



RESEARCH ARTICLE

High throughput sequencing for the detection and characterization of new virus found in arracacha (*Arracacia xanthorrhiza*)

Joao De Souza^{1,2} ; Segundo Fuentes⁴ ; Giovanna Müller⁴ ; Heidy Gamarra⁴ ; Mónica Guzmán⁵ ; Wilmer Cuellar³ ; Jan Kreuze^{4,*} 

¹ Escuela de Posgrado, Programa Doctoral de Ciencias e Ingeniería Biológica, Universidad Nacional Agraria La Molina, Lima 12. Peru

² Departamento de investigación, desarrollo e innovación, Ecofertilizing SAC. Lima. Peru.

³ International Center for Tropical Agriculture (CIAT), Apartado Aéreo 6713, Cali. Colombia.

⁴ International Potato Center (CIP), Apartado 1558, Lima 12. Peru.

⁵ Instituto de Biotecnología (IBUN), Universidad Nacional de Colombia, Bogota. Colombia.

* Corresponding author: j.kreuze@cgiar.org (J. Kreuze).

Received: 30 March 2021. Accepted: 19 September 2021. Published: 8 October 2021.

Abstract

Arracacha is a root crop cultivated in several countries of South America. The commercial product, which is the storage root, has multiple nutritional properties: its fine starch makes it easily digestible, and it contains high levels of calcium and vitamin A. High throughput sequencing (HTS) technology has been applied successfully for virus discovery in different agricultural crops, and it has been proposed to apply it in routine pathogen detection. Using HTS, novel sequences related to crinivirus and vitivirus were identified in apparently symptomless arracacha, which were assembled into contigs located in different positions of the genome. Based on those sequences' primers were designed to amplify corresponding sequences from further arracacha accessions and potato plant samples collected from farmers' fields in Colombia and Peru. It was possible to determine the near complete genomes of these viruses. In addition, sequences related to an enamovirus and ST9-like RNA were also identified in arracacha plants using HTS.

Keywords: Arracacha; siRNA sequencing; crinivirus; vitivirus; enamovirus.

DOI: <https://dx.doi.org/10.17268/sci.agropecu.2021.051>

Cite this article:

De Souza, J., Fuentes, S., Müller, G., Gamarra, H., Guzmán, M., Cuellar, W., & Kreuze, J. (2021). High Throughput Sequencing for the Detection and Characterization of new virus found in arracacha (*Arracacia xanthorrhiza*). *Scientia Agropecuaria*, 12(4), 471-480.

1. Introduction

Arracacha (*Arracacia xanthorrhiza* Bancroft) is a root crop originating in the Andean region which is also its main production area. Arracacha roots have beneficial nutritional properties (such as high calcium content, starch and vitamin A), and is industrialized for processing into infant foods in Brazil. Commercial production takes place in Colombia, Brazil, Ecuador and Venezuela, where the total coproduction area is estimated to exceed 30000 hectares (Hermann, 1997). In Peru and Bolivia, arracacha is also widely cultivated for subsistence and generally only some surpluses go to local markets (Hermann, 1997). Several arracacha viruses and their characteristics have been described, but it is not clear how they affect crop yield. To date, the known viruses that infect arracacha are *Arracacha virus A*, *Arracacha virus B*, *Arracacha potyvirus 1* (AP1), *Arracacha virus 3* (AV3), *Potato black ringspot virus* (PBRV), *Arracacha mottle virus* (AMoV), *Arracacha virus*

V and *Arracacha virus I* (Jones & Kenten, 1978; Kenten & Jones, 1979; Santa Cruz, 1994; Lizarraga, 1997; Lizarraga, Chuquillanqui & Jayasinghe, 1994; Orllio et al., 2009; Orllio et al., 2018; Oliveira et al., 2017). In addition, Brunt et al. (1996) described: *Arracacha potyvirus Y* and *Arracacha latent carlavirus* but it is not clear if they are distinct from AP-1 and AV-3, respectively.

Because arracacha is propagated vegetatively, it can accumulate viruses over generations leading to degeneration, resulting in reduced root production (Orllio et al., 2009). Also, it can be a reservoir of viruses for other crops, since it has been reported that different Andean roots and tubers can be infected by the same viruses, such as *Potato yellowing virus* (PYV) in yacon (Silvestre et al., 2020) or *Andean potato latent virus* (APLV), *Potato leafroll virus* (PLRV) and *Potato virus T* (PVT) in ulluco (Lizarraga, Santa Cruz & Jayasinghe, 1996a; Lizarraga, Santa Cruz & Salazar, 1996b; Lizarraga et al., 1999).

Thus, viruses that can be detected in arracacha could be a source of emerging viruses and cause problems, not only in this crop, but also nearby crops such as potatoes (Jones & Kenten, 1978; Kenten & Jones, 1979) due to global warming and the movement of vectors to places where they used to be they were not found. Some viruses are difficult to detect by inoculation to indicator host plants, because their vectors are unknown, they are difficult to transmit mechanically and/or difficult to graft to other plants. Because of that, next-generation or high throughput sequencing (HTS) technology has emerged as a powerful alternative to detect and identify viruses that have been difficult to identify using traditional methods (Wu et al., 2015). HTS technologies enable, in a relatively short period of time, the characterization of known plant viruses and the discovery of novel ones, without prior knowledge (Fitzpatrick et al., 2021; Kutnjak et al., 2021). This technology has been successfully used with several crop species, and its use has increase dramatically in recent years. Because HTS is based on agnostic sequencing of nucleic acids, the detection capacity of this technology is in principal independent of environmental conditions, virus genetic variability, and host responses (Kavalappara et al., 2021). It is well documented that HTS based detection is at least equivalent to conventional biological and molecular methods in sensitivity, specificity, repeatability and reproducibility (Bester et al., 2021; Soltani et al., 2021).

The objective of this study was to identify and characterize the sequences related to viruses in arracacha samples by HTS and screen field samples of potato for the sequences associated with viruses initially detected in arracacha.

2. Materials and methods

2.1. Plant material

Leaves of 4 asymptomatic arracacha plants grown in the greenhouse in CIP-Huancayo station (Junin, Peru) were collected and sent to CIP-Lima to extract total RNA using Trizol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's instructions. Origin of the samples are showed in Figure1.

2.2. Sequencing and *in-silico* assembly

Small RNAs were isolated as described previously (Fuentes et al., 2012), and sent to FASTERIS Life Sciences SA (Plan-les-Quates, Switzerland) for library preparation and sequencing (bulked together with 16 samples of potato and sweetpotato to save cost), on Illumina HiSeq2000. Reads obtained were assembled *denovo* using Velvet v0.6.04 (Zerbino & Birney, 2008) as described previously (Fuentes et al., 2012). Contigs with similarity to viruses were identified by BLASTx against GenBank viral sequences. The gapped regions were successfully amplified from the large RNA fraction of the arracacha RNA extracts by specifically designed primers, after which the corresponding amplified fragments were cloned into pGEM-T Easy vector (Promega, WI, USA) and sequenced (Macrogen-Korea). Sequences obtained were then edited and annotated using Vector NTI v.9 software package (Invitrogen). Attempts to amplify the 5' and 3' un-translated regions (UTRs) were done using Mod-One and Ban-two linkers. Phylogenetic and molecular evolutionary analyses were conducted using MEGA7 (Kumar, Stecher, & Tamura, 2016). Primers designed were used later to detect the identified viruses in potato and arracacha samples.

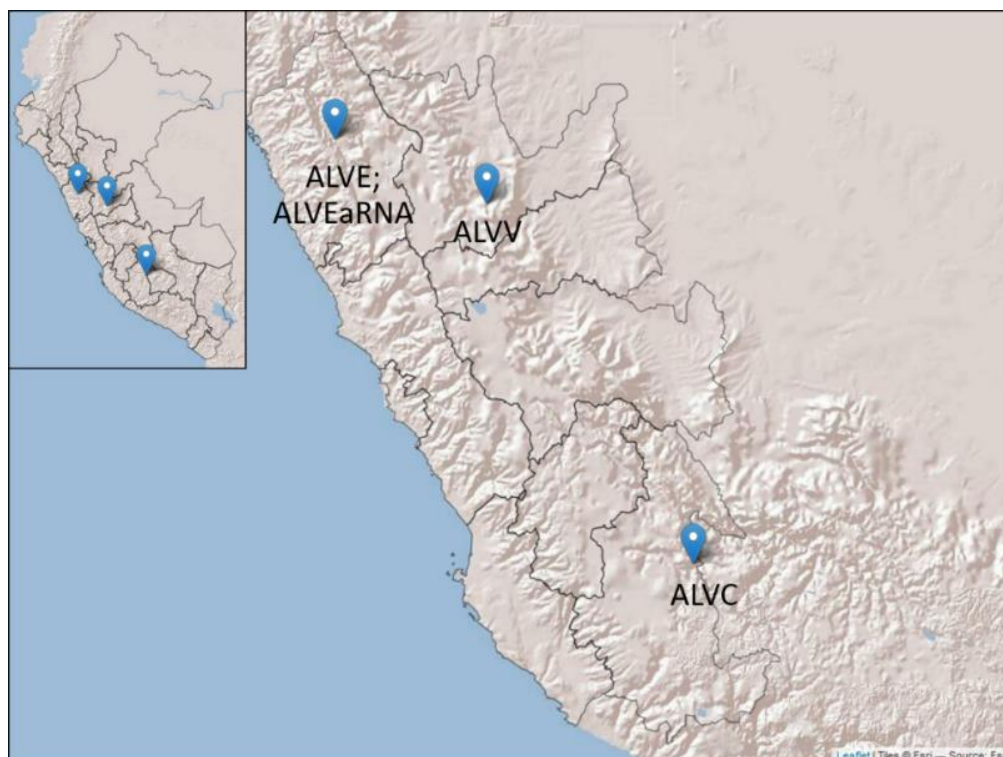


Figure 1. Geographic origin of sequenced viruses. ALVC (from Ayacucho), ALVV (from Huánuco), ALVE and ALVEaRNA (from Ancash).

2.3. Insect transmission experiments

The arracacha plants where virus-associated sequences were found were placed separately in a cage together with 400 vector insects in a glass greenhouse. *Myzus persicae*, *Bemisia tabaci*, *Trialeurodes vaporariorum*, *Empoasca* spp. were used to realize the transmission experiments. The insects were left with the plants for an acquisition period of 48 hours inside cages, to later place 20 insects to independent potato seedlings inside cages for an inoculation period of 48 h. After that time, the insects were eliminated manually and by applying insecticides. 20 repetitions were made in 2 potato varieties: Canchan and Perricholi (40 potato seedlings in total). One month later, the inoculated plants were evaluated by RT-PCR for vitivirus and crinivirus.

3. Results and discussion

3.1. Viruses genome assembly

Small RNA library sequencing produced a total of 11,286,650 reads between 21-24 nucleotides (nts) (https://research.cip.cgiar.org/confluence/display/cpx/GAF13-14/GAF14_21-24.fastq.gz). In total, 22 and 21 contigs corresponded to RNA1 and RNA2 of a novel crinivirus, 8 contigs corresponded to a vitivirus, 20 corresponded to an enamovirus and 8 to ST9-like RNAs. Primers designed to amplify the regions between assembled contigs are shown in **Table 1**. Amplification of 3' and 5' ends was attempted using various approaches but was only successful for the vitivirus 3'-end.

It is proposed to name the new crinivirus as Arracacha latent virus C (ALVC), the vitivirus as Arracacha latent virus V (ALVV), the enamovirus arracacha latent virus E (ALVE) and the ST9-like RNA Arracacha latent virus E associated RNA (ALVEaRNA). **Oliveira et al. (2017)** described the identification of another vitivirus from arracacha in Brazil, which they named Arracacha virus V (AVV). ALVV showed 74% nt identity to AVV over the complete genome, while it had 74% nt and 85% aa identity in the replicase, and 82% nt and 92% aa in the coat protein. Thus, AVV and ALVV represent distant isolates of the same species.

To verify presence of these viruses, 20 arracacha accessions grown in the field in CIP-Lima were screened using primers designed for each virus (targeting the ORF1a, CP and replicase region for ALVC, ALVV and ALVE, respectively). Results are provided in **Table 2**.

Subsequently, Illumina barcoded small RNA libraries of positive accessions were prepared as describe above at CIP Virology Lab and sent for sequencing (as a bulk sample of three plants infected with ALVV, ALVC or ALVE+ALVEaRNA, respectively) to Boyce Thompson Institute (Ithaca, NY, USA) on an Illumina HiSeq 2000 machine. It produced 1,558,633 reads between 21-24 nts (<https://research.cip.cgiar.org/confluence/display/cpx/GAF300>), of which 3210 reads aligned with ALVV (0.2% of the total siRNA), and 1483 reads aligned with ALVC [0.09% of the total siRNA: 678 reads aligned with RNA1 (0.04%) and 805 reads aligned with RNA2 (0.05%)]. Also, 3930 reads aligned with ALVE (0.25%) and 14137 aligned with ALVEaRNA (0.9%). The near complete sequences of ALVC

and ALVV, and partial sequences of ALVE and ALVEaRNA were deposited in GenBank (accession no. KY451034, KY451035 for ALVC, KY451036 for ALVV, MF136435 for ALVE and MF136436 for ALVEaRNA). Like other criniviruses, ALVC consisted of two single stranded RNAs of positive polarity. 7461nt of RNA1 and 7305nt of RNA2 were determined, containing most of the coding regions but lacking the 3' and 5' ends. An analysis of the coding capacity of RNA1 showed that it contains at least three ORFs (**Figure 2**) which corresponded to ORF1a, ORF1b, ORF2. The determined partial ORF1a contained 5613bp (1870aa, 21.3kDa), and contained conserved motifs associated with virus replication in closteroviridae: a methyltransferase domain (MTR), a helicase domain (HEL) (**Martelli et al., 2012**), and an RNA-dependent RNA polymerases (RdRp) domain in ORF1b which is thought to be translated by a +1 ribosomal frameshift from ORF1a. ORF3 (153pb, 51aa) encoded a predicted protein of 5.84 kDa (p6) but showed no significant similarity to corresponding proteins of other criniviruses.

Analysis of the sequenced RNA2 of ALVC revealed that it encoded at least seven ORFs (**Figure 1** homologous to HSP70h (61.27 kDa), cell-to-cell movement protein (p59, 59.61 kDa), CP (28.18 kDa), and CPm (79.03 kDa). Other ORFs were also predicted, which were similar in size (6.24 kDa (p6.5), 10.20 kDa (p9), and 28.06 kDa (p26)) and genomic positions to those found in other criniviruses but lacked significant aa similarity. ORF4 (p9) and ORF6 (mCP) begin with the codon GTG. Since it was unable to obtain sequences from the ends of the RNAs, it remains possible there are additional ORFs present 3' or 5' of the determined sequences as found as in some criniviruses. In the phylogenetic analysis based on HSP70h and CP coding region (**Figure 2**), ALVC clustered together with PYYV, BYVaV, SPaV and BPYV, which belong to group 1 of the criniviruses, which are transmitted by *Trialeurodes vaporariorum*, together with those of group 3 (**Martelli et al., 2012**).

ALVV had a similar genome structure as AVV (**Oliveira et al., 2017**). The determined sequence was predicted to have at least five open reading frames (ORFs), which correspond to the replicase, 20kDa-like protein (p20), movement protein (MP), coat protein (CP) and nucleic acid binding protein (NABP) based on homology with other vitiviruses. The replicase ORF was sequenced only partially (4944bp, 1648aa, 18.74 kDa) but contained three conserved motifs found among proteins associated with virus replication (MTR, HEL and RdRp). The predicted 20kDa-like protein (371bp, 123aa, 14.09 kDa) also found in other vitiviruses, begins with a GTG codon and it was similar to an ORF found also in AVV, but not reported by **Oliveira et al. (2017)**. The predicted MP (885bp, 295aa, 32.80kDa) and CP (576bp, 192aa, 21kDa) were clearly related with their homologs in other vitiviruses, whereas the nucleic acid binding protein (NABP; 330bp, 110aa, 12.58kDa) has weak homologies to proteins with RNA-binding properties. The two phylogenetic trees generated (**Figure 3**) confirmed the relationship of the vitivirus from arracacha with another species in this genus.

Table 1

List of primers used to amplify sequences related to the putative viruses detected in arracacha

Primer name	Sequence (5'-3')	Amplified size (bp)	Amplified region
Arracacha latent virus C – RNA1			
Crini-RNA1-F1	TGTCACGACGTCAAAAACCTTAGGG	446	ORF1a
Crini-RNA1-R1	CCCAAAATGACACCAAGGAACCC		
Crini-RNA1-F2	ATGGTTGGAGAATCGTAAGGAGAC	936	ORF1b
Crini-RNA1-R2	GCAATCATAGCGTCGACCAATACAC		
Crini1-Arra-Fa	GCAAGAACATCCATGTATGTTGCC	516	ORF1a
Crini1-Arra-Ra	CGTATTTACCATTTTTGAACCAACGC		
Crini1-Arra-Fb	GAGTGGATGAAGAAGCTCAGGAATACC	2446	ORF1a
Crini1-Arra-Rb	GCATCATCAATTACATTCCTCATATCCC		
Crini1-Arra-Fc	GCCAAATACAATGCAAAACATTCGG	1815	ORF1b
Crini1-Arra-Rc	CTCCCTTTAACCATCATCTTATCC		
Crini1-arra-F1	TCTCTCAAATGATCAGTTCGCTGC	1321	ORF1a
Crini1-arra-R1	GCAGTTACTGTCAACAATCTCTGGTGG		
Crini1_Arra_iF4	GAATGGAAGATTTCTGTACGGTGAGC	467	ORF1b
Crini1_Arra_iR4	YAAMMCGCGTATATGACTCTAAGGGAA		
Arracacha latent virus C – RNA2			
Crini-RNA2-F1	TGATGTACGATTTTCGGTGGTGGC	619	HSP70
Crini-RNA2-R1	GGGATAGACATAGCCTTGCGAACCC		
Crini-RNA2-F2	TGTGGTTAAATCACAACCTGATGCC	348	CP
Crini-RNA2-R2	CGTTGACGTAATCACCAGTAGAATTC		
Crini2-Arra-Fa	GCTTATCGATAACTGATGGTGAGACG	587	HSP70
Crini2-Arra-Ra	TGATGTTGTCAATATCAGGCCACC		
Crini2-Arra-Fb	AAGTCTGTGCTTATGGGTGTTCTATGC	1297	HSP70 - p59
Crini2-Arra-Rb	CCCTTGATAAACCCCTGATCAAGTTCCAG		
Crini2-Arra-Fc	GGAGGTGATATGAATGATTACTTGCTTCA	1530	p59 - CP
Crini2-Arra-Rc	TGATTGATTCAAAGCAGCTTGAGCC		
Crini2-Arra-Fd	CCAATCCATGGAGACAGTATATGAGGG	1955	CP - mCP
Crini2-Arra-Rd	TTCATAGCACCTTTACTGTCTTCCAC		
Crini2-d1F4	GGTTTGAACGATGGTATGCTTGAATTG	1471	mCP – p27
Crini2-d1R4	CAGTTGCCAACTAGTCAACGTCGG		
Crini2_Fn1a	TGTCATTGTGTGATCTTCTTCTTCTATTG	1014	HSP70
Crini2_Rn1	TCAAGAATGTTAAATGATGATTTTCCGAG		
Arracacha latent virus V			
Viti-CP-F1	CTTTAAAATTGGCTAAATGGG	305	CP - NABP
Viti-CP-R1	GTGAAAAGCCCCACACTTCAG		
Viti-Rep-i1F1b	AGGATGAAGCGGAGGTACATTCAG	497	Replicase
Viti-Rep-i1R1	AAAGCAATCATCGAGCTAACCATGAG		
Viti-RdRp-iFn	GTTTACAGTGATGCGGTTCTCTGG	2268	Replicase - CP
Viti-CP-iRn	TGAATCACTTGAGCTTCGCTAACGC		
Viti-F6	CTCTCTAGCATAGGGGTTTACTTTCAGTG	629	Replicase
Viti-R6	GACATGATGAGCCCCGAAACTCTC		
Viti-long-F	GCAAAACCGAGCCTCAATCAGCG	3545	Replicase
Viti-long-R	GCCAATGCAAAATACCTAAGAAGCG		
Viti-3UTR-Fn	CTATGGAGGCTCCGTTCTTGGG	416	NABP - 3'UTR
Arracacha latent virus E			
Enamo-F1	GTGCCTACCGACGACAAGCGTAG	674	replicase
Enamo-R1	CAAGTCATCCTCCATGAGCCAC		
Enamo-F2	CTCCTCATTTGAGAACAGACGATTC	988	replicase
Enamo-R2	CCTCCGTCGGCAGATGCCTG		
Enamo-F3	CCGGAGATCGCGTTACCTGT	800	Aphid protein
Enamo-R3	CCTCAGTCTCTCATACCAATCC		
Enamo-RdRp-F2	TGAGCCTGGATACTCTGGTTTGCC	1078	replicase
Enamo-RdRp-R2	TCGGAGCCATAGCTAGGCCAGC		
Enamo-RdRp-F3	TGCAAGGCCTGGTTCCACGC	817	replicase
Enamo-RdRp-R3	CAATCAGTCGGATAGCACCACGC		
Enamo-Aphid-F1	AGCCCCGAAGACGGAGAAGC	1030	aphid protein
Enamo-Aphid-R1	CAAATGTAGCCTCATTGGGGGTTTC		
Enamo-F5	CATCCACAATATAGCCATTTCAACTGAG	685	replicase
Enamo-R5	GTTGATTAGCCGATCCCATCCG		
Arracacha latent virus E associated RNA			
ST9-like_F1	GACAGCGGTGCGAGTGCAGTCTC	135	Replicase
ST9-like_R1	CCGTTCCACCTGCCACTTGC		
ST9-like_F2	GCAAGTGGCAGGTGGAACGG	684	Replicase
ST9-like_R2	CTCGGGCTAAAGCACGCTCG		
ST9-like_F3	CCAGCCGCGAAGTCCAAGGTTTC	989	Replicase
ST9-like_R3	CCGAAGCCATGTGAAAGGATAC		

Table 2

Detected viruses in accession of arracacha collected in Germplasm Bank from International potato Center

Accession number	Evaluated viruses				
	ALVV	ALVC	AMoV	PBRVS	ALVE
204004	-	-	-	-	+
204009	-	+	-	-	+
204011	-	-	+	-	+
204051	-	+	+	-	+
204059	-	-	-	-	+
204060	-	+	-	-	+
204061	-	+	-	-	+
204062	-	+	-	-	+
204063	+	-	+	-	+
204065	-	-	-	-	+
204067	-	-	-	-	+
204068	+	-	-	-	-
204069	-	-	+	-	+
204070	-	+	-	-	+
204071	-	+	-	-	+
204072	+	+	-	-	+
204073	-	+	+	-	+
204079	-	-	+	-	+
204080	-	+	+	-	+
204087	-	-	+	-	+

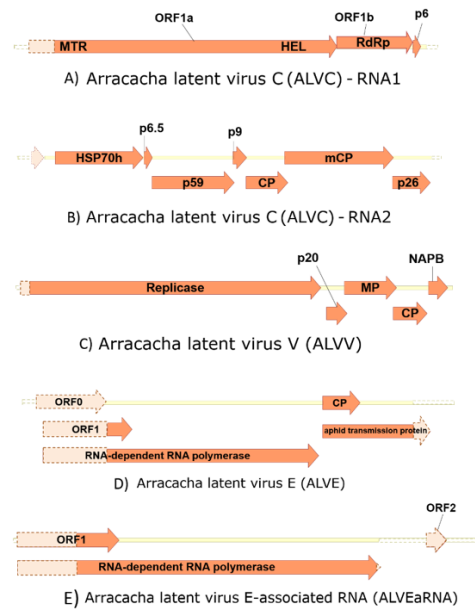
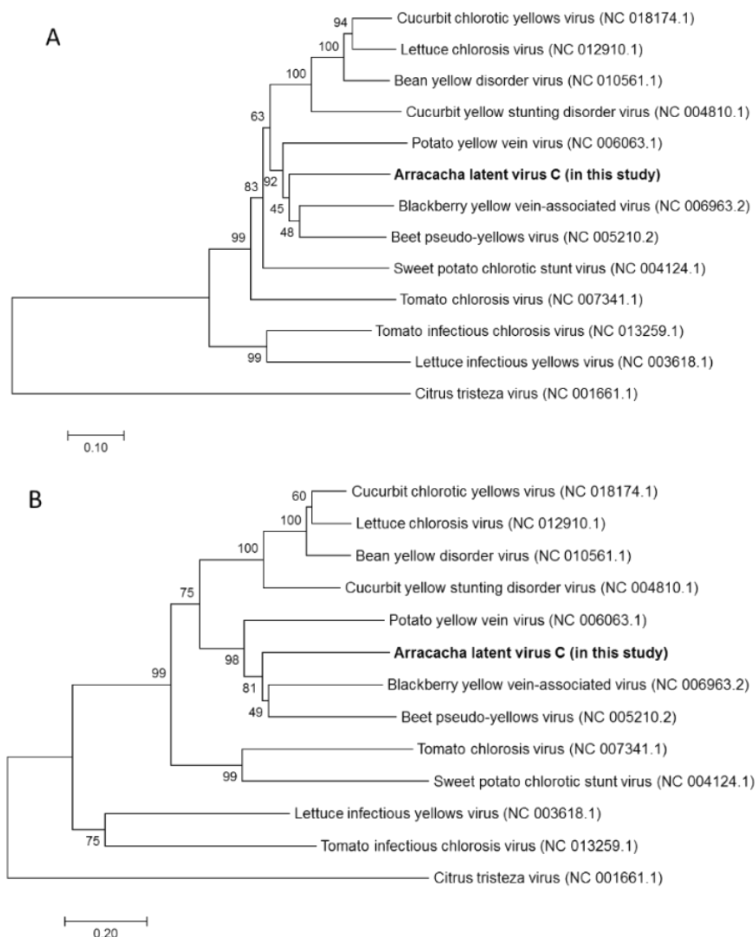
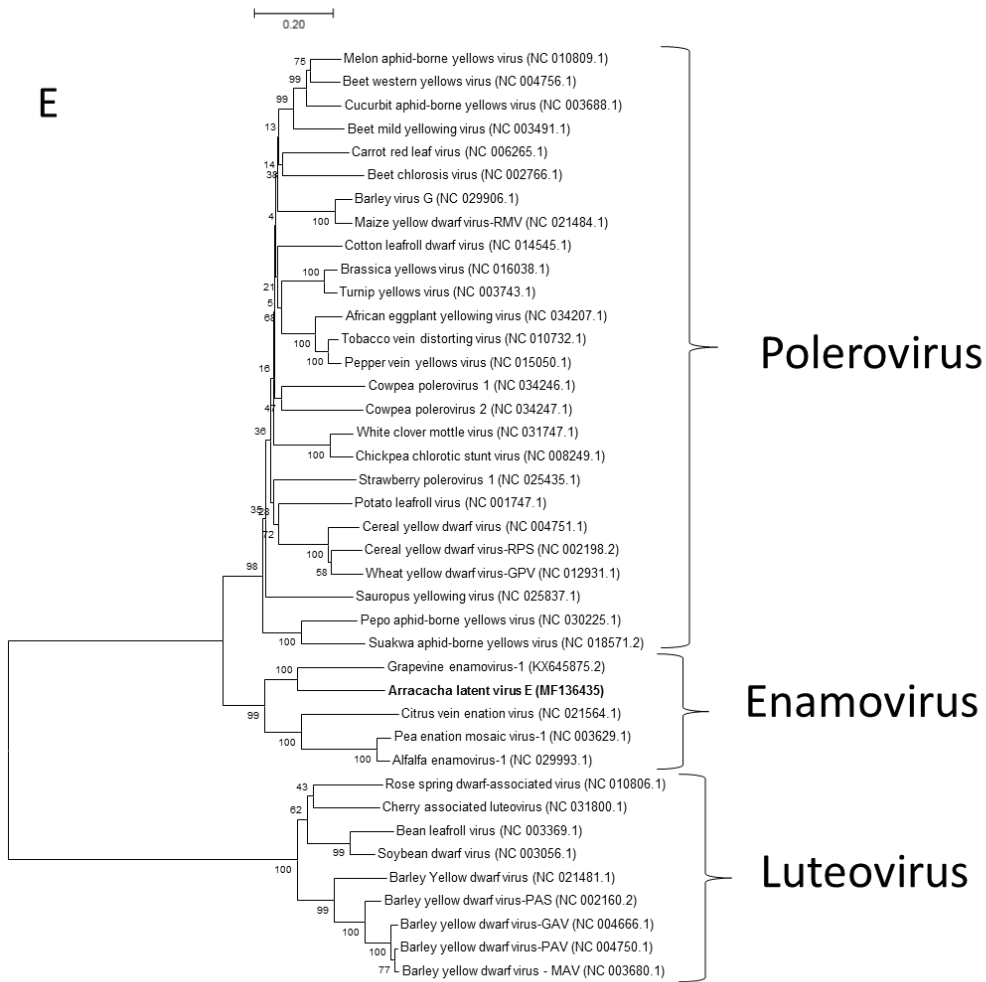
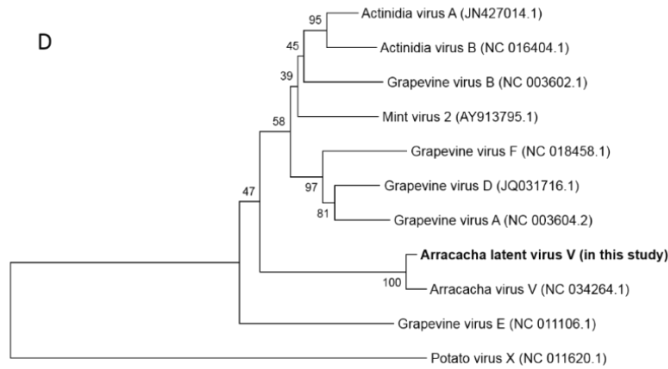
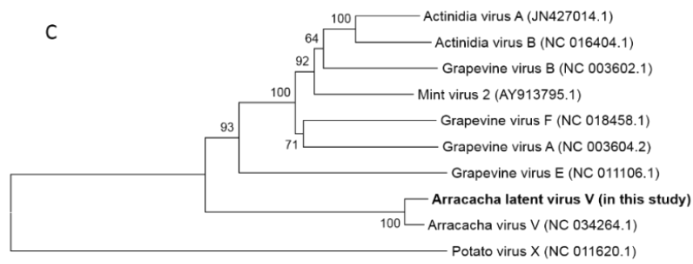


Figure 2. Genome structure of ALVC RNA1 (A), RNA2 (B), ALVV (C), ALVE (D) and ALVEaRNA (E). Orange open arrow indicate open reading frames, with the names of predicted gene products inside them. Regions surrounded by solid lines indicates they were sequenced by RT-PCR using primers designed from the sequences obtained by next generation sequencing technology, whereas lighter shaded areas with dotted lines represent unsequenced regions estimated according to related viruses.





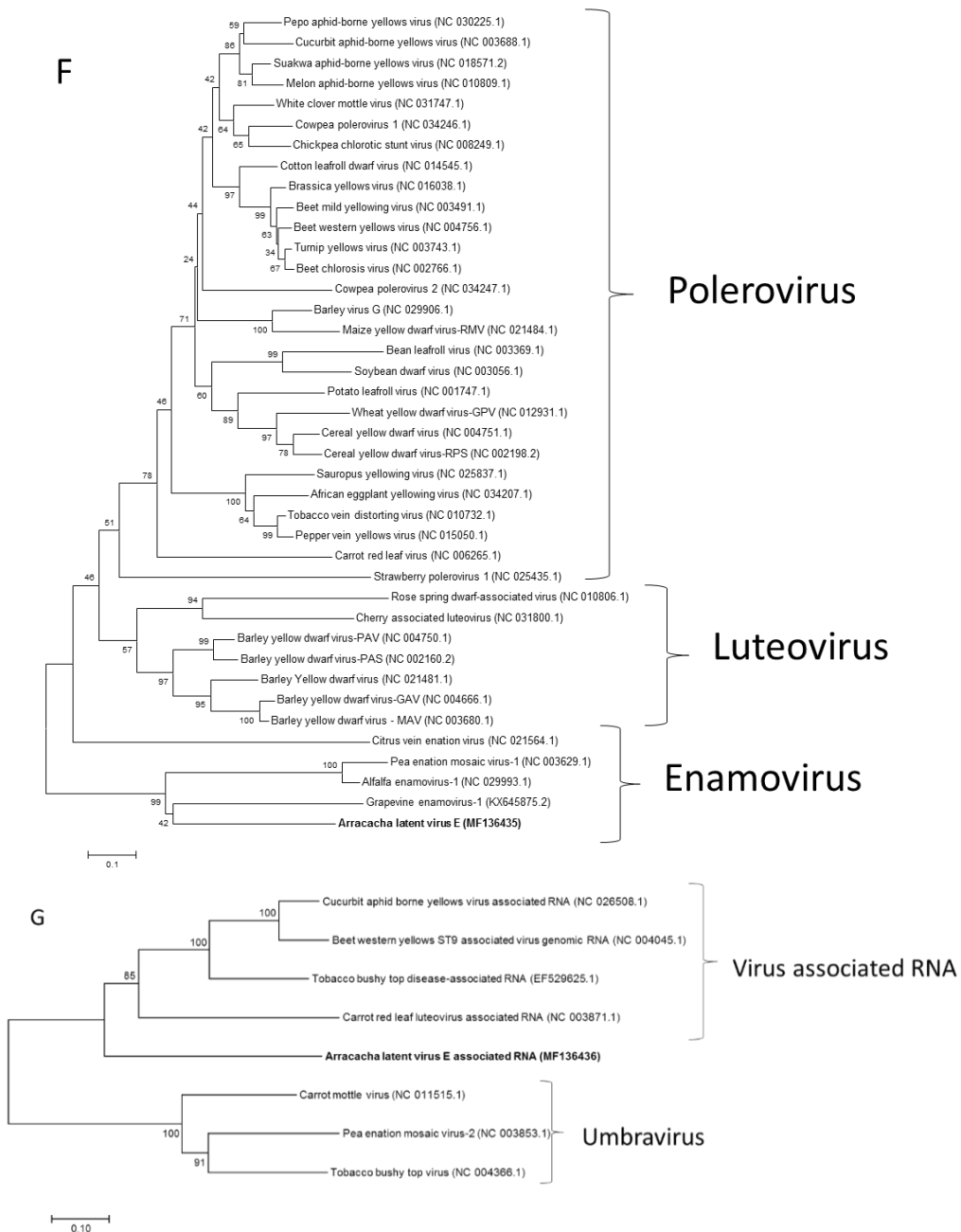


Figure 3. Phylogenetic trees based on an alignment of the HSP70 (A) and CP (B) for criniviruses; replicase (C) and CP (D) for vitiviruses; replicase (E) and coat protein (F) for enamoviruses; and (G) RdRp for virus-associated RNA. Arracacha latent virus C, V, E and Arracacha latent virus E associated RNA are indicated in bold in each phylogram. Sequences were aligned using the ClustalW algorithm, and the tree was generated using the Neighbour-Joining method. The evolutionary distances were computed using Poisson correction method and are in the units of the number of amino acid substitutions per site. Numbers on branches indicate percentage of bootstrap support out of 1000 bootstrap replications. Citrus tristeza virus, Potato virus X, Carrot mottle virus, Pea enation mosaic virus-2 and Tobacco bushy top virus were used as outgroup.

The determined sequence of 4035nts of ALVE was predicted to include four ORFs. The first two predicted ORFs were 5' incomplete and corresponded to P1 and P1-P2 fusion protein, which is translated through a -1 ribosomal frameshift from P1 (1185 aa, 152.7 kDa), and are involved in virus replication. The P1 contained a serine-like protease domain, and the P2 region the core domains of the viral RdRP as in other enamoviruses. The third ORF

corresponded to P3 (189 aa, 21.3 kDa) of luteovirids, and the last identified ORF, which was 3' incomplete is translated by readthrough of the P3 ORF to form a P3-P5 fusion protein. P3 is the coat protein (CP), whereas the CP readthrough extension (P5) of P3-P5 is thought to be an aphid transmission subunit of the virus. It was not possible to obtain sequence for the region encoding the ORFO characteristic of the members of family *Luteoviridae*.

In samples that were positive for ALVE, additional sequences were identified with similarity to Beet western yellows luteovirus ST9-associated RNA (BWYV ST9aRNA) and carrot red leaf virus-associated RNA (CRLVaRNA), by PCR test, which it was proposed to name Arracacha latent virus E-associated RNA (ALVEaRNA). The predicted partial ORF1 translated into a hypothetical protein with unknown function and a partial ORF2 into a putative RdRp is the results from readthrough of stop codon, similar to found in other ST9 like satellite RNAs. Additional ORFs can be found in related ST9 like satellite RNAs but were beyond the area sequenced in this study.

To evaluate if ALVC and ALVV were present in other plants, thirty-five additional arracachas maintained in greenhouse from CIP-Huancayo were evaluated for the presence of these viruses by RT-PCR. Also, because several potato viruses have been reported infecting Andean root crops and viceversa (Lizárraga et al., 1996a, 1996b; 1999), potato samples from Colombia (Chipaque - Cundinamarca, region near to Bogota) and from Peru (Huánuco and Cajamarca) were evaluated by RT-PCR to determine if these putative viruses were able to infect this crop. ALVC sequences were found in 25% of the tested arracachas (Fig. 3), whereas none of them tested positive for ALVV. From the evaluated Colombian potato samples, 20.61% and 5.15% resulted in positive amplification for ALVC and ALVV, respectively. Among the evaluated Peruvian potato samples, no sequence was detected in the 40 samples from Huánuco, while in the 41 samples from Cajamarca, 3 samples showed positive amplification for ALVC (7.31%) and 2 were positive for ALVV (4.87%). The PCR fragments were sequenced, aligned and submitted to phylogenetic analysis. The Colombian ALVC sequences formed a different group compared with the Peruvian ALVC sequences (Fig. 4a). On the other hand, all the ALVV sequences formed a single group (Figure 4b).

3.2. Virus transmission

Several attempts to transmit these viruses mechanically and by vectors were made. Mechanical inoculation on

Nicotiana clevelandii x *N. bigelovii*, *Gomphrena globosa*, *Chenopodium quinoa* and *Datura stramonium* were unsuccessful suggesting they may not be mechanically transmissible. This was not unexpected as members of the genus *Crinivirus* are phloem restricted (Martelli et al., 2012) and vitiviruses are known to be difficult to transmit mechanically (Adams et al., 2012). Graft transmission to *C. quinoa*, *C. amaranticolor*, *Datura stramonium* and *Physalis floridiana* using petioles and leaves as scions (due to morphological incompatibility of arracacha stems with these indicator plants) were also unsuccessfully attempted. Taking in consideration that whiteflies are generally the vectors of criniviruses from group 1 (Martelli et al., 2012) to which ALVC belongs, transmission by *Bemisia tabaci* biotype B and *Trialeurodes vaporariorum* were attempted with groups 20 insects twice according to Gamarra et al. (2010). For ALVV, the aphid *Myzus persicae* was chosen to attempt transmission, as vectors of vitivirus in herbaceous plants are generally aphids (Adams et al., 2012). Arracacha plants containing ALVC and ALVV sequences were placed in a separated cage with 400 adult insects under glasshouse conditions (20 °C, 70-95% RH, with 18h light and 6h dark). For the three-insect species, they were left on the source plants for a 48-hour acquisition access period before transferring 20 insects to each 20 individual healthy potato seedlings of two different cultivars (Canchán and Perricholi, 40 seedlings in total) in separate cages for a 48-hour inoculation access period (IAP). After this time, the insects were removed and plants sprayed with insecticide Applaud (Syngenta) and Confidor (Bayer) for whiteflies and aphids, respectively. Plants were evaluated by RT-PCR to confirm transmission one month after the IAP. ALVC and ALVV were not detected in any of the plants tested by PCR. Whereas transmission of ALVE was not attempted with whiteflies or aphids, leafhoppers (*Empoasca* spp.) found colonizing an arracacha field at CIP la Molina station were positive for ALVE by PCR. Subsequent attempts to transmit ALVE to indicator plants failed however, because the leafhoppers were so aggressive, they rapidly killed the plants.

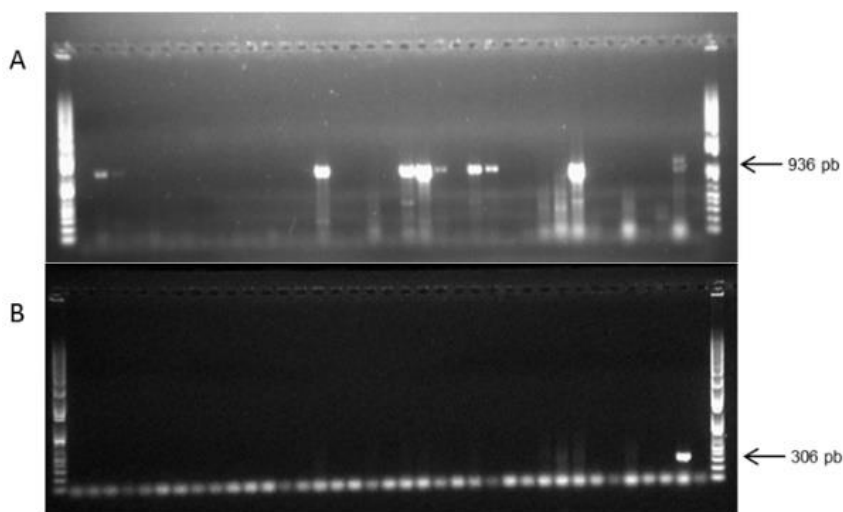


Figure 4. Electrophoresis gels showing arracacha accessions from CIP-Huancayo grown in greenhouse evaluated for sequences associated with crinivirus (A) and vitivirus (B), by RT-PCR (HSP70 and CP, respectively). The arrows indicate the size of the expected amplified product. First mixed sample was used as positive control. DNA ladder 1 kb plus (Invitrogen) was used as marker.

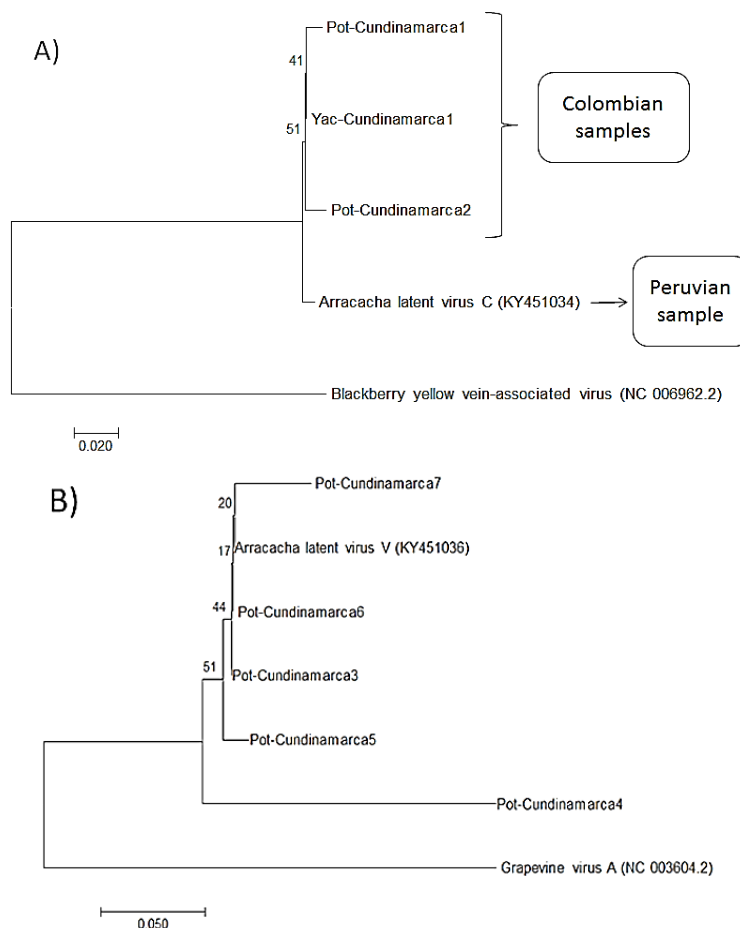


Figure 5. Phylogenetic tree showing the geographic relationships between Colombian and Peruvian samples that tested positive in RT-PCR to the new crinivirus and vitivirus detected in potato samples. Amplified region in ALVC correspond to ORF1a (A) and in ALVV to coat protein (B).

ALVC and ALVV belong to genera that typically are not known to have broad host ranges and it was therefore perhaps surprising to find them in such distantly related species as potato and arracacha (Figure 5). No obvious symptoms were associated with the presence of these viruses in either species and based on siRNA quantities, virus titers could be suspected to be quite low in arracacha at least and this may have affected results of our attempts at vector transmission. Obviously, the identified viruses are widespread in the Andean region at least from Colombia to Southern Peru, it is not necessarily the case that either arracacha or potato are their primary hosts even if both viruses can be maintained through vegetative cycles in arracacha at least. It would be interesting to screen weeds and wild plants found in association with crops in the Andes to identify plants that may serve as primary hosts and possible reservoirs from which these viruses enter cultivated crops.

4. Conclusions

In the present study we identified the presence of sequences corresponding to a new crinivirus (ALVC), a new enamovirus (ALVE) and a new ST9-like RNA (ELVEaRNA) in arracacha. Also, a vitivirus identified in this study (ALVV) was found to be closely related to AVV

identified from arracacha in Brazil and which was published while this manuscript was in preparation. Thus, this report expands the range of AVV to include Peru. Using the assembled sequences, primers were designed for all identified viruses and could be used to identify additional infected arracacha and potato plants. These primers present useful tools to screen germplasm or planting material to ensure plant health.

Sequence variation found between the amplified PCR fragments of samples suggest an actively evolving virus population in the Andes. Nevertheless, because presence of these virus sequences was not associated with any obvious symptoms, and virus titers appeared to be low, and we were unable to identify the means of transmission, many questions remain open. Future studies should focus on identifying alternate hosts, the likely means of transmission, and possible effects on yields in the different crops they infect.

Acknowledgements

This research was undertaken as part of, and funded by, the CGIAR Research Program on Roots, Tubers and Bananas (RTB) and supported by CGIAR Fund Donors (<http://www.cgiar.org/about-us/our-funders/>). This work was support by the financial contribution from ICGEB-TWAS-UNESCO/IBSP Joint Programme on Capacity Building in Basic Molecular Biology (Contract No.:

CRP/101005) and EU FP7 through the QBOL project (Grant No. 226482).

The authors gratefully acknowledge to Anngie Hernandez and Angela Villamil, from Universidad Nacional de Colombia, and to Rocío Silvestre, from International Potato Center, for the technical help. Also, we thank to Hugo Espinoza and Beder Guerrero for taking care the plants.

ORCID

J. De Souza  <https://orcid.org/0000-0002-3739-3122>

S. Fuentes  <https://orcid.org/0000-0001-8433-809X>

G. Müller  <https://orcid.org/0000-0003-0865-2946>

H. Gamarra  <https://orcid.org/0000-0003-2746-2849>

M. Guzmán  <https://orcid.org/0000-0002-2541-2031>

W. Cuellar  <https://orcid.org/0000-0003-4702-3237>

J. Kreuze  <https://orcid.org/0000-0002-6116-9200>

References

- Adams, M. J., Candresse, T., Hammond, J., Kreuze, J. F., Martelli, G. P., et al. (2012). Family Betaflexiviridae. In: King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (Eds.), *Virus Taxonomy - Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego, pp 920-941.
- Bester, R., Cook, G., & Maree, H. J. (2021). Citrus tristeza virus genotype detection using high-throughput sequencing. *Viruses*, 13(2), 168.
- Brun, A. A., Crabtree, K., Dallwitz, M. J., Gibbs, A. J., & Watson, L. (1996). *Viruses of plants*. CAB International, University Press, Cambridge, Great Britain. 1484 p.
- Fitzpatrick, A. H., Rupnik, A., O'Shea, H., Crispie, F., Keaveney, S., & Cotter, P. (2021). High Throughput Sequencing for the Detection and Characterization of RNA Viruses. *Frontiers in Microbiology*, 12, 190.
- Fuentes, S., Heider, B., Tasso, R. C., Romero, E., Zum Felde, T., & Kreuze, J. F. (2012) Complete genome sequence of a potyvirus infecting yam beans (*Pachyrhizus* spp.) in Peru. *Archives of Virology*, 157, 773-776.
- Gamarra, H. A., Fuentes, S., Morales, F. J., Glover, R., Malumphy, C., & Barker, I. (2010). Bemisia afer sensu lato, a vector of sweet potato chlorotic stunt virus. *Plant Diseases*, 94, 510-514.
- Hermann, M. (1997). Arracacha (*Arracacia xanthorrhiza* Bancroft). In: Andean roots and tubers: Ahipa, arracacha, maca and yacon. Promoting the conservation and use of underutilized and neglected crops. Hermann, M. and J. Heller (eds). 21. Institute of Plant Genetics and Crop Plant Research, Gatersleben/ International Plant Genetic Resources Institute, Rome, Italia, pp. 75-172.
- Jones, R. A. C., & Kenten, R. H. (1978). Arracacha virus A, a newly reconized virus infecting arracacha (*Arracacia xanthorrhiza*, Umbeliferæ) in the Peruvian Andes. *Annals of Applied Biology*, 90, 85-91.
- Kavalappara, S. R., Milner, H., Konakalla, N. C., Morgan, K., Sparks, A. N., et al. (2021). High throughput sequencing-aided survey reveals widespread mixed infections of whitefly-transmitted viruses in cucurbits in Georgia, USA. *Viruses*, 13(6), 988.
- Kenten, R. H., & Jones, R. A. C. (1979). Arracacha virus B, a second isometric virus infecting arracacha (*Arracacia xanthorrhiza*, Umbeliferæ) in the Peruvian Andes. *Annals of Applied Biology*, 93, 31-36.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7), 1870-1874.
- Kutnjak, D., Tamisier, L., Adams, I., Boonham, N., Candresse, T., et al. (2021). A primer on the analysis of high-throughput sequencing data for detection of plant viruses. *Microorganisms*, 9(4), 841.
- Lizarraga, C. (1997). Progress in the identification of the viruses infecting Andean root and tuber crops. Seminar, International Potato Center, Lima, Peru.
- Lizárraga, C., Chuquillanqui, C., & Jayasinghe, U. (1994). Una variante del virus del anillo necrótico de la papa (Potato black ringspot virus, PBRV) aislado de arracacha (*Arracacia xanthorrhiza*). *Fitopatología (Lima)*, 29(2), 144-149.
- Lizárraga, C., Santa Cruz, M., & Jayasinghe, U., (1996a) Detection of an isolate of Andean Potato latent virus in Ullucu (*Ullucus tuberosus* Caldas). *Plant Disease*, 80(3), 344.
- Lizárraga, C., Santa Cruz, M., & Salazar, L. F. (1996b) First report of Potato Leafroll virus in ullucu (*Ullucus tuberosus* Caldas). *Plant disease*, 80(3), 344.
- Lizárraga, C., Santa Cruz, M., Marca, J. L., & Salazar, L. F. (1999) La importancia de los virus que infectan a *Ullucus tuberosus* Caldas en el Perú. *Fitopatología*, 34(1), 22-28.
- Martelli, G. P., Agranovsky, A. A., Bar-Joseph, M., Boscia, D., Candresse, T., et al. (2012). Family Closteroviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., & Lefkowitz, E.J. (Eds.), *Virus Taxonomy - Classification and Nomenclature of Viruses. Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, San Diego. Pp: 987-1001.
- Oliveira, L. M., Orilio, A. F., Inoue-Nagata, A. K., & Blawid, R. (2017) A novel vitivirus-like sequence found in *Arracacia xanthorrhiza* plants by high throughput sequencing. *Arch Virol*, 162(7), 2141-2144.
- Orilio, A. F., Dusi, A. N., Madeira, N. R., & Inoue-Nagata, A. (2009). Characterization of a member of a new Potyvirus species infecting arracacha in Brazil. *Archives of Virology*, 154, 181-185.
- Orilio, A. F., Blawid, R., Costa, G. A., Gomes, S. S., Inoue-Nagata, T., et al. (2018) High-throughput sequencing reveals a novel closterovirus in arracacha (*Arracacia xanthorrhiza*). *Archives of virology*, 163(9), 2547-2550.
- Santa Cruz, M. (1994). Aislamiento y caracterización parcial de un potyvirus de arracacha (*Arracacia xanthorrhiza* Bancroft). Tesis para optar el título de Licenciado en Biología, Microbiología y Parasitología. Univ. Nac. Pedro Ruiz Gallo. Lambayeque, Perú. 109 p.
- Silvestre, R., Fuentes, S., Risco, R., Berrocal, A., Adams, I., et al. (2020). Characterization of distinct strains of an aphid-transmitted ilarvirus (Fam. Bromoviridae) infecting different hosts from South America. *Virus research*, 282, 197-944.
- Soltani, N., Stevens, K. A., Klaassen, V., Hwang, M. S., Golino, D. A., & Al Rwahnih, M. (2021). Quality Assessment and Validation of High-Throughput Sequencing for Grapevine Virus Diagnostics. *Viruses*, 13(6), 1130.
- Wu, Q., Ding, S. W., Zhang, Y., & Zhu, S. (2015). Identification of viruses and viroids by next-generation sequencing and homology-dependent and homology-independent algorithms. *Annual review of phytopathology*, 53, 425-444.
- Zerbino, D., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, 821-829.