Archives of Virology
https://doi.org/10.1007/s00705-017-3673-x

ANNOTATED SEQUENCE RECORD

Genome characterization of an Argentinean isolate of alfalfa leaf curl virus

Nicolás Bejerman[1,2] · Verónica Trucco[1] · Soledad de Breuil[1,2] · Patricia Rodriguez Pardina[1] · Sergio Lenardon[1,3] · Fabián Giolitti[1]

Received: 18 September 2017 / Accepted: 24 November 2017

Abstract

We investigated the molecular characteristics of an Argentinean isolate of alfalfa leaf curl virus (ALCV-Arg), a virus of the genus *Capulavirus* in the family *Geminiviridae* that was isolated from alfalfa plants showing dwarfism. The genome was found to be 2,750 nucleotides in length. In pairwise comparisons, this ALCV isolate shared 83.2% to 92.6% sequence identity with European ALCV isolates. Sequence comparisons and phylogenetic analysis showed that this isolate combines features of strains A and B of ALCV. Recombination analysis showed that ALCV-Arg is a recombinant isolate that was generated by intraspecific recombination between ALCV strains A and B. The results of this study not only show that ALCV-Arg is unique because it combines features of strains A and B but also show that ALCV naturally infects this forage crop on the American continent.

In Argentina, alfalfa (*Medicago sativa* L.) is a primary forage crop and a major feed component in dairy and beef cattle production systems. In 2010, we observed alfalfa plants showing symptoms of shortened internodes (bushy appearance), leaf puckering, and vein enations of varying size on abaxial leaf surfaces [2]. Deep sequencing of small RNAs of alfalfa plants collected in the central region of Argentina showing dwarfism symptoms revealed the presence of four RNA viruses: alfalfa mosaic virus (AMV), alfalfa dwarf virus (ADV), alfalfa enamovirus 1 (AEV-1) and bean leaf roll virus (BLRV) [3, 4, 15, 16]. Furthermore, two assembled sequences (contigs) analyzed using BlastX were found to be related to sequences encoding the capsid and replication-associated proteins of capulavirus alfalfa leaf curl virus (ALCV) [12].

The recently created genus *Capulavirus* (family *Geminiviridae*) consists of circular ssDNA viruses with ~ 2.7-kb monopartite genomes that are extremely divergent and have a genome organization that is unique amongst the geminiviruses [5, 6, 17]. Four distinct capulaviruses have been reported so far: euphorbia caput medusae latent virus (EcmLV), which was discovered in South Africa infecting a wild spurge [5], French bean severe leaf curl virus, detected in French bean crops in India, Plantago lanceolata latent virus, identified in Finland infecting *Plantago lanceolata* [14], and alfalfa leaf curl virus (ALCV), which was first observed in France infecting alfalfa [12]. ALCV is unique because it is the only geminivirus discovered so far that is aphid-transmitted, having *Aphis craccivora* (Koch), an invasive aphid species with a global distribution, as its vector [12]. ALCV was found infecting alfalfa not only in France but also in Spain, and it is thought to be freely moving across the Mediterranean region [6]. The molecular characterization of several ALCV isolates from France and Spain showed evidence for the existence of two distinct groups, which were named "strain A" and "strain B" [6].

In this work, we describe the genome characterization of an Argentinean isolate of ALCV (ALCV-Arg). DNA

Handling Editor: Jesus Navas-Castillo.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00705-017-3673-x) contains supplementary material, which is available to authorized users.

✉ Nicolás Bejerman
nicobejerman@gmail.com

[1] IPAVE-CIAP-INTA, Av. 11 de Septiembre 4755, 5020 Córdoba, Argentina

[2] Concejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 (1425 FBQ) CABA, Buenos Aires, Argentina

[3] Facultad de Agronomía y Veterinaria (FAV), Universidad Nacional de Río Cuarto (UNRC), Enlaces Rutas 8 y 36, 5800 Río Cuarto, Córdoba, Argentina

Published online: 05 December 2017

was extracted using the cetyltrimethylammonium bromide (CTAB) method [7] from 100 mg of fresh leaves from an alfalfa plant collected in central Argentina showing dwarfism symptoms. Total DNA was subjected to treatment with Phi29 DNA polymerase (TempliPhi^{TM}, GE Healthcare, USA) to obtain a copy of the complete genome of ALCV-Arg by rolling-circle amplification (RCA). A linear DNA fragment of about 2.7 kb was generated with digestion with EcoRI and cloned, and three independent colonies were sequenced bi-directionally using universal M13 forward and reverse primers. Additionally, primers were designed for primer-walking the full genome. The sequence fragments obtained were edited and assembled to obtain the complete genome sequence of the Argentinean isolate of ALCV (ALCV-Arg) using Geneious R9.1 (Biomatters, New Zealand), which was deposited in the GenBank database under accession number KX574859.

The ALCV-Arg genome is composed of 2,750 nucleotides (nt), which is within the range of genome lengths described for strain A members, but smaller than the genomes of strain B members (Supplementary Table 1). ALCV-Arg circular DNA contains three overlapping complementary-sense open reading frames (ORFs) (C1, C2 and C3), two intergenic regions and four virion-sense ORFs (V1, V2, V3 and V4). This genomic organization is similar to that reported for seven isolates belonging to strain A (Supplementary Table 1) [12]. However, most of the ALCV isolates reported so far (20/27) do not encode the V2 protein (Supplementary Table 1). Furthermore, the ALCV-Arg V2 protein is 29 amino acids (aa) larger than the V2 proteins encoded by seven ALCV isolates of strain A (Supplementary Table 1). This protein is suspected to be involved in movement [5], although its function has never been tested. Interestingly, the ALCV-Arg V4 protein, which is also suspected of being involved in movement [5], has the same size as the V4 proteins encoded by seven ALCV isolates of strain A but is 11 aa larger than the V4 proteins encoded by the remaining 20 ALCV isolates (Supplementary Table 1). The significance of this feature is not known and should be analyzed in further studies. In addition, while the RepA and C3 proteins encoded by ALCV-Arg are similar in size to those of ALCV isolates belonging to strain B (Supplementary Table 1), the length of its long intergenic region is similar to that of seven isolates of strain A. Interestingly, the Rep protein encoded by ALCV-Arg is similar in size to that of isolate 44-1E, which belongs to strain A (Supplementary Table 1). This size is similar to that described for the rep protein encoded by EcmLV [5], whereas the short intergenic region of ALCV-Arg is the shortest among all ALCV isolates described so far (Supplementary Table 1). The genomic analysis therefore reveals that ALCV-Arg combines features of both ALCV A and B strains.

Pairwise comparison using Sequence Demarcation Tool (SDT v1.2) [11] showed that ALCV-Arg was 83.2% to 92.6% identical to the other described ALCV isolates at the whole-genome-sequence level. The percentages of nucleotide sequence identity ranged from 83.2% to 84.1% with ALCV strain A members and from 90.9% to 92.6% with ALCV strain B members, as shown in the color-coded matrix of pairwise identity scores (Fig. 1). The degree of identity between ALCV-Arg and the ALCV isolates is above 78%, which is the species demarcation threshold recommended recently for capulaviruses [17]. Thus, this result indicates that the ssDNA virus identified in alfalfa in Argentina belongs to the species *Alfalfa leaf curl virus*. Despite its sequence similarity to ALCV isolates belonging to strain B, the genome organization of ALCV-Arg is similar to that of most of the isolates (7/11) belonging to strain A.

The complete genome sequences of ALCV-Arg was aligned with the 27 known complete genome sequences of ALCV (Supplementary Table 2) using MUSCLE [8] as implemented in MEGA 7 [9] to infer the evolutionary relationships of ALCV-Arg to European ALCV isolates. A maximum-likelihood (ML) phylogenetic tree was constructed using MEGA 7 [9], with the TN93+G nt substitution model chosen as the best-fit using jModelTest [13] and 1000 bootstrap replicates. The ML tree showed that ALCV-Arg belongs to a separate branch that clusters with ALCV isolates belonging to strain B (Fig. 2A). This supports the results obtained by pairwise comparison.

The predicted Rep and capsid protein (CP) aa sequences of ALCV-Arg were aligned with the corresponding sequences of the ALCV isolates listed in Supplementary Table 2, using MUSCLE [8] as implemented in MEGA 7 [9]. ML phylogenetic trees were constructed using MEGA 7 [9] with the WAG+G+I amino acid substitution model chosen as the best fit using ProtTest [1] and 1000 bootstrap replicates. Although the size of the rep protein encoded by ALCV-Arg is most similar to that encoded by the ALCV isolate 44-1E (strain A), in the Rep-based ML tree, ALCV-Arg belongs to a separate branch that clusters with ALCV isolates belonging to strain B (Fig. 2B), with topology similar to that observed in the tree based on the complete genome. However, in the CP-based ML tree, ALCV-Arg belongs to a separate branch that clusters with ALCV isolates belonging to both strains A and B (Fig. 2C). Therefore, this different clustering pattern of ALCV-Arg indicates incongruence in the comparative phylogenies based on the full genome and Rep sequences versus that based on the CP sequences, suggesting that this isolate might have arisen due to recombination. Thus, to test this hypothesis, the aligned complete genome sequences were scanned using the RDP4 program [10] with default parameters, with the exception that the default option of `linear´ was replaced with 'circular' genome architecture. Two putative recombinant events were

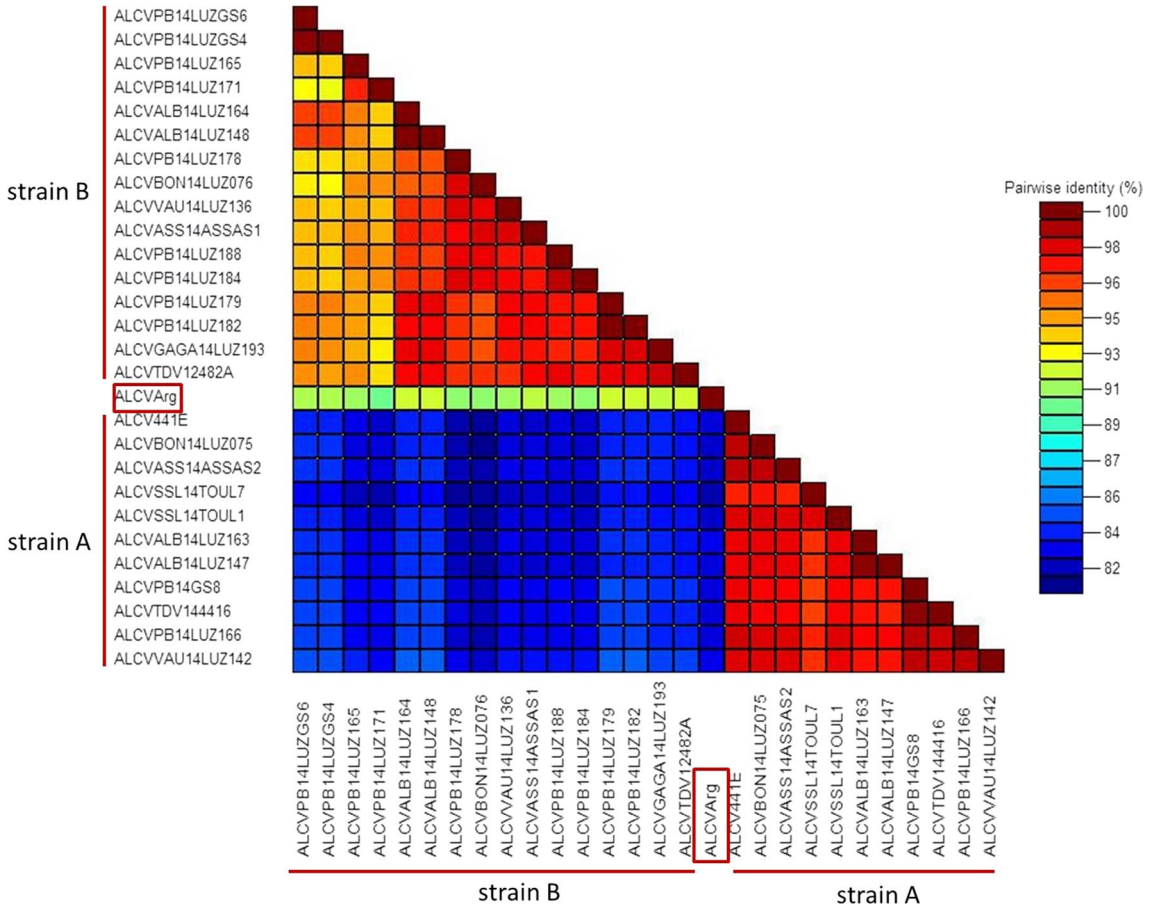

Fig. 1 Genome-wide percentage pairwise identities of ALCV calculated using SDT v1.2 [11]

considered authentic based on a threshold of its prediction by at least four of the seven RDP-implemented programs and strong statistical support ($P > 0.05$) for the detected events. Based on these criteria, one event showed ALCV-Arg to be a putative recombinant isolate, with the putative major and minor parents being ALCV-PB14-Luz178 (strain B) and ALCV-ALB14-Luz147 (strain A), respectively (Supplementary Table 3). This recombination event was further supported by analyzing nucleotide identity, using the segment corresponding to the recombination event and the region to each side of the predicted recombination breakpoints (Supplementary Table 4). The other identified event spanned the region from nt 54-839. In this detected event, ALCV-Arg may be the putative major parent and ALCV-TDV14-44 (strain A) the minor parent, and the recombinant sequence may be VAU14-Luz136 (strain B). However, based on the RDP results, it is unclear which of the sequences implicated in this event is the recombinant and which are the "parents". Comparison of more sequences will be needed to determine which one of the sequences is the recombinant one. It has been shown that European ALCV genomes also display evidence of recombination [6].

The recombination results show that ALCV-Arg might have originated from an intraspecific recombination event between ALCV isolates belonging to strains A and B, which supports the observation that ALCV-Arg, despite having a genomic organization similar to that of members of strain A, has the highest pairwise identity to members of strain B. More sequences will be needed to support this. Moreover, phylogenetic analysis of the complete genome sequence showed that ALCV-Arg branches basal to the strain B group, which reflects its genomic organization, and by extension, its status as a recombinant.

In conclusion, this is the first report of the genome characterization of an ALCV isolate, and of a capulavirus from the American continent and the Western Hemisphere. This finding extends the known geographical range of ALCV and capulaviruses to a new continent. Additional studies should be conducted to determine how ALCV spread between Europe and South America and to examine ALCV diversity in Argentina. Moreover, further work, including determination of the role of ALCV in alfalfa dwarf disease and how it contributes to symptom expression, is necessary to understand more about this widespread disease of alfalfa.

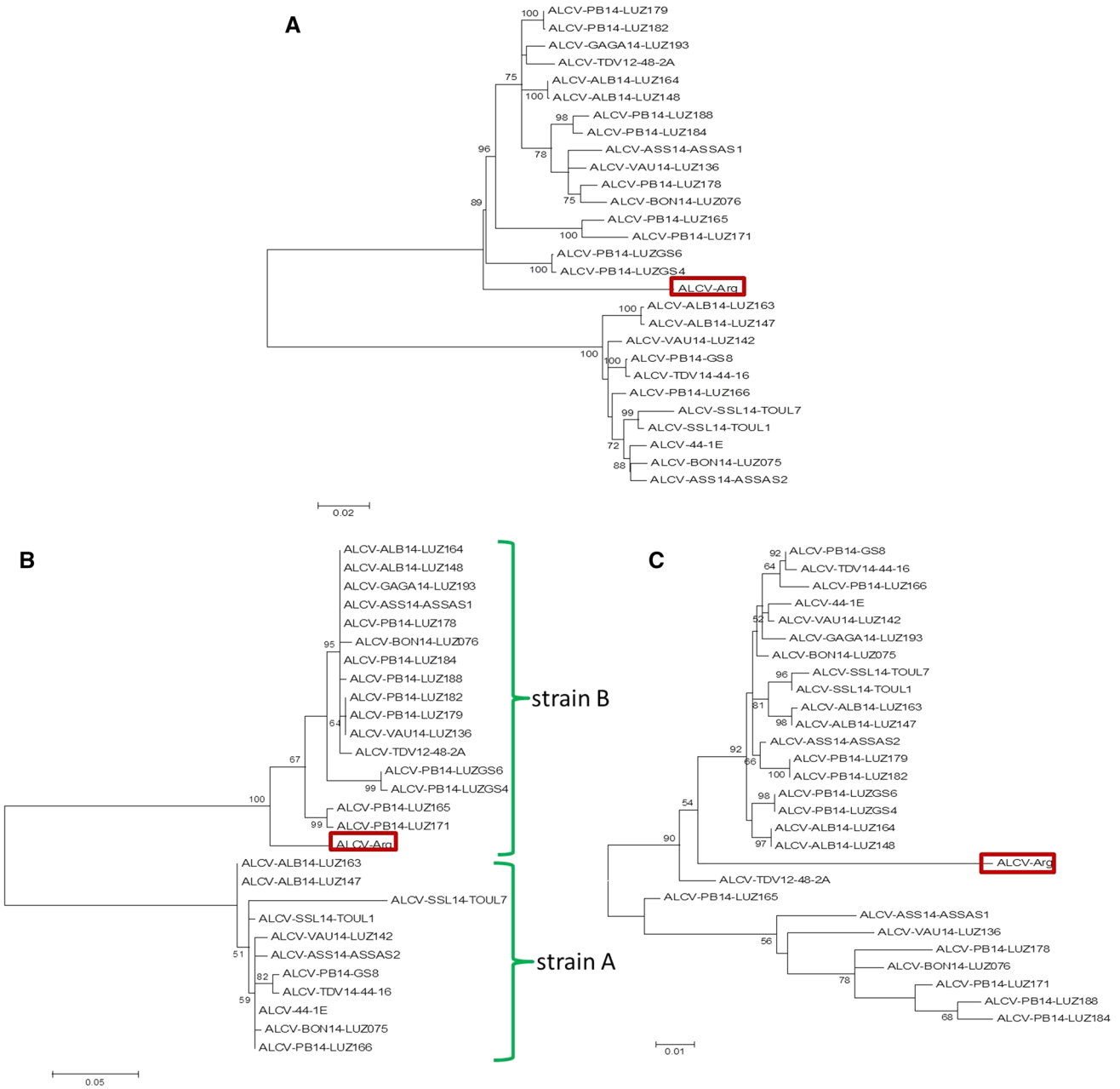

Fig. 2 **A.** Maximum-likelihood phylogenetic tree of ALCV based on full genome sequences. **B.** Maximum-likelihood phylogenetic tree based on the Rep amino acid sequences of ALCV isolates. **C.** Maximum-likelihood phylogenetic tree based on the CP amino acid sequences of ALCV isolates. All trees were constructed using MEGA 7 software. The TN93+G model was used to construct the phylogenetic tree based on the full genomes, whereas the model WAG+G+I was used to construct the phylogenetic trees based on CP and Rep amino acid sequences. Bootstrap values out of 1000 replicates are given at the nodes, but only values above 50% are shown. The bar below each tree represents substitutions per site. The accession number of each ALCV isolate used to construct the phylogenetic trees is listed in Supplementary Table 2

Acknowledgements This work was partially supported by the PNPV 1135022 project of INTA.

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Funding This work was partially supported by the PNPV 1135022 project of INTA.

References

1. Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105
2. Bejerman N, Nome C, Giolitti F, Kitajima E, de Breuil S, Pérez Fernández J, Basigalup D, Cornacchione M, Lenardon S (2011) First report of a Rhabdovirus infecting alfalfa in Argentina. Plant Dis 95:771
3. Bejerman N, Giolitti F, Trucco V, de Breuil S, Nome C, Lenardon S, Dietzgen R (2015) Complete genome sequence and integrated protein localization and interaction map for alfalfa dwarf virus, which combines properties of both cytoplasmic and nuclear plant rhabdoviruses. Virology 483:275–283
4. Bejerman N, Giolitti F, Trucco V, de Breuil S, Dietzgen R, Lenardon S (2016) Complete genome sequence of a new enamovirus from Argentina infecting alfalfa plants showing dwarfism symptoms. Arch Virol 161:2029–2032
5. Bernardo P, Golden M, Akram M, Naimuddin Nadarajan N et al (2013) Identification and characterisation of a highly divergent geminivirus: evolutionary and taxonomic implications. Virus Res 177:35–45
6. Bernardo P, Muhire B, Francois S, Deshoux M, Hartnady P et al (2016) Molecular characterization and prevalence of two capulaviruses: Alfalfa leaf curl virus from France and Euphorbia caput-medusae latent virus from South Africa. Virol 493:142–153
7. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15
8. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797
9. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874
10. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003
11. Muhire BM, Varsani A, Martin DP (2014) SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9:e108277
12. Roumagnac P, Granier M, Bernardo P, Deshoux M, Ferdinand R et al (2015) Alfalfa leaf curl virus: an aphid-transmitted geminivirus. J Virol 89:9683–9688
13. Santorum JM, Darriba D, Taboada GL, Posada D (2014) jmodeltest.org: selection of nucleotide substitution models on the cloud. Bioinformatics 30:1310–1311
14. Susi H, Laine A-L, Filloux D, Kraberger S, Farkas K et al (2017) Genome sequences of a capulavirus infecting *Plantago lanceolata* in the Åland archipelago of Finland. Arch Virol 162:2041–2045
15. Trucco V, de Breuil S, Bejerman N, Lenardon S, Giolitti F (2014) Complete nucleotide sequence of *Alfalfa mosaic virus* isolated from alfalfa (*Medicago sativa* L.) in Argentina. Virus Genes 48:562–565
16. Trucco V, de Breuil S, Bejerman N, Lenardon S, Giolitti F (2016) Bean leafroll virus (BLRV) in Argentina: molecular characterization and detection in alfalfa fields. Eur J Plant Pathol 146:207–212
17. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E et al (2017) Capulavirus and Grablovirus: two new genera in the family Geminiviridae. Arch Virol 162:1819–1831