences between lineages for both P1 and CP genes, and all of these values were greater than 0. In addition, K_{ST}^* values for both genes differed in continental populations. Statistically significant values were obtained between some continents, while some others were not significant, and some values were less than 0 (Online Resource 1, Table 5).

Among LYSV populations based on P1 and CP genes, significant Z^* values were calculated between all main lineages (S-type, L group, N group, N-type) populations. Furthermore, the Z^* values calculated for CP gene were significantly higher than P1 gene. In particular, Z^* value of S-type was statistically significant

and higher than other lineages, for both gene regions. Statistically significant results were obtained from F_{ST} values for all phylogroups in the two genes, and these values mostly were 1 or close to 1. F_{ST} (fixation index) values between main lineages for both gene regions were greater than 0.25 (Online Resource 1, Table 5).

The genetic differentiation and gene flow analysis performed between main nodes (S-type, L group, and N group) results may reveal that phylogroup populations for P1 and CP genes are genetically different from each other and also show that the reliability rates of dendograms created for both genes were high. They also showed that there was very frequent gene flow from S-type to L and N groups (N-type), and these phylogroups differentiated from each other over time. Therefore, these analyses may support conclusion by [8] that the sequences missing in P1 of S-type were present in the common ancestor of S-type and N-type.

Host differences, substantial nt and aa variation, inconsistent serological test results, and phylogenetic and diversity analyses results strongly indicated that LYSV should be differentiated into two types: S and N. Recognizing these two types, and other new variants that might emerge among LYSV isolates, could lead to development of more accurate diagnostic tools which in turn could help to control the virus epidemiologically.

Declarations

Funding

Authors are grateful to Scientific Research Projects Coordinator of Ankara University (Project No. 18H0447001) for the funding of this study.

Conflict of interest

The authors declare that they have no conflict of interest.

Availability of data and material

The authors declare that the data supporting the findings of this study are available within the article (and/or) its supplementary materials.

Ethics approval

This article does not contain any studies with human participant or animal performed by any of authors.

Authors' contributions

All authors contributed to the study conception and design. AIS performed phylogenetic analysis and wrote the manuscript. FRZ performed diversity and evolution analyses, and wrote the manuscript. FE supervised the analyses, and revised the manuscript. All authors read and approved the final manuscript.

References

1. Lunello P, Ducasse DA, Helguera M, Nome SF, Conci VC (2002) An Argentinean isolate of *Leek yellow stripe virus* from leek can be transmitted to garlic. *J Plant Pathol* 84(1):11–17. <https://doi.org/10.4454/jpp.v84i1.1082>
2. 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. *Arch Virol* 150(6):1135–1149. <https://doi.org/10.1007/s00705-004-0482-9>
3. Ayed C, Marais A, Faure C, Candresse T, Hamdi I, Najar A, Al-Mohandes Dridi B (2019) First report of *Onion yellow dwarf virus*, *Leek yellow stripe virus*, and *Shallot latent virus* on garlic (*Allium sativum*) in Tunisia. *Plant Dis.* 103(8):2143. <https://doi.org/10.1094/PDIS-09-18-1618-PDN>
4. Santosa AI, Ertunc F (2020) Identification, molecular detection and phylogenetic analysis of four viruses infecting *Allium cepa* in Ankara province. *Turkey J Plant Dis Prot* 127(4):561–569. <https://doi.org/10.1007/s41348-020-00347-5>
5. Gupta N, Islam S, Sharma SK, Baranwal VK (2017) Complete genome sequence of an isolate of *Leek yellow stripe virus* from garlic in India. *J Plant Pathol* 99(3):793–797. <https://doi.org/10.4454/jpp.v99i3.3945>
6. Valli A, Lopez-Moya J, Garcia JA (2007) Recombination and gene duplication in the evolutionary diversification of P1 proteins in the family *Potyviridae*. *J Gen Virol* 88:1016–1026. <https://doi.org/10.1099/vir.0.82402-0>
7. Adams MJ, Antoniw JF, Fauquet CM (2005) Molecular criteria for genus and species discrimination within the family *Potyviridae*. *Arch Virol* 150:459–479. <https://doi.org/10.1007/s00705-004-0440-6>
8. 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. *Arch Virol* 157(1):147–153. <https://doi.org/10.1007/s00705-011-1132-7>
9. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:1–5. <https://doi.org/10.1093/ve/vev003>
10. Wylie SJ, Li H, Saqib M, Jones MGK (2014) The global trade in fresh produce and the vagility of plant viruses: a case study in garlic. *PLoS One* 9(8):e105044. <https://doi.org/10.1371/journal.pone.0105044>
11. Wylie SJ, Luo H, Li H, Jones MGK (2012) Multiple polyadenylated RNA viruses detected in pooled cultivated and wild plant samples. *Arch Virol* 157:271–284. <https://doi.org/10.1007/s00705-011-1166-x>
12. Fajardo TVM, Nishijima M, Buso JA, Torres AC, Ávila AC, Resende RO (2001) Garlic viral complex: identification of potyviruses and carlavirus in central Brazil. *Fitopatologia Brasileira* 26:619–626
13. Tsuneyoshi T, Matsumi T, Natsuaki KT, Sumi S (1998) Nucleotide sequence analysis of virus isolates indicates the presence of three *Potyvirus* species in *Allium* plants. *Arch Virol* 143(1):97–113.

<https://doi.org/10.1007/s007050050271>

14. Chen J, Chen J, Adams MJ (2001) Molecular characterisation of a complex mixture of viruses in garlic with mosaic symptoms in China. *Arch Virol* 146(10):1841–1853.
<https://doi.org/10.1007/s007050170037>
15. Chen J, Chen JP, Adams MJ (2002) Characterisation of some carla- and potyviruses from bulb crops in China. *Arch Virol* 147(2):419–428. <https://doi.org/10.1007/s705-002-8330-y>
16. 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. *Arch Virol* 150(6):1135–1149.
<https://doi.org/10.1007/s00705-004-0482-9>
17. Navrátil M, Safárová D, Tkadlecová E, Klukáčková J (2006) Molecular characterization of Czech and Chinese *Leek yellow stripe virus* isolates from garlic. *Acta Virol* 50(3):207–210
18. Gupta N, Prabha K, Islam S, Baranwal VK (2013) First report of *Leek yellow stripe virus* in garlic from India. *J Plant Pathol* 95(s4):75. <https://doi.org/10.4454/JPP.V95I4.017>
19. 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. *Phytopathol Mediterr* 51(3):549–565
20. Tsuneyoshi T, Sumi S (1996) Differentiation among garlic viruses in mixed infections based on RT-PCR procedures and direct tissue blotting immunoassays. *Phytopathology* 86(3):253–259
21. Pérez-Moreno L, Córdova-Rosales, Barboza-Corona E, Ramírez-Malagón R, Ramírez-Lúa J, Ruiz-Castro S, Silva-Rosales L (2006) First report of *Leek yellow stripe virus* in garlic in the State of Guanajuato, Mexico. *Plant Dis* 90(11):1458. <https://doi.org/10.1094/PD-90-1458A>
22. Ha C, Revill P, Harding RM, Vu M, Dale JL (2008) Identification and sequence analysis of potyviruses infecting crops in Vietnam. *Arch Virol* 153(1):45–60. <https://doi.org/10.1007/s00705-007-1067-1>
23. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 35(6):1547–1549.
<https://doi.org/10.1093/molbev/msy096>
24. Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120.
<https://doi.org/10.1007/BF01731581>
25. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. *PLOS ONE* 9(9):e108277.
<https://doi.org/10.1371/journal.pone.0108277>
26. Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95–98
27. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol* 34:3299–3302. <https://doi.org/10.1093/molbev/msx248>

28. Hudson RR, Boos DD, Kaplan NL (1992) A statistical test for detecting geographic subdivision. *Mol Biol Evol* 9:138–151. <https://doi.org/10.1093/oxfordjournals.molbev.a040703>
29. Hudson RR (2000) A new statistic for detecting genetic differentiation. *Genetics* 155:2011–2014
30. Tsompana M, Abad J, Purugganan M, Moyer J (2005) The molecular population genetics of the *Tomato spotted wilt virus* (TSWV) genome. *Mol Ecol* 14:53–66. <https://doi.org/10.1111/j.1365-294X.2004.02392.x>
31. Gao F, Lin W, Shen J, Liao F (2016) Genetic diversity and molecular evolution of arabis mosaic virus based on the CP gene sequence. *Arch Virol* 161:1047–1051. <https://doi.org/10.1007/s00705-015-2729-z>
32. Dovas CI, Hatzibukas E, Salomon R, Barg E, Shibolet Y, Katis NI (2001) Comparison of methods for virus detection in *Allium* species. *J Phytopathol* 149:731–737. <https://doi.org/10.1046/j.1439-0434.2001.00705.x>
33. Etoh T, Simon PW (2002) Diversity, fertility and seed production of garlic. In: Rabinowitch HO, Currah L (eds) *Allium* crop science: recent advances. CABI Publishing, Oxford, pp 101–117

Table

Table 1 Similarity percentage of complete nt and aa sequences of different coding regions among LYSV phylogroups

Genomic region	Similarity (%)						Nucleotide identity threshold (%) for species discrimination [7]
	S-type		N-type		S-type vs N-type		
	nt	aa	nt	aa	nt	aa	
Polyprotein	76.8-99.4	40.2-98.5	90.8-92.3	72.7-80.1	75.0-79.3	38.1-51.1	76.0
P1	65.1-99.6	33.1-99.7	86.7-88.7	83.8-85.4	57.1-71.2	35.1-64.5	58.0
HC-Pro	74.7-100	82.0-100	90.7-91.6	93.9-94.5	74.9-80.8	79.2-91.0	76.0
P3	75.3-100	73.3-100	89.1-91.3	91.1-93.0	70.0-75.3	66.8-74.1	74.0
6K1	90.4-100	74.4-100	92.9-94.9	96.2-98.1	75.0-84.6	90.4-94.2	-
CI	79.2-100	93.2-100	91.9-93.3	98.7-99.2	79.3-83.2	92.9-95.1	78.3
6K2	71.7-100	69.2-100	90.6-91.8	90.6-94.3	70.4-82.4	67.3-86.8	-
VPg	75.5-100	81.2-100	89.8-91.3	95.8-96.4	74.1-79.9	75.9-90.6	76.0
Nla-Pro	77.8-100	91.7-100	92.8-94.8	98.3-99.2	76.7-82.5	86.4-92.1	76.5
NIb	77.3-100	89.6-100	91.9-93.0	98.2-99.0	78.1-82.6	89.5-93.2	75.0
CP ^α	76.9-100	77.5-100	88.8-100	90.9-100	76.6-86.6	75.4-91.7	76.0
3' UTR	90.4-100	78.0-100	92.4-93.5	78.9-82.3	87.4-93.5	60.6-88.6	76.0

Figures

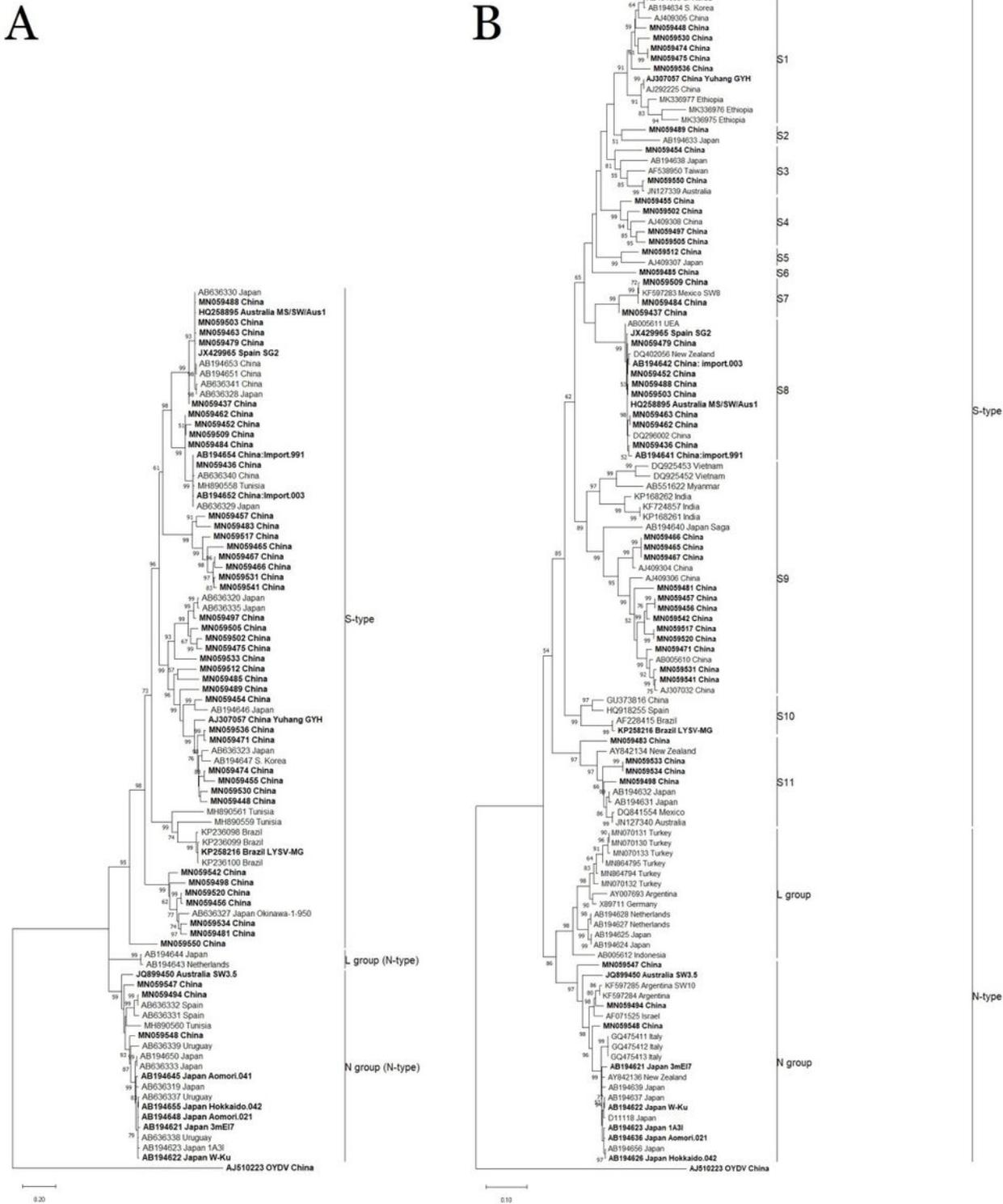


Figure 1

Phylogenetic tree constructed by Maximum Likelihood test of nucleotide sequences of LYSV P1 gene (A) and CP gene (B). Eighty-six and 115 isolates were applied in P1 and CP comparison, respectively. Isolates printed in bold were used in both P1 and CP comparisons. Bootstrap values on each branch were supported by 1000 replicates; only values greater than 50% were shown. An onion yellow dwarf virus (OYDV, Potyvirus) isolate was used as out-group

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PhylogeneticanddiversityanalysisofLYSVESM1.docx](#)