Short Communication

The complete genome sequences of a Peruvian and a Colombian isolate of Andean potato latent virus and partial sequences of further isolates suggest the existence of two distinct potato-infecting tymovirus species[1]

Jan Kreuze[a], Renate Koenig[b], Joao De Souza[a], Heinrich Josef Vetten[b], Giovanna Muller[a], Betty Flores[a], Heiko Ziebell[b], Wilmer Cuellar[a,c]

[a]International Potato Center, Apartado 1558, Lima 12, Peru
[b]Julius Kühn Institute, Institute for Epidemiology and Pathogen Diagnostics, Messeweg 11-12, D38104 Braunschweig, Germany
[c] Current address: International Center for Tropical Agriculture (CIAT), Apartado Aéreo 6713, Cali, Colombia

A B S T R A C T

The complete genomic RNA sequences of the tymovirus isolates Hu and Col from potato which originally had been considered to be strains of the same virus species, i.e. *Andean potato latent virus* (APLV), were determined by siRNA sequencing and assembly, and found to share only c. 65% nt sequence identity. This result together with those of serological tests and comparisons of the coat protein gene sequences of additional tymovirus isolates from potato suggest that the species *Andean potato latent virus* should be subdivided into two species, i.e. APLV and *Andean potato mild mosaic virus* (APMMV*).* Primers were designed for the broad specificity detection of both viruses.

Key words: Andean potato latent virus, Andean potato mild mosaic virus, broad specificity PCR primers for detecting potato-infecting tymoviruses, quarantine viruses in potatoes, siRNA sequencing

Andean potato latent virus (APLV), family *Tymoviridae*, genus *Tymovirus* was first described by Gibbs *et al.* (1966). The virus is adapted to multiply under the cool conditions of the tropical Andes where potatoes are grown at altitudes between 2000 and 4000 m above sea level (Jones and Fribourg, 1978). It has been pointed out by these authors that the name of the virus is misleading, because, especially in secondary infections, it often does cause symptoms in potato plants. The intensity of the symptoms may vary considerably depending on cultivars and growing conditions and the presence of other viruses. More than 20 different isolates originating from Colombia, Ecuador, Peru and Bolivia have been described which vary considerably in their symptomatological, serological and electrophoretic properties (Fribourg et al., 1977; Koenig et al., 1979). In serological tests pronounced differences were detected between isolates Hu and Ay from Peru and Bo 1 to 12 from Bolivia on the one side, and isolates Caj, Col and additional isolates from Bolivia, Ecuador and Colombia on the other (Koenig et al., 1979). Within this second group, especially the two isolates Col-2 and Col-3 from the Central Colombian Potato Collection (CCC) formed strong spurs when tested side by side with other isolates in serological agar gel double diffusion tests. Thus, there seemed to be three major strain groups which were designated as Hu, Col-Caj and CCC group. In double antibody sandwich

(DAS) ELISA, viruses of each of these groups reacted only weakly or not at all with antisera prepared to viruses of the other groups. The viruses in the Hu group were found to be serologically more closely related to eggplant mosaic virus (EMV) than those of the other two groups (Koenig et al., 1979). - Similar DAS ELISA results were recently obtained at the International Potato Center (CIP) in Lima/Peru. A newly prepared antiserum to the Hu isolate detected only the Hu, but not the Col and Caj isolates whereas the antisera 0002 (against Col) and 0003 (against Col-2) which were obtained from the *Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH* (DSMZ) in 2009 detected isolates Col, Caj and Col2, but not Hu.

In order to obtain more precise information on the relationships between the isolates Col and Hu, the complete genome sequences of these two isolates were determined at CIP through siRNA sequencing and assembly as described by Kreuze et al. (2009). The sequences of the coat protein (CP) genes of further isolates previously described by Koenig et al. (1979) were determined at the Julius Kühn Institute in Braunschweig/Germany from PCR products. The names, references and GenBank accession numbers of all potato-infecting tymovirus isolates studied in this paper are summarized in supplementary Table S1. For the work in Lima, total RNA was extracted from infected *Nicotiana bigelovii x N. clevelandii* plants using Trizol reagent and bulked together in equal proportions with 16 other samples in the case of the Hu isolate, or with two other samples in the case of the Col isolate. After bulking samples were processed as described previously (Fuentes et al., 2012) and sent for siRNA sequencing using the Illumina Hiseq2000 platform (Provider: Fasteris Life Sciences SA, Plan-les-Ouates, Switzerland). Sequences were assembled using Velvet's AssemblyAssembler (v1.4) script with a range of hash lengths from 13 to 25. Contigs were identified by BLAST against GenBank sequences and those corresponding to tymoviruses were extracted. Any gaps in the de-novo siRNA assemblies were separately sequenced using flanking primers, amplifying the corresponding regions and sequencing by the Sanger method at Macrogen (Korea). The 3' and 5' ends were confirmed by RNA linker mediated rapid amplification of cDNA ends (RLM-RACE; Schaefer, 1995) using the modBan (5'- A5'pp CTG TAG GCA CCA TCA AT/ddC/-3') and BanTwo (5'-ATC GTrA rGrGrC rArCrC rUrGrA rArA-3') linkers respectively. Finally assembled complete sequences were reconfirmed by alignment of siRNAs using Novoalign (v 2.07.09; Novocraft technologies, Malaysia) and MAQ (http://maq.sourceforge.net/index.shtml), and visualized using a custom script (colfreq.sh; available from authors upon request). Sequences were edited and annotated using CLC main workbench (v6.6.2; CLCBio) software packages. Phylogenetic and molecular evolutionary analyses were conducted using MEGA5 (Tamura et al., 2011).

From the bulk sample of 17 plants including one infected with the Hu isolate a total of 11,286,650 reads between 21-24 nts was obtained. A single contig of 6204nt with striking similarity to tymovirus sequences could be assembled from the siRNA library reads using Velvet's AssemblyAssembler. It showed 99 % identity to the partial 3' 1705 nt sequence of an unnamed APLV isolate with the GenBank accession number AF035402. The single contig appeared to span the complete genome of the Hu isolate. Re-alignment of the siRNA library to the obtained single contig using Novoalign confirmed the sequence and was able to call all nts with high quality scores at an average coverage of 1724x from 484,065 aligned reads (4.28% of the total siRNA reads). Coverage varied significantly over the genome ranging from 1 (for the extreme 3' nucleotide) to 12976 (at position 61; see supplementary Fig. S1). To verify the quality of the assembly, the CP gene was amplified using the primers Hu-CP-F and Hu-CP-R (for the composition of these and other primers see supplementary

Table S2) and the 3' and 5' ends were amplified by RLM-RACE using primers Hu-3'-F1 and BanOne and Hu-5'-1R, respectively. All fragments were then sequenced by the Sanger method and showed 100% identity with the sequence obtained by the siRNA method (Fig. S1). MAQ identified 14 putative single nt polymorphisms (SNPs) in the Hu sequence, i.e. ten C/T SNPs at positions 69, 2810, 3281, 5336, 5477, 5608, 5609, 5782, 5812 and 5860, two T/G SNP at positions 5 and 5599, and one A/T and A/C SNP each at position 1 and 3614, respectively.

From the bulk of three samples, one of which corresponded to siRNA from a plant infected with the Col isolate, a total of 1,305,287 reads between 21 and 24 nts were obtained. Eleven contigs ranging from 1502 to 82 nts in length with similarity to tymovirus RNAs were identified. Upon further analysis most contigs revealed overlaps with others in 7 - 12 nts enabling us to create two super-contigs of 5643 and 686 nts corresponding to the 5' and 3' terminal parts of the viral genome. A PCR product obtained with the primers Col-F2-CP and Col-R2-CP (Table S2) enabled us to close a gap of 7 nts between the two super-contigs. Another region in the RdRp gene with a limited overlap (7 nts) as well as the 3' end were also amplified using primers Col-F1-RdRp and Col-R1-RdRp and Hu-3'-F1 and BanOne (Table S2), respectively; their sequences confirmed the assembled super-contig sequence and the 3' RNA end (Fig. S1). Re-aligning siRNAs to the consensus sequence of the Col isolate using Novoalign confirmed the sequence calling all nts at an average coverage of 113x by 33,212 reads (2.54% of the total reads). Similar to the results from the Hu isolate, coverage varied significantly over the genome ranging from 1 (for the extreme 3' nucleotide and the first two 5' nucleotides) to 1607 (at position 503; see Fig. S1). MAQ identified six putative SNPs in the genome of APLV-Col: four T/A SNPs at positions 60, 1163, 2453 and 6336, and two C/A SNPs at positions 1398 and 1592, respectively.

The Hu as well as the Col isolates show the typical tymovirus genome organization (Fig. 1). The ORF encoding the movement protein starts seven nts upstream of the second ORF which encodes the replicase. This latter ORF contains close to its 3' end the highly conserved so called 'tymobox' (Ding et al., 1990), a 16 nt sequence which is probably part of the subgenomic promoter for the third ORF which encodes the CP (Dreher et al., 2012). With isolate Col the CP gene is in the same translational frame as the replicase gene, whereas with isolate Hu it is in the same frame as the movement protein gene (Fig. 1). For both RNAs, tRNA-like secondary structures with valine anticodons are predicted at the 3' ends (data not shown). There are pronounced differences, however, between the nt sequences of the two RNAs and also between the derived amino acid (aa) sequences of their three ORFs. The entire genomes share only 65 % nt sequence identity and the CPs only 57% aa identity (Fig. 1). Both values are far below the tymovirus species demarcation thresholds of 80% and 90%, respectively (Dreher et al., 2012). This strongly suggests that the Col and the Hu isolates represent two distinct virus species. Phylogenetic analysis confirmed the results of sequence similarity studies showing that the Hu and the Col isolates are clearly distinct viruses with the Hu isolate being more closely related to EMV than the Col isolate (Fig. 2). This is consistent with the results of previous serological tests (Koenig et al., 1979). Since APLV-Col was the first isolate of APLV described (Gibbs et al., 1966), we suggest the name APLV should be maintained for this virus and that the Hu isolate should be regarded as strain of a new virus species. Although the studies by Fribourg et al. (1977) and Jones and Fribourg (1978) show that all studied isolates (Caj, Ay and Hu) could induce severe symptoms in some wild potato species, the symptomatology in the tested cultivated potato varieties was usually symptomless systemic infection or mild mosaic. This is consistent with results from routine virus indexing performed on potato germplasm under greenhouse conditions at CIP, where

plants positive to antisera raised against the Hu isolate are nearly always symptomless, or in rare cases show mild mosaic, when in single infection. Therefore we propose the name Andean potato mild mosaic virus (APMMV) be adopted for the species represented by the Hu isolate.

Sequence analyses which were done independently and simultaneously in Germany at the Julius Kühn Institute supported the conclusion that the Hu isolate and also the closely related Ay isolate (Koenig et al., 1979) should no longer be regarded as strains of APLV, but rather as strains of the proposed new species APMMV. To compare the molecular properties of 11 of the 23 isolates studied previously (Fribourg et al., 1977; Koenig et al., 1979) RNA was extracted from purified preparations of these isolates which had been obtained 30 - 35 years ago and kept frozen since. In an alignment of all tymovirus sequences, which were available at the time when the molecular studies in Germany were started, genome areas were identified which showed a high degree of conservation. Such areas were used to design primers on the basis of the EMV sequence (GenBank accession No. NC_001480) which in previous serological tests had been found to be most closely related to APLV (Koenig et al., 2005). Primer AP5 was designed on the basis of the unnamed incomplete APLV sequence with the Genbank accession No. AF03502. The composition of the primers EM13, AP5, EM1, EM2, EM14, EM10 and EM3, which were found to be the most useful ones in our studies, is shown in Table S2, and their location on the EMV sequence in Fig. 3a. PCR products were sequenced either directly or after cloning into the pGEM-T vector (Promega).

The nt sequences of the CP genes and adjacent upstream and downstream sequences were found to be distinct for each of the eleven isolates studied. The sequences obtained for the isolates Col and Hu in Braunschweig were almost 100% identical to those obtained by siRNA sequencing in Lima. The average percentages of nt sequence identities for the CP genes and of amino acid (aa) sequence identities for the CPs are shown in Fig. 3b. The sequence differences between those isolates which belong to the Col/Caj (Col, Caj, Col-4, Col-5, Bo-14, Bo-15 and Ec-1) and the CCC (Col-2, Col-3) strain groups (Koenig et al., 1979) are higher on the nt than on the aa level. The average percentages of CP aa sequence identities for these isolates range from 89% to 100% and are thus mostly above or only very slightly below the species demarcation threshold of 90% suggested for tymovirus CP aa sequences (Dreher et al., 2012). The CP gene nt and CP aa sequences of the Ay isolate and also of the unnamed APLV isolate with the GenBank No. AF035402, however, differ greatly from those of the 'true' APLV isolates. They are closely related to those of the Hu isolate (Fig. 3b). The cluster formed by the CPs of these three isolates shares an average of aa sequence identity of only 59% with the cluster formed by the 'true' APLV isolates and of 68% with EMV thus confirming the notion that the Hu isolate and the related Ay and AF035402 isolates should be regarded as strains of APMMV, a distinct tymovirus species which is more closely related to EMV than to the various APLV strains (Fig. 3b).

The so far available APLV and APMMV sequences suggest additional differences between the two virus species. The nt sequences of the CP genes of the APMMV isolates and of EMV end with a UAA stop codon, whereas those of the typical APLV isolates are terminated by a UAG stop codon (result not shown). In the sequences of the true APLV isolates the putative stop codons of the replicase genes and the start codons of the CP genes are separated by three nucleotides (nts), whereas in the sequences of the APMMV isolates they are separated by only two nts (Fig. 3c). This produces the frame shift of the CP genes with respect to the replicase genes shown in Fig. 1. In addition, the

sequences of the APMMV isolates lack two codons upstream of the replicase stop codons which are present in the true APLV isolates (Fig. 3c).

APLV is a regulated pest in the European Union (Council Directive 2000/29/EC, Annex I A I, section I d-2). Quarantine testing requires reliable detection methods. With the primer combinations EM10/AP5 and EM14/EM13 we obtained single bands for PCR products of the expected sizes with all our isolates. The primer combinations EM3/AP5, EM3/EM1 or EM3/EM2, which all encompass the CP gene region, sometimes produced additional bands that might complicate the interpretation of the results in routine diagnosis. The region encompassed by primers EM13 and EM14 would seem to be especially suitable for the differentiation of isolates, because in this region there are many differences between the Col and the Hu sequences and between tymovirus sequences in general. Several of our isolates, however, appeared to contain mixtures of sequences in this region which had not been evident in the CP gene regions. After cloning and sequencing the EM13/EM14 PCR products, several different sequences were detected from some of our isolates and in one case a recombination of two different sequence portions was clearly evident (result not shown). It thus seems that an exchange of 5' regions varying in lengths may have occurred with some of the isolates. Further sequencing studies, possibly with isolates in which different components have been separated by local lesion passages will, therefore, be necessary to obtain more detailed information on various sources of APLV and APMMV and also to enable studies on their pathogenicities.

It should also be noted that further variation than reported here may exist among tymoviruses infecting potatoes and related crops. Thus, an isolate U from ulluco (*Ullucus tuberosus*) (Lizarraga et al., 1996) not included in this study only gave a weak ELISA reaction (as compared to the other viruses) with one of the three antisera tested. This suggests that this isolate, although more closely related to the Col-Caj and CCC groups, may represent a more distantly related strain group. Genome sequence analysis of such isolates will elucidate their exact taxonomic status.

Acknowledgements

We gratefully acknowledge the excellent technical assistance of Hugo Espinoza and Petra Lüddecke-(formerly Rähse), the gift of virus isolates by Cesar Fribourg and 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).

References

Ding, S.W., Howe, J., Keese, P., Mackenzie, A., Meek, D., Osorio-Keese, M., Skotnicki, M., Srifah, P., Torronen, M., Gibbs, A., 1990. The tymobox, a sequence shared by most tymoviruses: its use in molecular studies of tymoviruses. Nucl. Acids Res. 18, 1181–1187.

Dreher, T.W., Edwards, M.C., Gibbs, A.J., Haenni, A.-L., Hammond, R.W., Jupin, I., Koenig, R., Sabanadzovic, S., Martelli, G.P., 2012. Tymovirus. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds), Virus Taxonomy, Ninth Report of the ICTV. Elsevier/Academic Press, London, pp. 946-952.

Fribourg, C.E., Jones, R.A.C., Koenig, R., 1977. Host plant reactions, physical properties and serology of three isolates of Andean potato latent virus from Peru. Ann. Appl. Biol. 86, 373-380.

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. Arch. Virol. 157,773-776. *DOI 101007/s00705-011-1214-6*

Gibbs, A.J., Hecht-Poinar, E., McKee, R.K., 1966. Some properties of three related viruses: Andean potato latent, dulcamara mottle and ononis yellow mosaic. J. Gen. Microbiol. 44, 177-193.

Jones, R.A.C., Fribourg, C.E., 1978. Symptoms induced by Andean potato latent virus in wild and cultivated potatoes. Potato Res. 21, 121-127.

Koenig, R., Fribourg, C.E., Jones, R.A.C., 1979. Symptomatological, serological and electrophoretic diversity of isolates of Andean potato latent virus from different regions of the Andes. Phytopathology 69, 748-752.

Koenig, R., Pleij, C.W.A., Lesemann, D.E., Loss, S., Vetten, H.J., 2005. Molecular characterization of isolates of anagyris vein yellowing virus, plantago mottle virus and scrophularia mottle virus — comparison of various approaches for tymovirus classification. Arch. Virol. 150, 2325–2338.

Kreuze, J.F., Perez, A., Untiveros, M., Quispe, D., Fuentes, S., Barker, I., Simon, R., 2009. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. Virology 388, 1-7, DOI 101016/jvirol200903024

Lizarraga, C., Santa Cruz, M., Jayasinghe, U., 1996. Detection of an isolate of Andean potato latent virus in ulluco (*Ullucus tuberosus* Caldas). Plant Dis 80, 344.

Saitou, N., Nei, M., 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425.

Schaefer, B.C., 1995. Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-Length cDNA ends. Anal. Biochem. 227, 255-273.

Sneath, P.H., Sokal, R., 1973. Numerical Taxonomy: the principles and practice of numerical classification. Freeman, San Francisco.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. Mol. Biol. Evol. 28, 2731-2739.

242 Legends to Figures:

243 **Fig. 1.** Arrangment of the open reading frames in the genomes of the typical APLV isolate Col, the
244 deviating isolate Hu (APMMV) and of EMV. The percentages of nt sequence identities of the enire
245 RNAs and of the aa sequence identities of the movement (MP), replicase (ReP) and coat proteins (CP)
246 are shown.

247 **Fig. 2.** Phylogenetic tree generated from aligned complete nt sequences using the neighbour-joining
248 and Kimura 2 parameter models. Trees with similar topology were generated when using maximum
249 likelihood or minimum evolution algorithms (Tamura et al., 2011). Bootstrap values obtained
250 following 1000 bootstrap replications are given at each node when they exceeded 50%; the highest
251 and lowest values are given and when they differed between the three methods used.

252 **Fig. 3.** Comparison of some molecular properties of eleven potato-infecting tymoviruses (Table S1;
253 Koenig et al., 1997), of EMV and of an incompletely sequenced APLV isolate (GenBank accession No
254 AF035402). **a** Locations on the EMV nt sequence of sense and antisense primers which were found
255 to be especially useful for the broad specificity detection of APLV isolates as well as APMMV isolates.
256 Primer sequences are listed in Table S2. **b** Average percentages of nt and aa sequence identities for
257 the CP genes and the CPs, respectively. The trees were generated by the DNAMAN software using
258 the UPGMA method (Sneath and Sokal, 1973). Phylogenetic trees almost identical in topology were
259 obtained by the neighbor joining method (Saitou and Nei, 1987). **c** Sequence alignment of the area
260 downstream of the tymobox and upstream of the start codon of the CP genes in the various isolates;
261 nucleotides differing from those in the corresponding positions in the majority of the other viruses
262 are highlighted by white letters on a black background. - The names of the atypical isolates which
263 represent the proposed new species APMMV are underlined.

264

265

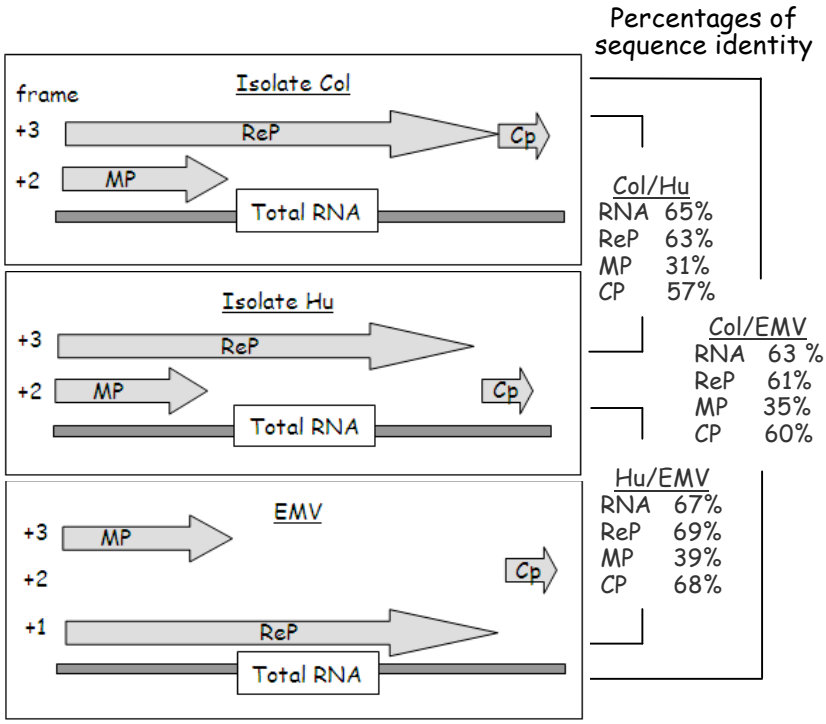
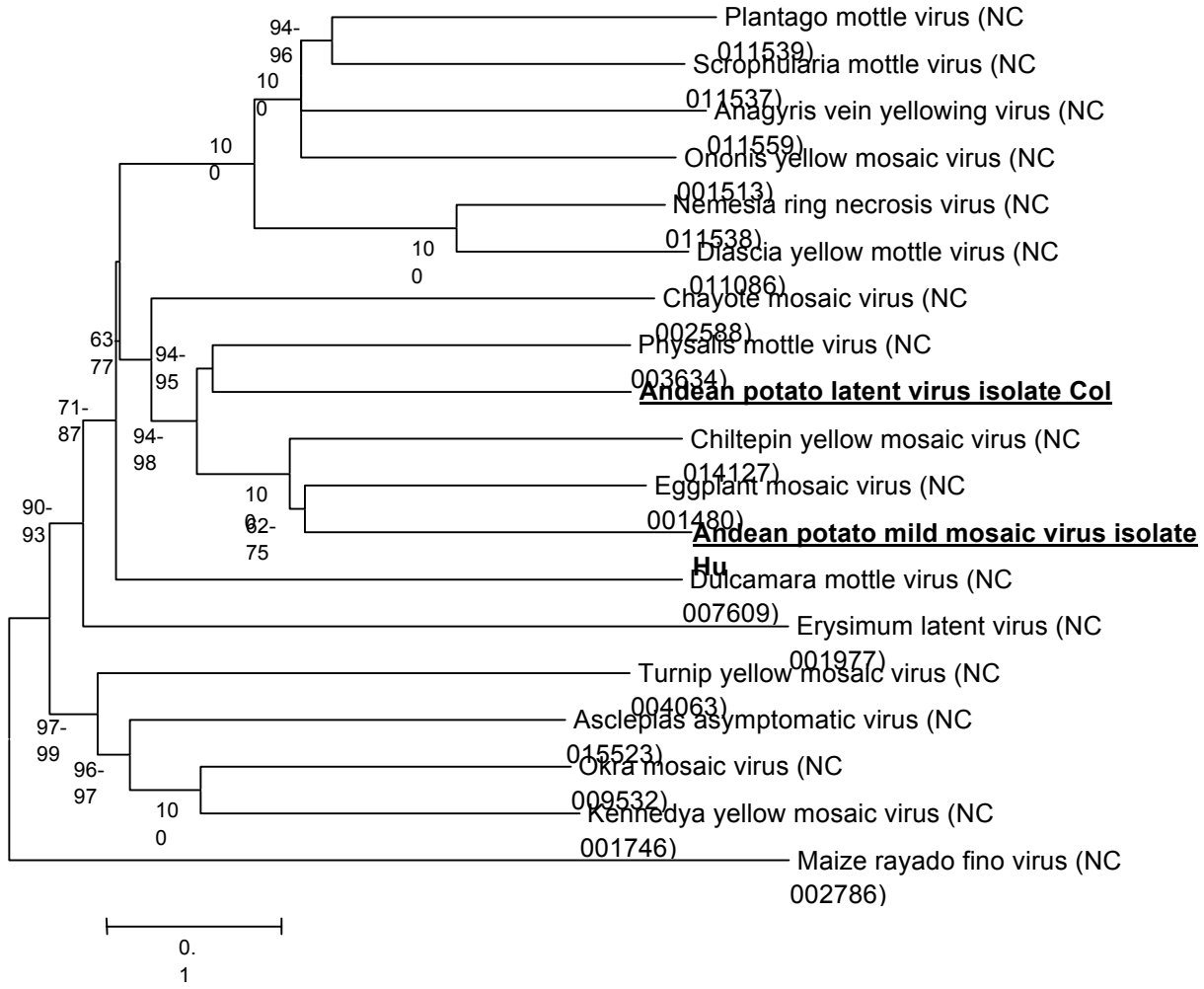

Percentages of sequence identity
frame
Isolate Col
+3
ReP
Cp
+2
MP
Total RNA
Col/Hu
RNA 65%
ReP 63%
MP 31%
CP 57%
Isolate Hu
+3
ReP
+2
MP
Cp
Total RNA
Col/EMV
RNA 63 %
ReP 61%
MP 35%
CP 60%
EMV
+3
MP
+2
Cp
+1
ReP
Total RNA
Hu/EMV
RNA 67%
ReP 69%
MP 39%
CP 68%

Fig. 1.

281

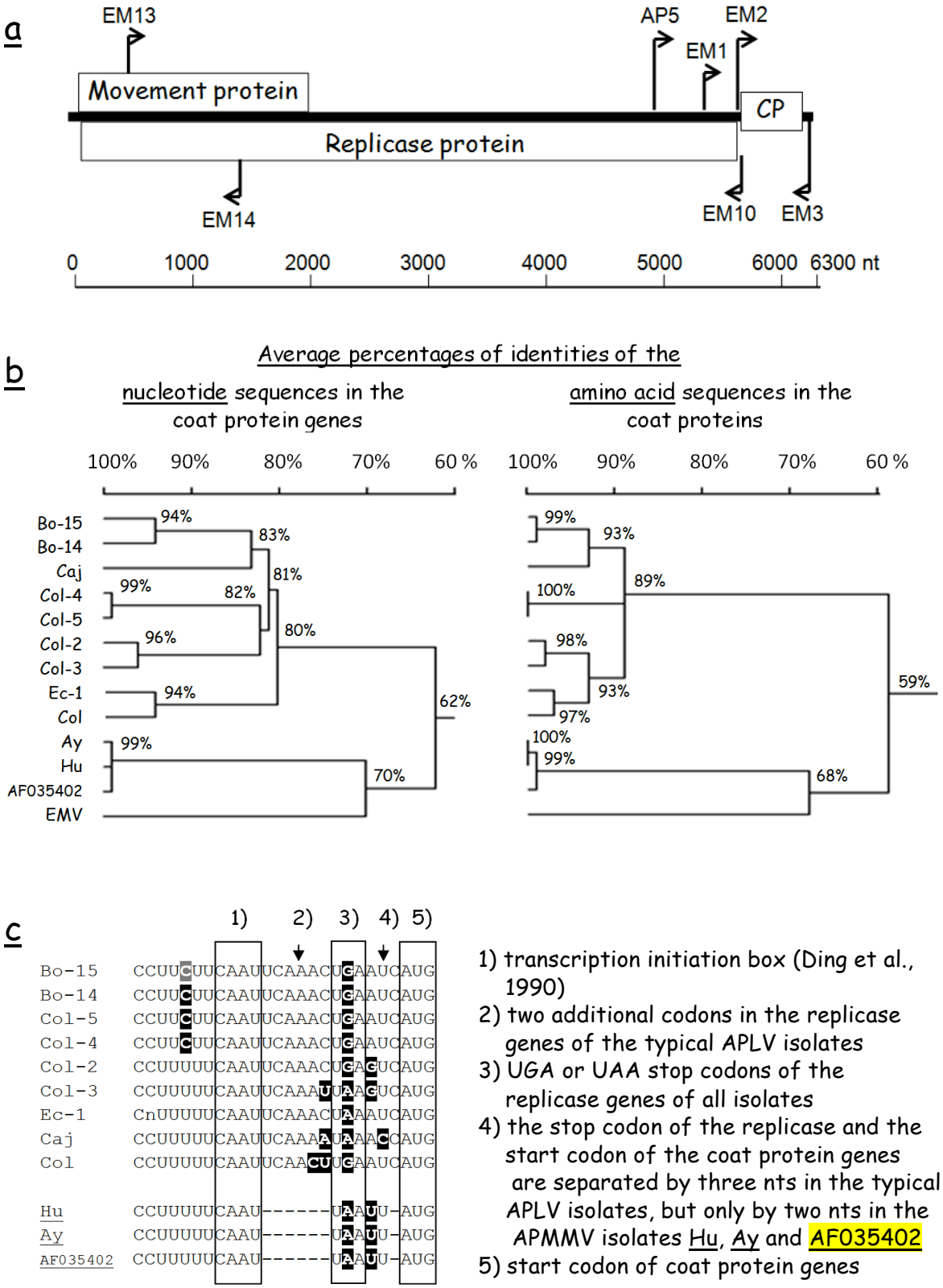
282



283

284 **Fig. 2.**

285



1) transcription initiation box (Ding et al., 1990)
2) two additional codons in the replicase genes of the typical APLV isolates
3) UGA or UAA stop codons of the replicase genes of all isolates
4) the stop codon of the replicase and the start codon of the coat protein genes are separated by three nts in the typical APLV isolates, but only by two nts in the APMMV isolates Hu, Ay and AF035402
5) start codon of coat protein genes

Fig. 3.