Variability of P26 and P10 genes in Colombian isolates of *Potato yellow vein virus* (PYVV)

Variabilidad de los genes P26 y P10 en aislamientos colombianos del *Potato yellow vein virus* (PYVV)

Jhon Calderón[1], Teresa Mosquera Vásquez[1], and Ángela María Vargas[2]*

ABSTRACT

Potato yellow vein virus (PYVV) is the causal agent of the potato yellow vein disease and can reduce potato production up to 50%. This virus also infects tomatoes and can remain asymptomatic in plants. PYVV transmission is mediated by vegetative seed, the vector *Trialeurodes vaporariorum,* and grafts. Its genome has the P26 and P10 genes that are orthologues in the *Crinivirus* genus, which have been characterized as pathogenic factors and have not been studied in PYVV. We analyzed the variability of P26 and P10 from 45 and 48 sequences, which were obtained by RT-PCR amplification of the total RNA of symptomatic potato leaves from the provinces of Nariño, Cundinamarca, and Boyaca (Colombia). We included sequences of each gene of the PYVV genome of potato and tomato isolates from GenBank. The variability in these genes is influenced by the flow and uncontrolled use of vegetative seed between different provinces, that favor the dispersion of viral variants. In addition, the variability analysis based on maximum likelihood trees, haplotypes, and diversity indices showed that P26 is more variable than P10 and both are more variable in Andigena than in Phureja potatoes. The Tajima and Fu and Li tests revealed that these genes are subject to negative selection.

Key words: tripartite genome, *Crinivirus, Solanum tuberosum,* negative selection.

RESUMEN

El virus *Potato yellow vein virus* (PYVV), en español virus del amarillamiento de las venas de la papa, es el agente causal de la enfermedad conocida como amarillamiento de venas de la papa y puede reducir la producción hasta un 50%. Este virus también infecta tomate y puede permanecer en plantas asintomáticas. Su transmisión está mediada por semilla vegetativa, el vector *Trialeurodes vaporariorum* e injertos. Su genoma codifica los genes P26 y P10 que son ortólogos en el género *Crinivirus,* en el cual se han caracterizado como factores de patogenicidad y no han sido estudiados en PYVV. Se analizó la variabilidad de P26 y P10 a partir de 45 y 48 secuencias respectivamente, obtenidas de la amplificación por RT-PCR del RNA total de hojas sintomáticas de papa de los departamentos de Nariño, Cundinamarca y Boyacá (Colombia), incluyendo tres secuencias de cada gen de los genomas de PYVV de aislamientos de papa y tomate reportados en GenBank. La variabilidad en estos genes está influenciada por el flujo y uso no controlado de semilla vegetativa entre diferentes departamentos, lo que favorece la dispersión de variantes virales. Además, los análisis de variabilidad basados en árboles de máxima verosimilitud, haplotipos e índices de diversidad mostraron que P26 es más variable que P10 y que ambos son más variables en papa Andígena que en Phureja. Las pruebas de Tajima y Fu and Li revelaron que estos genes están sometidos a la selección negativa.

Palabras clave: genoma tripartito, *Crinivirus, Solanum tuberosum,* selección negativa.

Introduction

Potato (*Solanum tuberosum*) is cultivated in Colombia, especially in the provinces of Antioquia, Boyaca, Cundinamarca and Nariño, where Cundinamarca and Boyaca are the most important areas for potato production (Ñústez, 2011). According to Fedepapa (2014), 250 potato cultivars are produced in Colombia; but the most important are: Criolla Colombia, Criolla Dorada, Criolla Ocarina, Pastusa Suprema, Superior, Diacol Capiro, Ica Puracé, Tuquerreña, and Roja Nariño.

The *Crinivirus* genus comprises an important group of emergent plant viruses widely distributed in the world belonging to the Closteroviridae family (Martelli *et al*., 2013). Within their genomes two proteins are coded that oscillate between 26 and 28 KDa (group P26) and 8 and 10 KDa (group P10). These proteins are unique in the genus and are orthologous in the 15 species sequenced and reported in the GenBank (Kiss *et al*., 2013; Genbank, 2019). These proteins have been characterized in some criniviruses, in which their participation in infectious processes, cell movement

Received for publication: 3 June, 2018. Accepted for publication: 30 August, 2019 Doi: 10.15446/agron.colomb.v37n2.72638

[1] Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, Bogota (Colombia).
[2] Corporación Colombiana de Investigación Agropecuaria - Agrosavia, C.I. Nataima (Colombia).
* Corresponding author: avargas@agrosavia.co

(Grimsley *et al*., 1986; Medina *et al*., 2003; Medina *et al*., 2005; Wang *et al*., 2009; Kiss *et al*., 2013), and physical self-interaction (Stewart *et a*l., 2009) have been demonstrated.

Potato yellow vein virus (PYVV) is a re-emergent *Crinivirus* that causes potato yellow vein disease (PYVD), which is characterized by chlorosis in veins and can remain in asymptomatic plants. It also infects *Solanum lycopersicum*, *Rumex* sp., *Catharanthus* sp. and *Polygonium* sp. plants (Salazar *et al*., 2000; Muñóz *et al*., 2017). The virus is transmitted by infected tubers, grafts, and by the whitefly *Trialeurodes vaporariorum* in a semi-persistent way (Salazar *et al*., 2000; Gamarra *et al*., 2006; Wale *et al*., 2008; Barragán and Guzmán-Barney, 2014; Hernández and Guzmán-Barney, 2014).

PYVV was first reported in Antioquia (Colombia) in 1943 in plants of the *S. tuberosum* group Phureja (Diploid) (Salazar *et al*., 2000), and it was later reported in *S. tuberosum* group Andigena (Tetraploid). In 1996 the PYVV virus dispersed to Ecuador, Peru and in 1998 to Venezuela (Salazar *et al*., 2000; EEPO, 2019).

PYVV is a virus limited to the phloem and accompanying cells. It has a tripartite genome composed by three molecules of single-stranded RNA, positive polarity and two defective RNAs (Eliasco *et al*., 2006; Muñoz *et al*., 2017). RNA 1 (8,035 kb) has three open reading frames (ORFs), RNA 2 (5,339 kb) and has five ORFs, and RNA 3 (3,892 Kb) has three ORFs. The P10 and P26 proteins of PYVV are present in both RNA 2 and RNA 3.

The *Closteroviridae* family is considered to be invariable, but there are differences among the genera. For example, *Ampelovirus* and *Closterovirus* genera are the most variable, and *Crinivirus* the least variable (Rubio *et al*., 2013; Erkiş-Güngör and Bayram, 2019). Low variability and genetic diversity have been reported for PYVV through single strand conformation polymorphisms (SSCP), restriction fragment length polymorphisms (RFLP) and sequence analysis of major capsid protein gene (CP), minor capsid gene (CPm) and heat shock protein homologue gene (HSP70h), although CPm is more variable and has a tendency for recombination (Offei *et al*., 2004; Guzmán *et al*., 2006; Rubio *et al*., 2013; Chaves-Bedoya *et al*., 2014; Cubillos and Guzmán-Barney, 2015).

Due to the importance of P10 and P26 gene orthologues group in pathogenicity, its viral cycle in *Crinivirus*, and the fact that these genes have not been studied in PYVV, it is necessary to know about the variability, evolution and presence of molecular viral variants in potato plants generated by selection pressure, mutations, genetic drift, migration and also inter and intraspecific recombination events. These factors may impact the PYVV infectious capacity within its host and the ability to colonize other plant species, increasing the hosts' range as in tomato plants (Ruiz *et al*., 2018). Our objective in this study was to analyze the P26 and P10 genes of PYVV isolates from two contrasting geographic regions of Colombia to determine their genetic variability, evolution, and possible relationships with the host.

Materials and methods

Plant material

Sampling was performed in potato plots with a high incidence of plants affected by PYVD symptoms in Nariño, Cundinamarca, and Boyaca (Colombia). From each crop plot leaf samples from the middle third of different plants of the potato *S. tuberosum* Andigena group (StA) and Phureja group (StP) were collected. The samples were covered with absorbent paper towels, stored in perforated plastic bags, and transported to the Biotechnology Laboratory of the Facultad de Ciencias Agrarias of the Universidad Nacional de Colombia, Bogota campus for the analysis.

Primer design

To obtain a set of primers for specific amplification of P10 and P26 genes of PYVV, we used the Primer 3 program v.0.4.0 (Koressaar and Remm, 2007; Untergasser *et al*., 2012) and Primer BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). As templates, we used accessions AJ557129.1 (for P10 gene) and AJ508757.2 (for P26 gene) of PYVV Peruvian genome sequences from GenBank. The two primer pairs with the best characteristics in terms of length (20 to 25 bp), > 40% guanine-cytosine percentage, minimum dimerization and temperature delta between primers, flanking location for covering 100% of each gene, and melting temperature between 50 and 60°C were selected. The AmplifX program version 1.5.4 (http://ifrjr.nord.univ-mrs.fr/AmplifX) was used to simulate the migration of fragments of the P10 and P26 genes with the selected primers in a virtual agarose gel.

Total RNA extraction and verification of PYVV presence

Total RNA was extracted with the RNeasy Plant Minikit (Qiagen™) following the manufacturer's instructions, except for the use of the QIAshredder® column. The same column was used to purify and concentrate two replicates of each sample, and these were unified. RNA was diluted

in 30 µL of RNAase free water (Qiagen™) and maintained at -20°C until use. To verify the presence of PYVV in the samples, the CP gene was amplified as a control, according to the protocol by Hernández and Guzman-Barney (2014).

Primer verification, RT-PCR and sequencing

All the reagents used for the reverse transcription (RT) and amplification (PCR) process were from the Invitrogen™ brand (Thermo Fisher Scientific, MA, USA), and RT-PCR was performed in a thermal cycler (C1000, BioRad®, Berkeley, CA).

Two samples from Nariño and Cundinamarca were selected for primer verification. SuperScript® III kit (First-Strand Synthesis System for RT-PCR®) was used for P10 and P26 RT in positive samples for amplification of PYVV-CP, following the manufacturer's instructions. In separated reactions we mixed 3 µL of total RNA, 0.5 µM of P10 and P26 primer reverse (designed in this investigation, Tab. 2), 1 mM of dNTPs, 1X of buffer, 10 mM of ditioeritrol (DTT), 1.6 u μL^{-1} of RNaseOut®, 5 mM of $MgCl_2$, 8 u μL^{-1} of reverse transcriptase (Superscript®III), and RNAse free water to complete the final volume of 10 µL. After RT was performed, PCR for P10 and P26 genes were carried out in separated reactions with PCR® kit, 1X of buffer, 2.5 mM of $MgCl_2$, 0.4 µM of forward and reverse primers (Tab. 2), 0.4 mM of dNTPs, 1 u μL^{-1} of high fidelity Taq Polymerase Platinum®, 0.6 µL of cDNA, and diethyl pyrocarbonate (DEPC) treated water to complete a final volume of 15 µL. The amplification cycle was as follows: 1 min at 94°C, followed by 34 cycles of 1 min at 94°C, 30 s at 58°C for annealing and 1 min at 72°C for extension, and 10 min at 72°C for final extension. For the amplicon visualization, 2 µL of amplified product was loaded in 1% agarose gel in TAE buffer. The gel was stained with 0.02 mg mL^{-1} of ethidium bromide. The kit Pure Link PCR Purification was used to purify P10 and P26 gene PCR products. Purified products were diluted in 30 µL DNAse free water and sent to Macrogen® for forward and reverse sequencing.

Variability and diversity of P26 and P10 genes of PYVV

Forward and reverse sequences obtained from each amplicon were assembled and a consensus sequence was generated using CAP3 of the PRABI-Doua- program (Huang and Madan, 1999). Complete P26 and P10 gene sequences were submitted to the EMBL/GenBank to verify the identity using the P26 and P10 genes of PYVV. After verifying the genes' identity, the best primers were selected for RT-PCR performance and sequencing of all the samples in the same way as described above.

For the variability analysis, the contigs of P26 and P10 genes of PYVV were grouped as follows: (1) potato genotype: Andigena (Tetraploid) and Phureja (Diploid); (2) geographical origin: Nariño (N), Cundinamarca (C) and Boyaca (B): and (3) total contigs. The Sequence Demarcation Tool was used for evaluating nucleotide identity through the pairwise matrix (Muhire *et al.,* 2014). The Mega 7 Program (Kumar *et al.,* 2016) was used for multiple alignments by codons with Clustal algorithm and to estimate the best-fit nucleotide substitution model by the Akaike Information Criterion (AIC). The best model was used for the construction of the maximum likelihood tree for each gene with the bootstrap test with 1000 replicas for generating the consensus tree (Nei and Kumar, 2000). The Mega 7 program was used to estimate the average of non-synonymous (dN) and synonymous (dS) mutations and their ratio (dN/dS). The program DNaSP 5.0 (Librado *et al.,* 2009) was used for genetic variability in which we estimated total variable sites (s), total mutations (η), nucleotide diversity (π), total haplotypes (h), and haplotype diversity (dH). The neutrality test of Tajima's D (Tajima, 1989) based on the difference between segregating sites and the average of nucleotide differences, and the Fu and Li's F* test (Fu and Li, 1993) based on the differences between singletons and the average of the number of pairwise nucleotide differences were also used. The program Network 5 (Bandelt *et al*., 1999) was used for graphic representation of haplotypes for the visualization of phylogenetic networks through the Median-Joining (MJ) algorithm and the statistical parsimony method as described by Templeton *et al*. (1992).

Results

Plant material

A total of 46 samples from symptomatic *S. tuberosum* Phureja and Andigena groups (Fig. 1), from six varieties (Tetraploids: Pastusa Suprema, Diacol Capiro, and Superior. Diploids: Criolla Colombia, Criolla Guaneña and Mambera) were sampled in eight municipalities from Nariño, six from Cundinamarca and one from Boyaca (Tab. 1).

Primer selection for amplification of P10 and P26 genes of PYVV

P26 and P10 genes were amplified during the preliminary test (Fig. 2) from samples CT0426 and NT1926 (Tab. 1) (positive to the PYVV-CP gene) through four pairs of primers (Tab. 2). We decided to select P10F and P10R primers for the P10 gene amplification and P26F gene and P26R for the P26 gene.

FIGURE 1. Potato plants affected by Potato Yellow Vein Disease (PYVD). The symptoms in the potato crop are observed as yellow plants distributed in foci. A) potato crop with high incidence of plants affected by PYVD, B) potato plant with initial PYVD symptoms, C) potato plant with advanced PYVD symptoms, D) detail of a potato leaf affected by PYVD.

TABLE 1. Isolate codification of P26 and P10 genes of *Potato yellow vein virus* (PYVV) obtained from Cundinamarca, Nariño and Boyaca.

P26 gene code	P10 gene code	Province	Municipality	Host	Variety
CT0426	CT0410	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1126	CT1110	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1226	CT1210	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1326	CT1310	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1426	CT1410	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1526	CT1510	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1626	-----	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1726	CT1710	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
CT1826	CT1810	Cundinamarca	Subachoque	Andigena	Pastusa Suprema
NT1926	NT1910	Nariño	Carlosama	Andigena	Diacol Capiro
NT2026	-----	Nariño	Carlosama	Andigena	Diacol Capiro
NT2126	NT2110	Nariño	Carlosama	Andigena	Diacol Capiro
NT2226	NT2210	Nariño	Ipiales	Andigena	Diacol Capiro
NT2326	NT2310	Nariño	Cumbal	Andigena	Diacol Capiro
NT2426	NT2410	Nariño	Cumbal	Andigena	Diacol Capiro

Continue

P26 gene code	P10 gene code	Province	Municipality	Host	Variety
NT2526	NT2510	Nariño	Pupiales	Andigena	Diacol Capiro
NT2626	NT2610	Nariño	Gualmatan	Andigena	Diacol Capiro
NT2726	NT2710	Nariño	Pasto	Andigena	Diacol Capiro
NT2826	NT2810	Nariño	Pasto	Andigena	Diacol Capiro
ND2926	ND2910	Nariño	Ipiales	Phureja	Guaneña
ND3026	ND3010	Nariño	Ipiales	Phureja	Guaneña
-----	ND3110	Nariño	Carlosama	Phureja	C. Colombia*
-----	ND3210	Nariño	Carlosama	Phureja	C. Colombia*
ND3326	ND3310	Nariño	Tuquerres	Phureja	Guaneña
ND3426	ND3410	Nariño	Tuquerres	Phureja	Guaneña
ND3526	ND3510	Nariño	Guaitarrilla	Phureja	Mambera
ND3626	ND3610	Nariño	Guaitarrilla	Phureja	Mambera
ND3726	ND3710	Nariño	Tuquerres	Phureja	C. Colombia*
ND3826	ND3810	Nariño	Tuquerres	Phureja	C. Colombia*
CT3926	-----	Cundinamarca	Villa Pinzon	Andigena	Diacol Capiro
CT4026	CT4010	Cundinamarca	Villa Pinzon	Andigena	Diacol Capiro
CT4126	CT4110	Cundinamarca	Guasca	Andigena	Diacol Capiro
CT4226	CT4210	Cundinamarca	Villa Pinzon	Andigena	Diacol Capiro
CT4326	CT4310	Cundinamarca	Suesca	Andigena	Diacol Capiro
CT4426	CT4410	Cundinamarca	Suesca	Andigena	Diacol Capiro
-----	CT4510	Cundinamarca	Usme	Andigena	Diacol Capiro
CT4726	CT4710	Cundinamarca	Guasca	Andigena	Diacol Capiro
-----	CT4610	Cundinamarca	Suesca	Andigena	Diacol Capiro
CT4926	CT4910	Cundinamarca	Suesca	Andigena	Diacol Capiro
-----	CT5010	Cundinamarca	Usme	Andigena	Diacol Capiro
CT5126	CT5110	Cundinamarca	Usme	Andigena	Diacol Capiro
CT5226	CT5210	Cundinamarca	Usme	Andigena	Diacol Capiro
CT5326	CT5310	Cundinamarca	El Rosal	Andigena	Pastusa Suprema
CT6226	-----	Cundinamarca	Usme	Andigena	Diacol Capiro
-----	BT6710	Boyaca	Ventaquemada	Andigena	Pastusa Suprema
-----	BT7010	Boyaca	Ventaquemada	Andigena	Pastusa Suprema

*Criolla Colombia.

TABLE 2. Sequences of primers obtained for P10 and P26 genes amplification of *Potato yellow vein virus* (PYVV).

Primer	Sequence	%GC	3'	ΔT°	MT	Amp	AT
P10F	GAAAGACATGACAGATGAGGAAGTG	44	0	0	52°C	628	55°C
P10R	CTGTCTGCTCTAACCTGAATCTTTG	44	0				
P10F2	AAGGTTACACACTGAGAAGAGAA	39	0	0	50°C	367	55°C
P10R2	GGATCCATTGTTCTAGTACCTCA	43	1				
P26F	GGCATTGAACAGTCCGAACAC	52	0	0	53°C	900	55°C
P26R	ATCACTCGTACTAGACCTCGGG	54	0				
P26F2	CGAAGACACATGCCAACAAG	50	0	1	51°C	998	55°C
P26R2	TCGTACTAGACCTCGGGTAAATAA	42	0				

%GC: Guanine-Cytosine percentage; 3´:3´ complementarity; *ΔT°*: temperature delta between primers; *MT*: melting temperature; *Amp*: expected amplicon size in base pairs (bp); *AT*: annealing Temperature. Grey highlight: selected primer sequences for amplification of all samples.

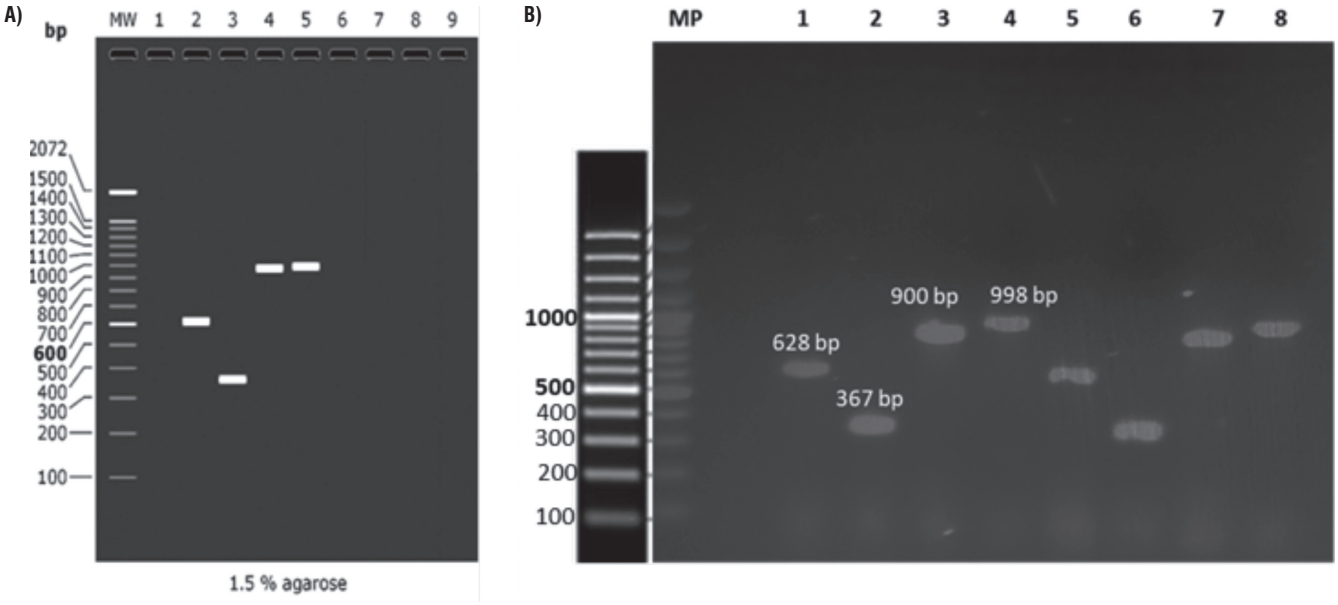

FIGURE 2. Representative amplification gel of P10 and P26 genes of *Potato yellow vein virus* (PYVV) by RT-PCR and amplicon migration comparison A) in silico and B) experimental amplification. A) In silico migration obtained by AmpliFx: lane 1: molecular weight marker, 2 and 3: P10 gene primers set, 4 and 5: P26 gene primers set. B) Representative experimental gel of amplification for P10 gene (Lanes 1, 2 and 5, 6) and P26 gene (lanes 3, 4 and 7, 8), CT0426 sample (lanes 1 to 4) and NT1926 sample (lanes 5 to 8).

Analysis of sequences

Forty-two contig sequences of P26 gene and 45 of the P10 gene of PYVV from potato leaflets were obtained (Tab. 1). Three reference sequence accessions for P26 and P10 genes of PYVV genomes were included. The first one was obtained from the *S. tuberosum* Andigena group (Cajamarca, Peru) (Livieratos *et al*., 2002), the second one from the *S. tuberosum* Phureja group (Antioquia) (Muñoz *et al*., 2016), and the third one from tomatoes (Antioquia) (Gutiérrez *et al*., 2017). In addition, the sequences of the P26 and P10 genes of the *Crinivirus* LIYV were used as an out-group in the phylogenetic analysis (Tab. 3).

Haplotype diversity of P26 and P10 genes of PYVV and networks

Eleven haplotypes (haplotypes 1 to 11) for the P10 gene of PYVV were obtained. These haplotypes had frequencies between 6.6% and 62.3% (Tab. 4). Haplotype 4 was the most frequently formed with 28 isolates from Cundinamarca and Nariño. This haplotype was the most ancestral, according to the haplotype network (Fig. 3). For the P26 gene of PYVV, 23 haplotypes (Haplotypes 1 to 23) were obtained with frequencies between 2.4% and 16.7% (Tab. 5) in which haplotypes 4, 7 and 9 were the most frequent and haplotype 4 was formed by four isolates from

TABLE 3. P10 and P26 gene reference accessions of *Potato yellow vein virus* (PYVV) and lettuce infectious yellow virus orthologues.

P26 gene Accession	P10 gene Accession	Province/country	Municipality	Host	Reference
AJ508757.2	AJ557129.1	Cajamarca/Peru	Chota	Potato Andigena	Livieratos *et al.,* 2002
KX573903.1	KX573902.1	Antioquia/Colombia	Marinilla	*S. lycopersicum*	Gutiérrez *et al.,* 2017
KR998195.1	KR998194.1	Antioquia/ Colombia	La Union	Group Phureja	Muñoz *et al.,* 2016
U15441.1	U15441.1	Maryland/USA	-	*Nicotiana clevelandii*	Klassen *et al.,*1994

TABLE 4. Haplotypes for the P10 gene of *Potato yellow vein virus* (PYVV).

P10 gene haplotypes	Quantity of sequences	Codes of P10 and P26 gene sequences	Frequency (%)
Hap 1	6	KR998194.1, AJ557129.1, CT1310, CT1810, CT4110, CT4210	13.3
Hap 2	1	KX573902.1	2.2
Hap 3	3	CT0410, NT2210, NT2410	6.6
Hap 4	28	CT1210, CT1110, CT1410, CT1510, CT1710, NT2110, NT2510, NT2610, NT2710, NT2810, ND3010, ND3110, ND3310, ND3410, ND3510, ND3610, ND3710 ND3810, CT4010, CT4310, CT4410, CT4510, CT4610, CT4710, CT5010, CT5110 CT5210, CT5310.	62.3
Hap 5	1	NT1910	6.6
Hap 6	1	NT2310	6.6
Hap 7	1	ND2910	6.6
Hap 8	1	ND3210	6.6
Hap 9	1	CT4910	6.6
Hap 10	1	BT6710	6.6
Hap 11	1	BT7010	6.6

TABLE 5. Haplotypes for the P26 gene of *Potato yellow vein virus* (PYVV).

Haplotype P26	Quantity of sequences	Sequence code	Frequency (%)
Hap 1	1	AJ508757.2	2.38
Hap 2	1	KX573903.1	2.38
Hap 3	1	KR998195.1	2.38
Hap 4	4	CT0426, NT2326, NT2426, CT4226	9.52
Hap 5	1	CT1126	2.38
Hap 6	1	CT1226	2.38
Hap 7	4	CT1326, CT1526, CT4726, CT5326.	9.52
Hap 8	3	CT1426, CT1626, CT1826.	7.14
Hap 9	4	CT1726, CT4126, CT4326, CT4426	9.52
Hap 10	2	NT1926, NT2026.	4.76
Hap 11	1	NT2126	2.38
Hap 12	1	NT2226	2.38
Hap 13	1	NT2526	2.38
Hap 14	2	NT2626 NT2826.	4.76
Hap 15	1	NT2726	2.38
Hap 16	2	ND2926; ND3026	4.76
Hap 17	2	ND3326; ND3426	4.76
Hap 18	1	ND3526	2.38
Hap 19	2	ND3626; ND3726	4.76
Hap 20	1	ND3826	2.38
Hap 21	1	CT3926	2.38
Hap 22	1	CT4026	2.38
Hap 23	1	CT4926	2.38
Hap 24	1	CT5126	2.38
Hap 25	1	BoT5226	2.38
Hap 26	1	CT6226	2.38

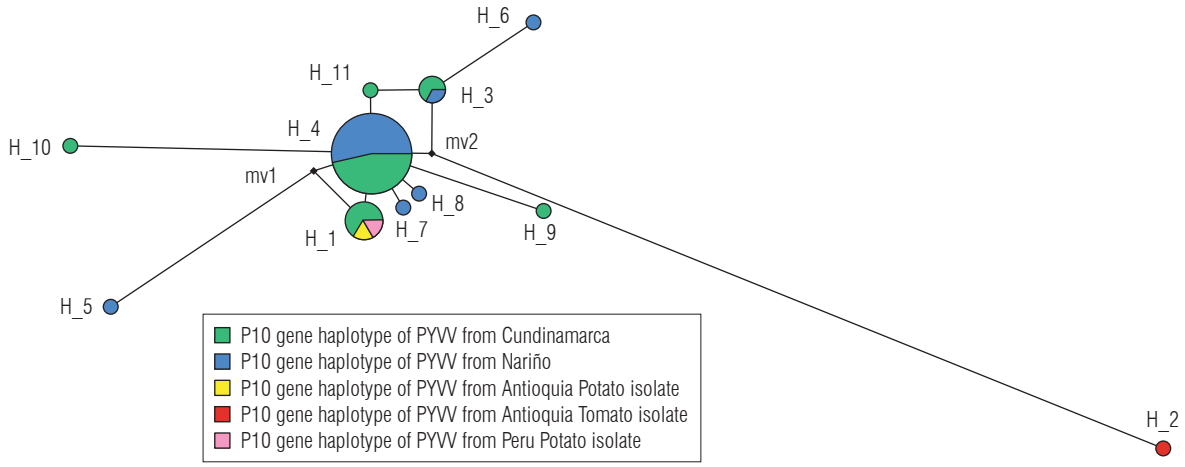

FIGURE 3. Haplotype network for the P10 gene of *Potato yellow vein virus* (PYVV).

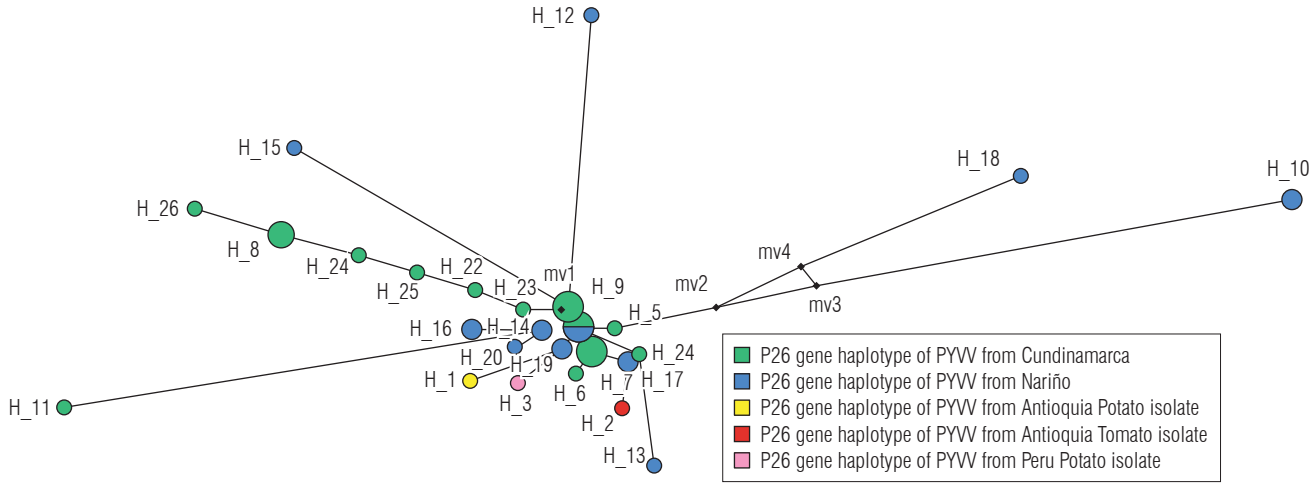

FIGURE 4. Haplotype network for the P26 gene of *Potato yellow vein virus* (PYVV).

Cundinamarca and Nariño. This represents the ancestral haplotype for the P26 gene (Fig. 4), and haplotypes 7 and 9 were formed by sequences only from Cundinamarca.

Phylogenetic relationships

The consensus maximum likelihood tree (MLT) for the P10 gene of PYVV (Fig. 5) showed three clades; GI, GII and GIII related to the host (StA or StP). The GI clade was formed by viral sequences from StP plants. The GII and GIII clades were formed by viral sequences from StA plants. The GII was formed by sequences from Cundinamarca and Peruvian and Colombian accessions, and the GIII was formed by viral sequences from Cundinamarca and Nariño and a viral isolate accession from the tomato.

The MLT for the P26 gene of PYVV (Fig. 6) showed seven clades (GI to GVII). In contrast to the P10 gene, there is no grouping by host. Additionally, there is no specific relationship by geographical origin. Nevertheless, the viral sequences from Nariño are in groups GIV, GVI, and GVII; and the Cundinamarca sequences are in groups GIII and GV. Groups GI and GII are formed by sequences from the three provinces. The three reference accessions from PYVV-P26 gene were also in this group.

Genetic distances between sequences of the P10 and P26 genes from PYVV isolates

The similarity percentage for the P10 gene was between 96 and 100% except for sequences KX573902.1 and NT1910,

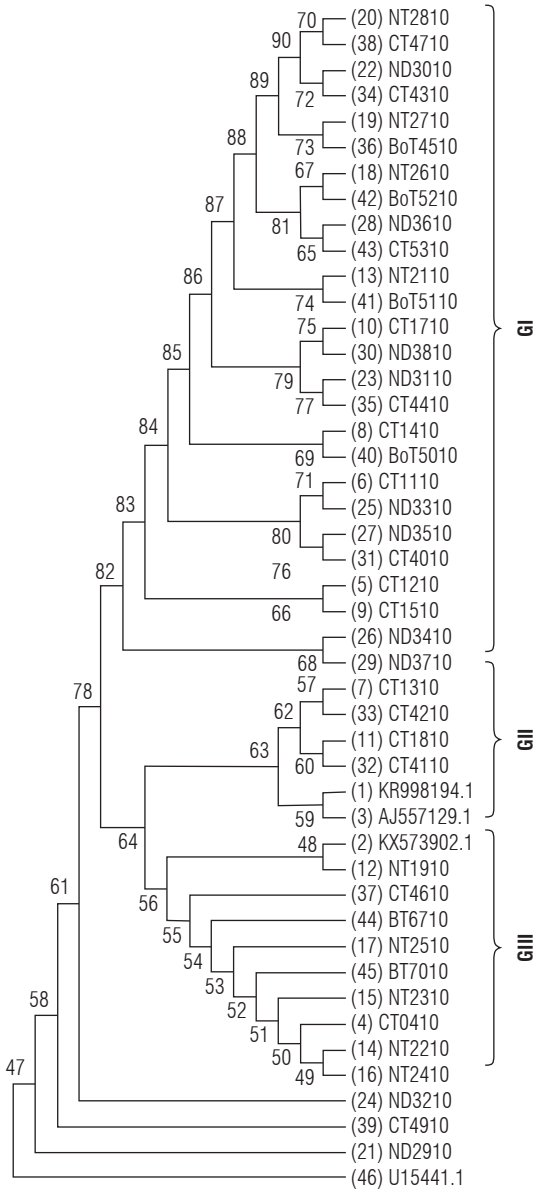

FIGURE 5. Consensus maximum likelihood tree of the P10 gene of *Potato yellow vein virus* (PYVV).

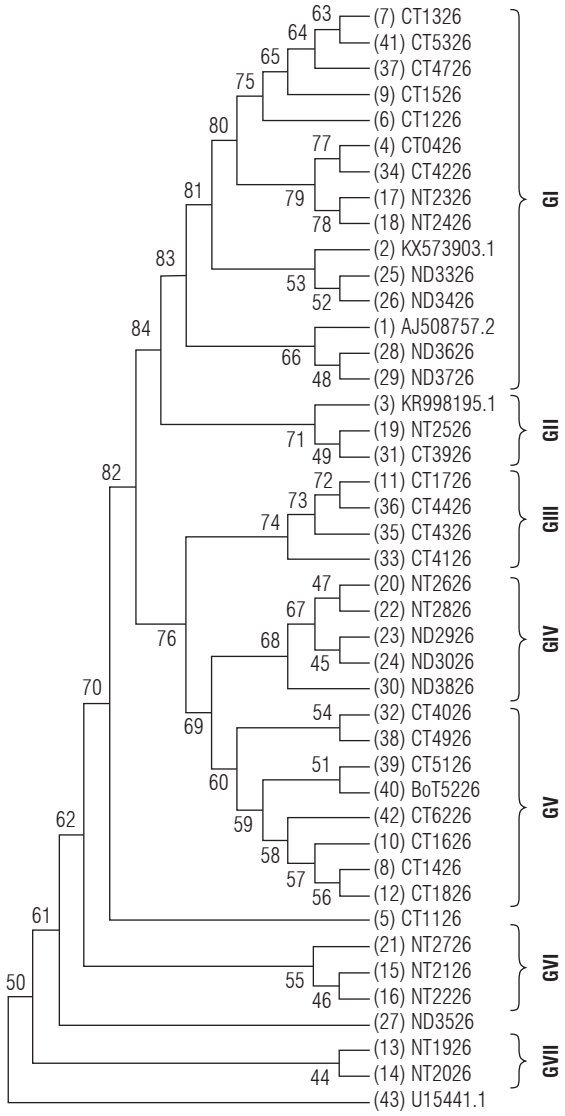

FIGURE 6. Consensus maximum likelihood tree of the P26 gene of the Potato yellow vein virus (PYVV).

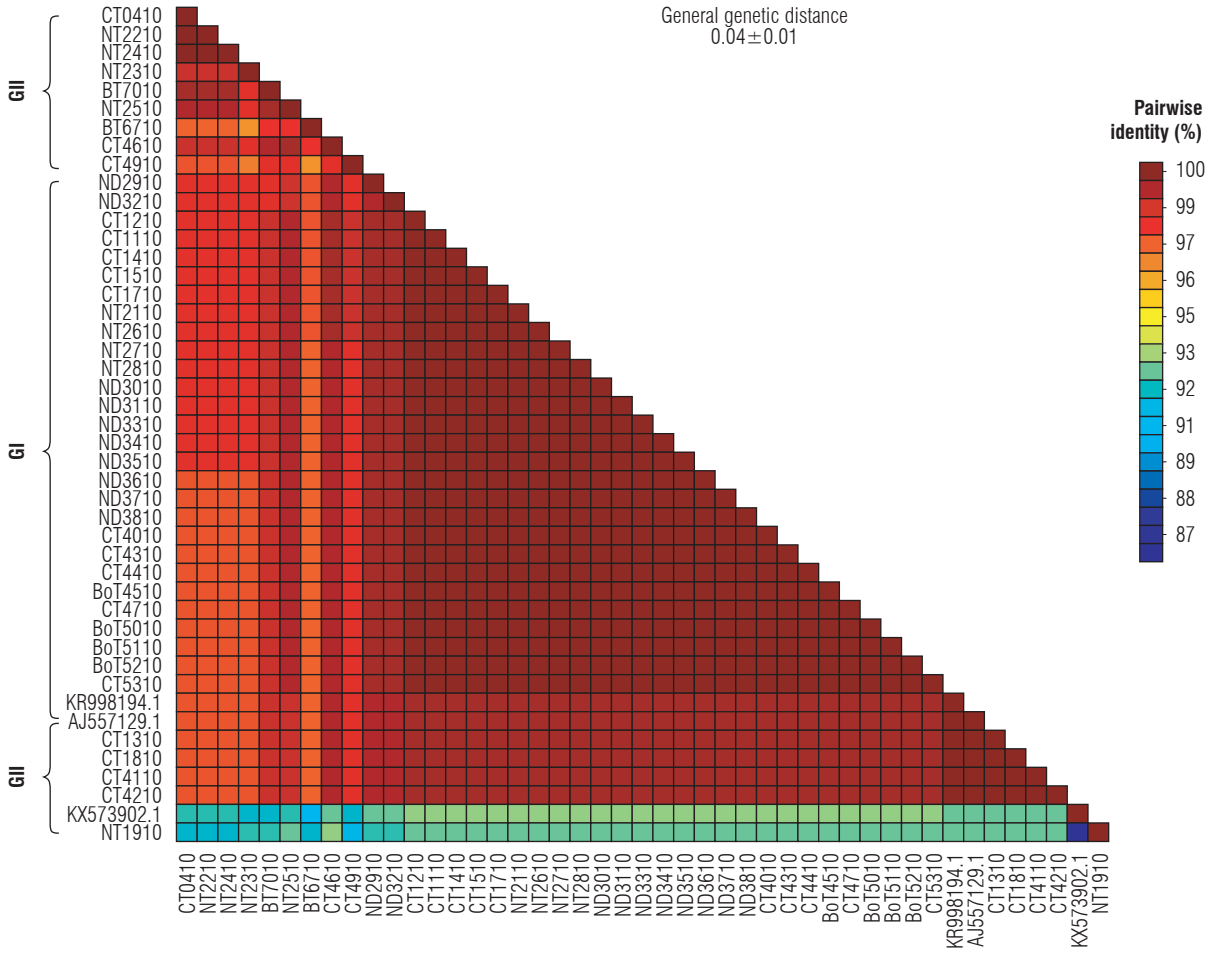

FIGURE 7. Matrix of genetic distances between sequences for the P10 gene isolates of *Potato yellow vein virus* (PYVV).

which show a similarity of between 87 and 93% respectively. In addition, the general average of genetic distances for the complete sequences was 0.04±0.01 (Fig. 7). Similarly, the matrix for the P26 gene was between 93 and 100% except for sequences NT1926, NT2026, NT2126, NT2226, ND3526, and NT2726 with percentages of similarity between 76 and 93%. The general mean of genetic distances for this gene was 0.09 ± 0.01 (Fig. 8).

Variability, diversity and neutrality tests of the P10 and P26 genes of PYVV

In general, the P26 gene of PYVV showed 252 variable sites (S) from 693 total sites (36.4% variable sites in the gene), 333 total mutations (η), a nucleotide and haplotype diversity of 0.0603 and 0.97, and a rate between non-synonymous and synonymous mutations (ω(dN/dS)) of 0.18. In the Tajima's D and Fu and Li tests significant negative values of -2.27 (<0.01) and -1.07 were found (Tab. 6, grey highlight).

The P10 gene of PYVV had 30 S of 255 total sites (11.8% variable sites in the gene), 31 η, nucleotide and haplotype diversity of 0.028 and 0.60, and a rate between non-synonymous and synonymous mutations (ω(dN/dS)) of 0.35. In the Tajima's D and Fu and Li tests, significant negative values of -2.45 (<0.001) and -4.70 were found (Tab. 6).

Internally in the groups generated for the P10 and P26 genes, greater variability was observed in isolates from tetraploid (StA) plants than in diploid (StP).

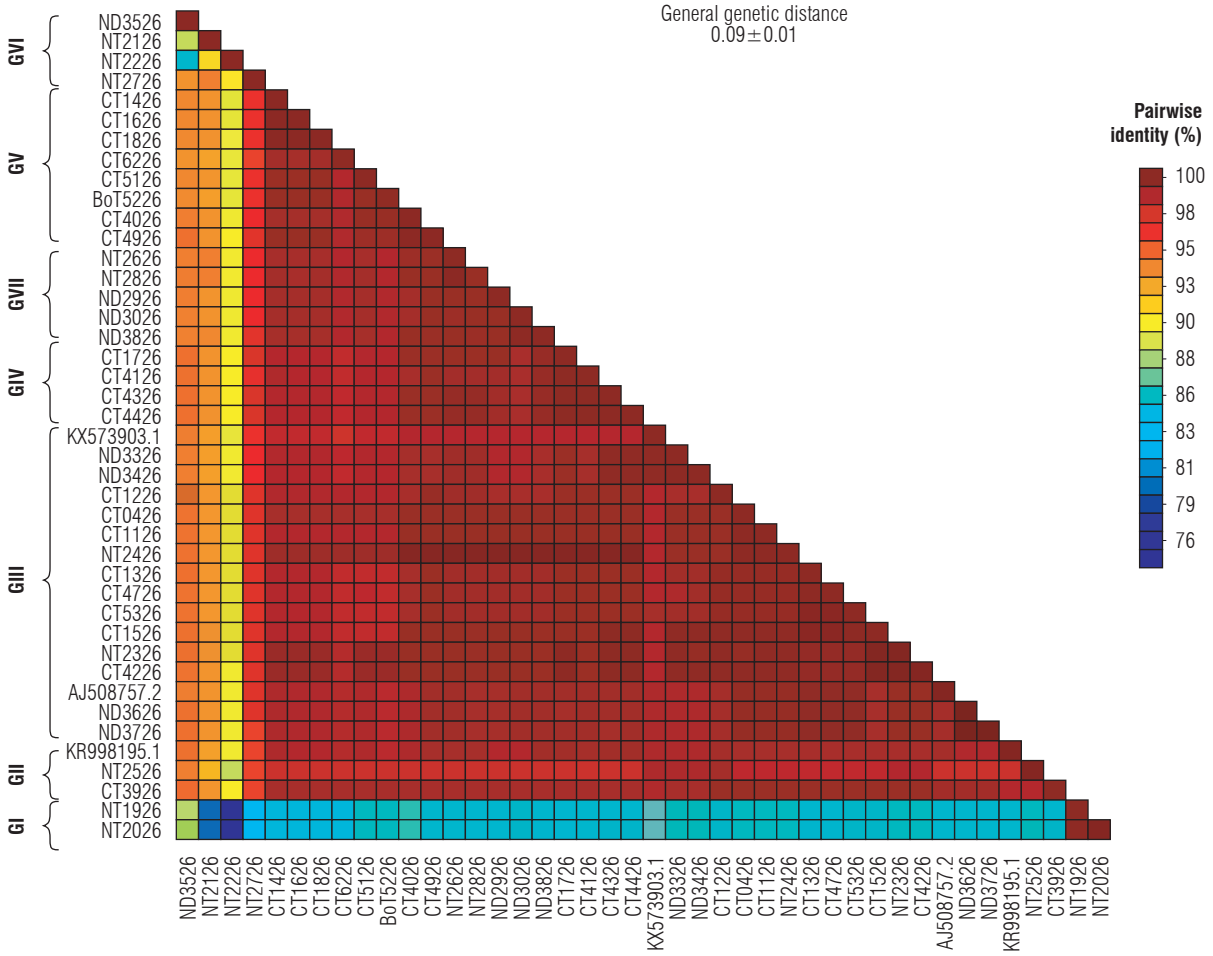

FIGURE 8. Matrix of genetic distances between sequences for the P26 gene isolates of *Potato yellow vein virus* (PYVV).

TABLE 6. Estimated parameters of genetic variability for the P26 and P10 genes of *Potato yellow vein virus* (PYVV).

Gene	Genotype	P	S	η	π	h	Hd	dN	dS	ω(dN/dS)	TD	FD
	StA	C+N	252	333	0.0926	16	0.95	0.025	0.0894	0.28	-2.30 *	-2.02
	StP	C+N	48	50	0.0199	6	0.92	0.017	0.0428	0.40	-1.53	-1.52
P26	StA+StP	C	236	239	0.0440	11	0.95	0.004	0.0253	0.17	-2.61***	-4.14
	StA+StP	N	253	290	0.124	12	0.96	0.054	0.1824	0.29	-1.50	-0.11
	StA+StP	C+N	132	138	0.0603	26	0.97	0.064	0.3501	0.18	-2.27**	-4.63
	StA	C+N	93	107	0.0437	9	0.56	0.154	0.3845	0.40	-2.55 ***	-4.15
	StP	C+N	3	3	0.0021	4	0.49	0.005	0.0197	0.25	-1.600	-1.87
P10	StA+StP	C	89	95	0.053	7	0.45	0.243	0.4571	0.53	-2.35**	-2.93
	StA+StP	N	19	20	0.012	7	0.61	0.026	0.0783	0.33	-1.86*	-2.34
	StA+StP	C+N	30	31	0.028	11	0.60	0.143	0.358	0.35	-2.45***	-4.70

Genotype: StA: *S. tuberosum* Andigena group; StP: *S. tuberosum* Phureja group. C: Cundinamarca; N: Nariño; P: Geographical origin; S: total variable sites; η: number of total sites; π: nucleotide diversity, average of nucleotide differences per site; h: total haplotype number; Hd: haplotype diversity; dN: average number of even differences by synonymous; dS: average number of even differences by nonsynonymous sites; ω(dN/dS): dN/dS ratio; TD: Tajima's D test; FD: Fu and Li D test. *$P<0.05$, **$P<0.01$, ***$P<0.001$.

Discussion

RNA viruses can accumulate mutations in their genomes because their RNA-dependent RNA polymerases have weak corrective activity, unlike DNA polymerases. Consequently, some viral species can have greater genetic variability (Domingo and Holland 1997; Sanjuán *et al.*, 2010; Sanjuán and Domingo, 2016). This makes viruses with RNA genomes more variable than those with a DNA genome. Therefore, populations of RNA viruses tend to evolve rapidly and adapt to the environmental conditions derived from the interaction with their vectors (García-Arenal and Fraile, 2001; García-Arenal and Fraile, 2011). Mutation and recombination are processes that influence genetic diversity since the first one introduces specific changes in the sequences generating new variants, and the second one is a process by which segments of different genetic variants are crossed during the replication process that allows the movement of variants to produce new haplotypes. In this sense, recombination does not create mutations but creates new combinations between pre-existing mutations (Pérez-Losada *et al*., 2015). The characterization of the genetic variability of viral populations supplies important information about the processes involved in the evolution of the virus and its dispersion, where epidemiology is crucial for the design of reliable diagnostic tools that allow selecting those genes that are more conserved in time and space, as well as for the development of efficient and long-lasting disease control strategies.

According to Rubio *et al*. (2013), the *Closteroviridae* family is not very variable because it is subject to negative selection, long-distance migrations, recombination processes, interaction between viral strains, and host-virus and virus-vector interaction. These are the main forces that exert pressure in the *Closteroviridae,* according to analyses carried out on structural genes such as CP. Moncef (2010) performed an analysis of the complete genomes for some species of genera in the *Closteroviridae* and reported that these viruses are subject to positive selection processes with dN/dS values ranging from 1.508 to 2.599. In addition, that author observed that only the *Closterovirus* (Citrus Tristeza Virus or CTV) and the *Ampelovirus* (Grapevine Leafroll-associated Virus 3 or GLRaV-3) are the only ones that have recombination processes; but this author did not detect any recombination processes for criniviruses.

In this study, we reported for the first time the amplification by RT-PCR and the acquirement of nucleotide sequences of the P26 and P10 genes of PYVV, from symptomatic and non-symptomatic samples of different potato varieties from Nariño, Cundinamarca, and Boyacá in Colombia for the variability study.

The haplotype analysis showed that both the P26 and P10 genes have high haplotype numbers with high diversity and low frequency (haplotype 4 with a frequency of 62.3% for the P10 gene and haplotypes 4, 7, and 9 with a frequency of 9.5% for P26 gene). Haplotype 4 for both genes corresponds to the most ancestral gene, and diversification has been generated from it in the different provinces evaluated in this study. In a study conducted by Chávez *et al.* (2013), 33 different haplotypes for the CP gene of PYVV were found, in which haplotype 28 was the most frequent with a value of 16.7. The most frequent haplotype groups for P10 and P26 genes of PYVV do not group sequences by host genotype or by geographic origin for the P10 gene, except for haplotypes with a single sequence. However, the P26 gene haplotypes do discriminate both by origin and genotype (except for the ancestral haplotype). This is supported by the maximum likelihood tree that shows the clades to be associated with the sequences of the specific haplotypes. This situation is not seen in the P10 gene tree, since no clear differentiation between the groups was observed by genotypes or geographical origin. This could be due to the increase in the use of vegetative seeds infected by the PYVV virus of different geographical origins, which allowed the spread of the virus in the potato-producing regions and, therefore, the distribution of variants in the geographical areas analyzed in this study. This, in turn, favors the development of new variants and recombinants when subjected to pressure in different environments. It is worth mentioning that the sequences NT1910 of the P10 gene and NT1926, NT2026, NT2126, NT2226, ND3526, NT2726 of the P26 gene showed differences between 87% and 93% for the P10 gene and 76% and 93% for P26 gene. This indicates that speciation processes are being developed in PYVV both in diploid and tetraploid plants from Nariño (Townsend, 2014; Koloniuk *et al*., 2018).

The Tajima's D and Fu and Li global tests showed a significant negative value for both genes, indicating that in the PYVV populations for these two genes, there was an excess of low-frequency haplotypes (rare alleles in high frequency). This low frequency could indicate that it is a population that is diversifying after a bottleneck or that there was a selection process that decreased the frequency of abundant haplotypes, increasing those that were rare at the time and that could be explained again by the flow of infected material between geographic regions. How-

ever, when analyzing the isolates of the StP host, negative values (but not significant) were observed for the Tajima test, indicating that this subpopulation of the PYVV was undergoing a neutral process or contraction period, or that the number of sequences obtained from StP was limited to offer enough statistical robustness.

When observing the estimated parameters of genetic variability and diversity, we can affirm that the P26 gene has a global nucleotide diversity of 0.0603, but when the diversity of the hosts is observed, it can be seen that in plants of the species StA is much higher than 0.0926 compared to the global and plants of the StP species. This indicates that in this study the sequences obtained from the StA potato plants of the P26 gene provide most of the diversity. This result is repeated with the P10 gene, in which the global diversity is 0.028 and that provided by the isolates from the StA and the StP is 0.0437 and 0.0021, respectively. Similar results were obtained by Chávez-Bedoya *et al.* (2014), who reported the diversity of the CP, Hsp70 and CPm genes of the PYVV, which are higher for StA hosts in all cases (0.010, 0.016 and 0.046, respectively) and lower in the StP (0.008, 0.005 and 0.084, respectively) except for the CPm gene. This may occur since the StP group is diploid and the StA group is tetraploid, which implies that the virus could be subjected to higher selection pressure by having to evade the silencing machinery in a larger genome. On the other hand, it is possible to state that the nucleotide diversity is greater in the P26 gene than in the P10 gene. However, that diversity is generated by a larger number of synonymous mutations (dS) that neither affects the composition of amino acids nor the three-dimensional structure of the protein. This maintains protein function and consequently takes the populations of the PYVV to a process of negative selection. The fact that the P26 and P10 genes are more diverse than the CP, CPm, and HSP70 of the PYVV defines the functions that each one offers to the virus, since the P26 and P10 are associated with pathogenicity in *Crinivirus*, while those studied by Chávez-Bedoya *et al.* (2014) are structural genes that must be highly conserved to guarantee the correct assembly of the virion and the transmission through the vector.

This research allows the determination of the variability of the P10 and P26 genes, which is higher in isolates from the *S. tuberosum* Andigena group than in the Phureja group infected with PYVV. In addition, the influence of humans as a vector for the dispersion of the virus among potato producing regions in Colombia was established. This research also allows the planning of new studies aimed at evaluating the functionality of PYVV proteins P10 and P26 and the relationship of the variants to the presence or absence of symptoms in the field, the degree of severity, and the impact on the production and colonization of new hosts.

Acknowledgments

The authors thank the financial support of the International Development Research Centre (IDRC) and Global Affairs Canada (GAC), through the Canadian International Food Security Research Fund (CIFSRF), which funded the Project Scaling up the Production of More Nutritious Yellow Potatoes in Colombia No. 108125 (More Nutritious Potatoes Project).

Literature cited

Bandelt, H.J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37-48. Doi: 10.1093/oxfordjournals.molbev.a026036

Barragán, C.E. and M. Guzmán-Barney. 2014. Molecular detection of *Potato yellow vein virus* in the natural whitefly vector *Trialeurodes vaporariorum*, Westwood. Rev. Prot. Veg. 29(3), 168-176.

Chaves-Bedoya, G., M. Guzmán-Barney, and L. Ortiz-Rojas. 2013. Genetic structure and evidence of putative Darwinian diversifying *selection* in *Potato yellow vein virus* (PYVV). Agron. Colomb. 31(2), 161-168.

Cubillos, K. and M. Guzmán-Barney. 2015. Molecular variability of three genes of *Potato vein yellow virus* infecting *Solanum tuberosum*, using single strand conformational polymorphism. Acta Biol. Colomb. 20(1), 233-237. Doi: 10.15446/abc.v20n1.40993

Domingo, E. and J. Holland. 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51,151-178. Doi: 10.1146/annurev.micro.51.1.151

Eliasco, E., I.C. Livieratos, G. Müller, M. Guzman, L.F. Salazar, and R.H. Coutts. 2006. Sequences of defective RNAs associated with *Potato yellow vein virus.* Arch. Virol. 151, 201-204. Doi: 10.1007/s00705-005-0625-7

Erkiş-Güngör, G. and B. Çevik. 2019. Genetic Diversity and phylogenetic analysis of *Citrus tristeza virus* isolates from Turkey. Adv. Virol. 2019, 7163747. Doi: 10.1155/2019/7163747

EEPO (European and Mediterranean plant protection Organization). URL: https://gd.eppo.int/taxon/PYVV00 (accessed 7 July 2019).

Fu, Y.X. and W.H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133, 693-709.

Gamarra, H., C. Chuquillanqui, and G. Müller. 2002. Transmisión del virus del amarillamiento de las venas de la papa en variedades y clones de *Solanum tuberosum* L. Proceedings of the Entomological National Convention. Potato International Center (CIP), Lima.

García-Arenal, F. and A. Fraile. 2011. Population dynamics and genetics of plant infection by viruses. pp. 263-281. In: Caranta, C., M.A. Aranda, M. Tepfer, and J.J. Lopez-Moya (eds.). Recent Advances in Plant Virology. Caister Academic Press, Norfolk, UK.

García-Arenal, F., A. Fraile, and J.M. Malpica. 2001. Variability and genetic structure of plant virus populations. Annu. Rev. Phytopathol. 39, 157-186. Doi: 10.1146/annurev.phyto.39.1.157

Grimsley, N., B. Hohn, T. Hohn, and R. Walden. 1986. "Agroinfection," an alternative route for viral infection of plants by using the Ti plasmid. Proc. Natl. Acad. Sci. USA. 83 (10), 3282-3286. Doi: 10.1073/pnas.83.10.3282

Guzmán, M., E. Ruiz, N. Arciniegas, and R. Coutts. 2006. Occurrence and variability of *Potato yellow vein virus* in three departments of Colombia. J, Phytopathol. 154, 748-750. Doi: 10.1111/j.1439-0434.2006.01174.x

Genbank. 2019. URL: https://www.ncbi.nlm.nih.gov/genome/?term=Potato+yellow+vein+virus (accessed June 2019).

Guzmán-Barney, M., P.A. Rodríguez-Burgos, and J. Calderón-Romero. 2013. Detección por inmunoimpresión de *Potato yellow vein virus* (PYVV) en diferentes órganos de papa: herramienta sencilla y útil en diagnóstico del virus de amarillamiento de nervaduras de papa y certificación de semillas. Editorial UN, Bogota.

Hernández, A. and M. Guzmán-Barney. 2014. *Potato yellow vein virus* detection in different organs of *Solanum tuberosum* Phureja group cv Criolla Colombia by conventional and real time qRT-PCR. Rev. Colomb. Biotecnol. 16(1), 74-85. Doi: 10.15446/rev.colomb.biote.v16n1.44226

Huang, X. and A. Madan. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9, 868-877. Doi: 10.1101/gr.9.9.868

Kiss, Z., V. Medinaand, and W.B. Falk. 2013. *Crinivirus* replication and host interactions. Front. Microbiol. 4(99), 1-11. Doi: 10.3389/fmicb.2013.00099

Koressaar, T. and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23(10), 1289-1291. Doi: 10.1093/bioinformatics/btm091

Koloniuk, I., T. Thekke-Veetil, J.S. Reynard, P. Mavrič, J. Přibylová, J. Brodard, I. Kellenberger, T. Sarkisova, J. Špak, J. Lamovšek, S. Massart, T. Ho, J.D. Postman, and I.E. Tzanetakis. 2014. Molecular characterization of divergent Closterovirus isolates infecting *Ribes* species. Viruses 10(7), 2-11. Doi: 10.3390/v10070369

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7), 1870-1874. Doi: 10.1093/molbev/msw054

Librado, P. and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451-1452. Doi: 10.1093/bioinformatics/btp187

Livieratos, I., E. Eliasco, G. Muller, R. Olsthoorn, L. Salazar, W. Pleij, and R.H Coutts. 2004. Analysis of the RNA of *Potato yellow vein virus*: evidence for a tripartite genome and conserved 3'-terminal structures among members of the genus *Crinivirus*. J. Gen. Virol. 85(7), 2065-2075. Doi: 10.1099/vir.0.79910-0

Martelli, A., N. Abou Ghanem-Sabanadzovic, A.A. Agranovsky, M. Al Rwahnih, V.V. Dolja, C.I. Dovas, M. Fuchs, P. Gugerli, J.S. Hu, W. Jelkmann, N.I. Katis, V.I. Maliogka, M.J. Melzer, W. Menzel, A. Minafra, M.E. Rott, A. Rowhani, S. Sabanadzovic, and P. Saldarelli. 2012. Taxonomic revision of the family *Closteroviridae* with special reference to the grapevine leafroll-associated members of the genus *Ampelovirus* and the putative species unassigned to the family. J. Plant Pathol. 94(1), 7-19.

Medina, V., G. Rodrigo, T. Tian, V.V. Dollja, and M. Achon. 2003. Comparative cytopathology of *Crinivirus* infections in different plant hosts. Ann. Appl. Biol. 143, 99-100. Doi: 10.1111/j.1744-7348.2003.tb00274.x

Medina, V., M.R. Sudarshanab, T. Tianb, K.S. Ralstonb, H.H. Yehb, and B.W. Falk. 2005. The *Lettuce infectious yellows virus* (LIYV)-encoded P26 is associated with plasmalemma deposits within LIYV-infected cells. Virology 333(2), 367-373. Doi: 10.1016/j.virol.2005.01.012

Moncef, B. 2010. Selective pressure, putative recombination events and evolutionary relationships among members of the family *Closteroviridae*: A proposal for a new classification. Biochem. Syst. Ecol. 38(6), 1185-1192.

Muhire, B.M., A. Varsani, and D.P. Martin. 2014. SDT: a Virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9: e108277. Doi: 10.1371/journal.pone.0108277

Muñoz, D., P.A. Gutiérrez, and M. Marín. 2016. Detección y caracterización molecular del *Potato virus Y* (PVY) en cultivos de papa (*Solanum tuberosum* L.) del norte de Antioquia, Colombia. Rev. Prot. Veg. 31, 9-19.

Nei, M. and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, USA.

Ñústez, C.E. 2011. Variedades de papa liberadas en Colombia. URL: http://www.papaunc.com/variedades-liberadas-por-la-universidad-nacional-de-colombia (accessed 7 July 2019).

Offei, S., N. Arciniegas, G. Müller, M. Guzman-Barney, L. Salazar, and R. Coutts. 2003. Molecular variation of *Potato yellow vein virus* isolates. Arch. Virol. 149(4), 821-827. Doi: 10.1007/s00705-003-0250-2

Pérez-Losada, M., M. Arenas, J.C. Galán, F. Palero, and F. González-Candelas. 2015. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30, 296-307. Doi: 10.1016/j.meegid.2014.12.022

Rubio, L., J. Guerri, and P. Moreno. 2013. Genetic variability and evolutionary dynamics of viruses of the family *Closteroviridae*. Front. Microbiol. 4(151). Doi: 10.3389/fmicb.2013.00151

Ruiz, L., A. Simón, C. García, L. Velasco, and D. Janssen. 2018. First natural crossover recombination between two distinct species of the family *Closteroviridae* leads to the emergence of a new disease. PLoS One 13(9), e0198228. Doi: 10.1371/journal.pone.0198228

Salazar, L.F., G. Müller, M. Querci, L.J. Zapata, and R.A. Owens. 2000. *Potato yellow vein* disease: its host range, distribution in South America, and identification as a *Crinivirus* transmitted by *Trialeurodes vaporariorum*. Ann. Appl. Biol. 137(1), 7-19. Doi: 10.1111/j.1744-7348.2000.tb00052.x

Stewart, L.R., V. Medina, M.R. Sudarshana, and B.W. Falk. 2009. *Lettuce infectious yellows* virus-encoded P26 induces plasmalemma deposit cytopathology. Virology 388, 212-220. Doi: 10.1016/j.virol.2009.03.016

Sanjuán, R., M.R. Nebot, N. Chirico, L.M. Mansky, and R. Belshaw. 2010. Viral mutation rates. J. Virol. 84(19), 9733-9748. Doi: 10.1128/JVI.00694-10

Sanjuán, R. and P. Domingo-Calap. 2016. Mechanisms of viral mutation. Cell. Mol. Life Sci. 73, 4433-4448. Doi: 10.1007/s00018-016-2299-6

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585-595.

Templeton, A.R., K.A. Crandall, and C.F. Sing. 1992. A cladistic analysis of phenotypic association with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132, 619-633.

Townsend, A. 2014. Defining viral species: making taxonomy useful. J. Virol. 11(1) 131. Doi: 10.1186/1743-422X-11-131

Untergasser, A., I. Cutcutache, T. Koressaar, J. Ye, B.C. Faircloth, M. Remm, and S.G. Rozen. 2012. Primer3 - new capabilities and interfaces. Nucleic Acids Res. 40(15), e115. Doi: 10.1093/nar/gks596

Wale, S., B. Platt, and N. D. Cattlin. 2008. Diseases, pests and disorders of potatoes: a colour handbook. CRC Press, Boca Raton, USA. Doi: 10.1111/j.1365-3059.2008.01936.x

Wang, J., M. Turina, L.R. Stewart, J.A. Lindbo, and B.W. Falk. 2009. Agroinoculation of the *Crinivirus, Lettuce infectious yellows virus*, for systemic plant infection. Virology 392(1), 131-136. Doi: 10.1016/j.virol.2009.06.034