Sequence analysis of Potato leafroll virus isolates reveals genetic stability, major evolutionary events and differential selection pressure between overlapping reading frame products

Sébastien Guyader and Danièle Giblot Ducray

INRA - UMR BiO3P, Domaine de la Motte, BP 35327, 35653 Le Rheu Cedex, France

In order to investigate the genetic diversity of Potato leafroll virus (PLRV), seven new complete genomic sequences of isolates collected worldwide were compared with the five sequences available in GenBank. Then, a restricted polymorphic region of the genome was chosen to further analyse new sequences. The sequences of PLRV open reading frames (ORFs) 3 and 4 were also compared with those of two other poleroviruses and the non-synonymous to synonymous substitution ratio distribution was analysed in overlapping and non-overlapping regions of the genome using maximum-likelihood models. Results confirmed that PLRV sequences from around the world are very closely related and showed that the region encoding protein P0 allowed the detection of three groups of isolates. When compared to other poleroviruses, PLRV was the most conserved in both ORFs 3 and 4. However, the results suggest that important events, such as deletion, mutation at a stop codon and intraspecific homologous recombination events, have occurred during the evolution of PLRV. Finally, it was shown that the translation products of ORFs 0 and 3 are significantly more conserved than those of the overlapping ORFs 1 and 4, respectively. All together, the results allow the proposal of new hypotheses to explain the apparent genetic stability of PLRV and its evolution.

Introduction

Potato leafroll virus (PLRV), the type species of the genus Polerovirus, family Luteoviridae, is an RNA virus that infects potato crops worldwide causing important damage (Robert & Lemaire, 1999). It is transmitted in a persistent manner by a limited number of aphid species, Myzus persicae Sulz. being the most efficient and important vector (Harrison, 1984). Its natural host range is mainly restricted to a few solanaceous plants, including Physalis floridana (Rydb.), Datura stramonium L. and Lycopersicon esculentum (Mill.), in addition to the potato (Solanum tuberosum L.) (Thomas, 1993). Moreover, a few plants of other families can be infected by PLRV, either experimentally (Harrison, 1984) or naturally (Lizarraga et al., 1996). The genome of PLRV consists of a positive-sense single-stranded RNA harbouring six main open reading frames (ORFs) numbered from 0 to 5 (reviewed by Miller et al., 1997).

The proteins translated from these ORFs are referred to as P0 to P5. There is strong evidence to suggest that P0 is essential for replication of the PLRV genome (Sadowy et al., 2001b) and that P1 is a proteinase-containing polyprotein whose self-cleavage releases VPg (Prüfer et al., 1999; Sadowy et al., 2001a; van der Wilk et al., 1997). P2 is translated by a rarely occurring ribosomal frameshift from ORF1 and carries the conserved motifs typical of RNA-dependent RNA polymerases (RdRp). P3, P4 and P5 are translated from a subgenomic RNA and, respectively, correspond to the major capsid protein (CP), the putative movement protein (MP) and the readthrough domain (RTD), which is translated by suppression of the ORF3 stop codon. Ashoub et al. (1998) identified two further putative ORFs encoding two proteins, P6 and P7, the roles of which remain unknown.

PLRV exhibits some variability with regard to biological properties. Firstly, different strains can be characterized by the symptoms, which range from mild to severe (Harrison, 1984), caused in P. floridana. Secondly, some PLRV strains differ in their ability to infect some plant species. For example, the
PLRV-TYT strain causes the tomato yellow top disease, whereas tomato plants infected by typical potato strains are almost symptomless (Thomas, 1984). PLRV-TYT, on the other hand, causes no or only weak symptoms on potato plants. Finally, transmissibility by aphid vectors can range from low to high levels, according to both the strain and the aphid clone used to transmit it (Bourdin et al., 1998).

Despite such diversity in biological properties, some studies suggest that PLRV is not highly variable at the sequence level. Comparison of nucleotide sequences of PLRV isolates from Canada (PLRV-C), Scotland (PLRV-S), the Netherlands (PLRV-N) and Australia (PLRV-A) have shown that PLRV-C, -S and -N share about 98% of the residues and that PLRV-A, the most divergent isolate, shows an overall identity of 93% with other isolates (Keese et al., 1990). PLRV-A is particularly divergent from other isolates by about 600 nt in the polymerase-coding region. Another isolate from Poland (PLRV-P) is also more closely related to the three former isolates than to the latter (Palucha et al., 1994). Similarly, sequencing of ORF3 of Cuban and Brazilian isolates suggests that these isolates are more closely related to the Northern temperate country isolates than to the Australian isolate (López et al., 1994; Caram de Souza-Dias et al., 1999). Finally, comparison of ORF3 and ORF5 sequences between Scottish isolates has shown that these isolates are as divergent from each other as they are from overseas isolates, except for the one from Australia (Jolly & Mayo, 1994).

Viral RdRps are known to be error-prone because they lack proofreading functions. As a result, they introduce mutations at a rate close to one substitution per genome per replication cycle (Drake & Holland, 1999). Other mechanisms, such as recombinations and reassortments, in addition to redistributing neutral mutations among more or less closely related genomes, allow both the elimination of deleterious mutations and the creation or spread of beneficial combinations of changes (Worobey & Holmes, 1999). The role of these mechanisms in the evolution of RNA viruses is reinforced by the short replication time of the viral genome and account for the great diversity and adaptability of RNA viruses. However, some RNA viruses show relatively long-term evolutionary stasis, although they are capable of rapid evolution in particular circumstances (Holland et al., 1992). Such cases are documented for animal viruses, such as Vesicular stomatitis virus (VSV) (Nichol et al., 1993; Novella et al., 1999; Rodriguez et al., 1996) and Venezuelan equine encephalitis virus (Powers et al., 1997), and plant viruses, such as Tobacco mild green mosaic virus (Fraile et al., 1997; Moya et al., 1993) and Citrus tristeza virus (Albiach-Martí et al., 2000).

The apparent low genetic diversity of PLRV may be an indication of evolutionary stasis. However, such stasis does not seem to be a characteristic of other virus members of the family Luteoviridae. Thus, other species like Barley yellow dwarf virus-PAV (BYDV-PAV) (Bencharki et al., 1999; Mastari & Lapierre, 1999) and those belonging to the beet polerovirus complex (Hauser et al., 2000) are known to be more variable. However, as only five complete PLRV sequences are available in the international databases, there is insufficient information to draw firm conclusions. In this paper, we investigate the genetic diversity of PLRV by obtaining and analysing new complete and partial sequences. Our results confirm that the level of sequence variation among PLRV isolates from around the world is low. However, we show that important events, such as deletions and intraspecific recombinations, have occurred during the evolutionary history of PLRV. We also find that differential selection applies to overlapping reading frame products. New hypotheses are proposed to explain the apparent genetic stability of PLRV and its evolutionary pathway.

**Methods**

- **Virus isolates.** A total of 12 PLRV isolates originating from different countries and maintained on *P. floridanus* plants and propagated by cuttings was taken from our collection (Table 1). Eight isolates, collected from potato (*Solanum tuberosum* ssp. *tuberosum* and *andigena*) or *P. floridanus* were also provided by donors from several countries (Table 1).

- **PCR and sequencing.** Total RNAs were extracted from infected leaf samples with the RNeasy Plant Mini kit (Qiagen), according to the manufacturer’s protocol. Reverse transcription was performed using AMV reverse transcriptase and PCR was performed using Taq DNA polymerase (primer sequences are presented as supplementary data on JGV Online (see http://vir.sgmjournals.org), which allowed the amplification of overlapping fragments encompassing the entire genome of PLRV. Sequencing reactions (ABI Prism Big Dye D RHodamine Terminator Cycle Sequencing Ready Reaction kit) (Applied Biosystems) were carried out directly on purified PCR products (Concert Rapid PCR Purification system) (Gibco BRL) from at least two independent RT–PCRs for each isolate and analysed on an ABI 310 Automated sequencer (Applied Biosystems). Overlapping sequences were assembled with contig (Huang, 1992) and then aligned with clustal w, version 1.8 (Thompson et al., 1994), both available in the mirror package (Hall, 1999).

- **Phylogenetic analyses.** Unrooted phylogenetic trees based on complete sequence alignments were reconstructed by the maximum-likelihood (ML) approach, as implemented in TREEFUGUE, version 5.0 (Strimmer & von Haeseler, 1996). The Tamura-Nei model of substitution was used (Tamura & Nei, 1993) and both the transition to transversion ratio (ts:tv) and the shape parameter (α) of a Γ distribution of rate variation among sites (with eight categories) were estimated during the tree reconstruction. The distribution of variability along the aligned complete sequences was examined by plotting the entropy against each position of the alignment. The entropy at a position of the alignment is a measure of the complexity (thus of the variability) in nucleotide composition, taking into account both the number and the frequency of different nucleotides observed at this position. It was calculated with the mirror package as H(l) = −Σ(b,l)ln[(b,l)], where f is the frequency at which residue b is found at position l, and plotted with Microsoft Excel. This method was used to help in the choice of a variable region (concentration of close sites with high entropy) for which new partial sequences were obtained and analysed, as described above. Average nucleotide diversity (π) within subpopulations was calculated by the method of Nei and Jin with the program dnasp (written by Naoko Takezaki, http://iubio.bio.indiana.edu/soft/molbio/evolve/). For the
Table 1. Newly sequenced PLRV isolates

Isolates from the local collection (coll) are indicated. Partial sequences (934 nt) encompass complete ORF0 and partial ORF1.

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Geographical origin</th>
<th>Year of isolation</th>
<th>Symptoms on P. floridana</th>
<th>Source</th>
<th>Sequence</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zim13</td>
<td>Zimbabwe</td>
<td>1985</td>
<td>Severe</td>
<td>Coll</td>
<td>Complete</td>
<td>AF453388</td>
</tr>
<tr>
<td>OP</td>
<td>Spain</td>
<td>1986</td>
<td>Mild</td>
<td>Coll</td>
<td>Complete</td>
<td>AF453389</td>
</tr>
<tr>
<td>Noir</td>
<td>France</td>
<td>1996</td>
<td>Severe</td>
<td>Coll</td>
<td>Complete</td>
<td>AF453390</td>
</tr>
<tr>
<td>F1</td>
<td>France</td>
<td>1983</td>
<td>Weak</td>
<td>Coll</td>
<td>Complete</td>
<td>AF453391</td>
</tr>
<tr>
<td>CIP01</td>
<td>Peru</td>
<td>1974</td>
<td>–</td>
<td>L.F. Salazar</td>
<td>Complete</td>
<td>AF453392</td>
</tr>
<tr>
<td>Cu87*</td>
<td>Cuba</td>
<td>1987</td>
<td>Very severe</td>
<td>Coll</td>
<td>Complete</td>
<td>AF453393</td>
</tr>
<tr>
<td>14.2*</td>
<td>France</td>
<td>1985</td>
<td>Mild</td>
<td>Coll</td>
<td>Complete</td>
<td>AF453394</td>
</tr>
<tr>
<td>Au16</td>
<td>Australia</td>
<td>2000</td>
<td>–</td>
<td>J.E. Thomas</td>
<td>Partial</td>
<td>AF453395</td>
</tr>
<tr>
<td>Au40b</td>
<td>Australia</td>
<td>1981</td>
<td>–</td>
<td>J.E. Thomas</td>
<td>Partial</td>
<td>AF453396</td>
</tr>
<tr>
<td>TYTV2</td>
<td>Australia</td>
<td>1987</td>
<td>–</td>
<td>J.E. Thomas</td>
<td>Partial</td>
<td>AF453397</td>
</tr>
<tr>
<td>Au252</td>
<td>Australia</td>
<td>1992</td>
<td>–</td>
<td>J.E. Thomas</td>
<td>Partial</td>
<td>AF453398</td>
</tr>
<tr>
<td>K5</td>
<td>Spain</td>
<td>1986</td>
<td>Mild</td>
<td>Coll</td>
<td>Partial</td>
<td>AF453399</td>
</tr>
<tr>
<td>L18</td>
<td>Poland</td>
<td>1985</td>
<td>–</td>
<td>Coll</td>
<td>Partial</td>
<td>AF453400</td>
</tr>
<tr>
<td>L7</td>
<td>Poland</td>
<td>2000</td>
<td>–</td>
<td>D. Hulanicka</td>
<td>Partial</td>
<td>AF453401</td>
</tr>
<tr>
<td>PLRV-V</td>
<td>Scotland</td>
<td>–</td>
<td>Mild</td>
<td>M.A. Mayo</td>
<td>Partial</td>
<td>AF453402</td>
</tr>
<tr>
<td>L13B</td>
<td>Poland</td>
<td>1985</td>
<td>Mild</td>
<td>Coll</td>
<td>Partial</td>
<td>AF453403</td>
</tr>
<tr>
<td>L13D</td>
<td>Poland</td>
<td>1985</td>
<td>Mild</td>
<td>Coll</td>
<td>Partial</td>
<td>AF453404</td>
</tr>
<tr>
<td>14.1</td>
<td>France</td>
<td>1985</td>
<td>Mild</td>
<td>Coll</td>
<td>Partial</td>
<td>AF453405</td>
</tr>
<tr>
<td>Br1</td>
<td>Brazil</td>
<td>2000</td>
<td>–</td>
<td>J.A. Caram de Souza Dias</td>
<td>Partial</td>
<td>AF453406</td>
</tr>
<tr>
<td>1457</td>
<td>Poland</td>
<td>2000</td>
<td>–</td>
<td>D. Hulanicka</td>
<td>Partial</td>
<td>AF454283</td>
</tr>
</tbody>
</table>

* ORFs 3 (CP) and 5 (RTD) sequences published already under accession numbers AF271214 and AF271215 for isolates 14.2 and Cu87, respectively.

Table 2. ML models used for analysing the ratio $\omega = d_N/d_S$ distribution

<table>
<thead>
<tr>
<th>Model code</th>
<th>Number of parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>1</td>
<td>One ratio across all sites</td>
</tr>
<tr>
<td>M1</td>
<td>1</td>
<td>Two categories: $\omega = 0$ and $\omega = 1$</td>
</tr>
<tr>
<td>M2</td>
<td>3</td>
<td>Three categories: $\omega = 0$, $\omega = 1$ and $\omega = 1$</td>
</tr>
<tr>
<td>M3</td>
<td>5</td>
<td>Three categories: $\omega = 0$, $\omega = 1$ and one free category</td>
</tr>
<tr>
<td>M7</td>
<td>2</td>
<td>Discrete $\beta$ distribution with 10 categories of $0 \leq \omega \leq 1$</td>
</tr>
<tr>
<td>M8</td>
<td>4</td>
<td>Same as M7 but with one more category of $\omega = 1$</td>
</tr>
</tbody>
</table>

A recombination event was suspected to have occurred for at least one isolate when examining, comparison of PLRV sequences with those of two other poleroviruses, we used hidden Markov Model-based alignments of CP and of the putative MP sequences (SAM-T99 server at http://www.cse.ucsc.edu/research/compbio/HMM-apps/T99-tuneup.html) and converted them to nucleotide alignments using the PROTA2DNA server (http://bioweb.pasteur.fr/seqanal/interfaces/protal2dna.html).

Identification of selection pressure. In order to investigate the type of selective pressure affecting the substitution process, the ratios $\omega$ for the rates of non-synonymous substitution ($d_N$) to synonymous substitution ($d_S$) were estimated using codon-based Markov models of substitution implemented in the codeml program of the PAML package, version 3.8d (Yang et al., 2000), which allows heterogeneity of this ratio across codon sites. Six different models (Table 2) were tested and the likelihood-ratio tests (LRTs) were used according to the authors’ recommendations in order to find the model that best fits the data: the value $2\Delta\ln L$ (twice the difference in log-likelihood between a model accounting for positively selected sites and the simpler model without positively selected sites) is compared to a $\chi^2$ distribution with the degrees of freedom equal to the difference in the number of parameters between the two models. The topology of the phylogenetic tree used by the program as input data were reconstructed using the ML search procedure of the program BASEML in PAML, with parameters estimated using TREEPUZZLE. Yang et al. (2000) have shown that this type of analysis is largely insensitive to tree topology.

Validation of a recombination event. A recombination event was suspected to have occurred for at least one isolate when examining,
Comparison of nucleotide diversity within each group showed clustering in the central star-like region of the tree. Remaining isolates, originating from diverse countries and continents, clustered in the central star-like region of the tree. A first group encompassed the three European isolates L13B, L13D and 14.1 with CIP01. A second group contained exclusively Australian isolates, whereas the remaining isolates, originating from diverse countries and continents, clustered in the central star-like region of the tree.

Results

Sequence diversity

The complete sequences of seven isolates were obtained and aligned with the five sequences available in GenBank. Analyses confirmed the low variability of PLRV, as the overall nucleotide identity among these complete sequences is 96.5% (see JGV Online for supplementary data, http://vir.sgmjournals.org). Accordingly, the ML tree reconstructed from these sequences showed a lack of phylogenetic structure (Fig. 1a). As reported already, the Australian isolate was the most divergent, with an average of 94% nucleotide identity with the other isolates, but no correlation appeared between the patterns of variability and the geographical origin of the isolates. When the distribution of variability along the genome was assessed with an entropy plot (Fig. 2), it appeared that the most variable sites concentrated in ORF1 (including overlapping regions with ORFs 0 and 2), the 3’ third of ORF2, the central non-coding region, the 5’ overlapping regions between ORFs 3 and 4, and ORF5. However, most of the variability observed in the non-overlapping parts of ORFs 1 and 2 was attributed to PLRV-A, whose divergence from the other isolates introduced biases into the analysis.

Nucleotide identities for each of the ORFs ranged from 95.9 to 98.6%, whereas amino acid identities ranged from 94.3 to 98% (see JGV Online for supplementary data, http://vir.sgmjournals.org). Phylogenetic trees were also constructed for each ORF (data not shown) and confirmed that the Australian isolate was the most divergent in nearly all ORFs, particularly in ORF1. However, the Peruvian isolate CIP01 was, interestingly, more divergent in ORF0. The variability observed in ORF0 with the entropy plot, together with the divergence of the Peruvian and Australian isolates in this ORF, led us to sequence a region of 934 nt for 13 new isolates, that isolates in group 2 (π = 0.0241) were more diverse than those in either group 1 or 3 (π = 0.0123 and 0.0190, respectively).

Comparison with two other poleroviruses

The beet polerovirus complex (Hauser et al., 2000) and the Cereal yellow dwarf virus-RPV (CYDV-RPV) are two other poleroviruses close to PLRV. Therefore, we compared all of the ORF3 and ORF4 sequences obtained in this study as well as those available in GenBank for PLRV (19 sequences) with those of the beet polerovirus complex (23 sequences) and those of CYDV-RPV (three sequences). Here, we shall consider the beet polerovirus complex, which contains the three recently proposed species Beet mild yellowing virus (BMV), Brassica yellowroads virus (BrYV) and Beet chlorosis virus (BChV) (Hauser et al., 2000) as a single species, as the authors have shown that the minimum amino acid sequence identity among these viruses was 89.8% in CP sequences. Moreover, ORF3 sequences for these three viruses are perfectly aligned with no gaps.
In our analyses, PLRV appeared to be the most conserved virus, in both ORFs 3 and 4 (Table 3) with a mean nucleotide identity of 98·2 and 98·4%, respectively, compared to 94·8 and 94·6% for the beet polerovirus complex and 91·9 and 92·8% for the three CYDV-RPV sequences. The minimum identities were 96·5 and 96·6%, respectively, for ORFs 3 and 4 of PLRV and we found a minimal identity of 91·6% in ORF3 (89·3% aa) for beet poleroviruses, which is comparable to the value obtained by Hauser et al. (2000). Within the beet polerovirus subgroup, designated BrYV, we found a mean nucleotide identity of 96·2%, with a minimum of 92·5%, which is in agreement with our grouping of the beet poleroviruses in a single species.

One striking feature concerning PLRV as well as the two other poleroviruses compared in this study is that, for the overlapping ORFs 3 and 4, nucleotide sequences were more conserved than protein sequences (Table 3). The difference was more pronounced in ORF4. Additionally, similar patterns were observed for PLRV in all the other ORFs (see JGV Online for supplementary data, http://vir.sgmjournals.org). However, when PLRV-coding regions were split into overlapping (31% of the genome length) and non-overlapping ones, we found that nucleotide sequences were more variable than amino acid sequences in non-overlapping regions, as expected due to the degeneracy of the genetic code, but that the opposite was found in overlapping regions where, in both reading frames, nucleotide substitutions were more often non-synonymous.

**Identification of differential selection pressures in overlapping frames**

The patterns of non-synonymous ($d_{\omega}$) and synonymous ($d_{S}$) substitutions and their ratio ($\omega$) were analysed codon by codon (25 sequences) in the two overlapping frames of ORFs 0 and 1. Six ML models were tested for their goodness of fit to the aligned sequences. The model M0 was always significantly less-well fit to the data than the others (data not shown); therefore only the results concerning the other models and the LRTs for their comparison are presented (Table 4). For ORF0, no evidence for positively selected sites was found, as the models built to accommodate such codons with $\omega$ values greater than 1 (M2, M3 and M8) were not significantly better than the simpler models M1 and M7 ($P > 0.05$). In this ORF, about one-third of the codons have an $\omega$ value close to 1 and the remaining sites have an $\omega$ value close to 0. Conversely, for ORF1, the models accounting for codons with an $\omega$ value greater than 1 were significantly better than those that do not ($P < 0.05$). Parameters estimated with model M8 for ORF1 indicate that a large number of codons (14%) have an $\omega$ value greater than 3. The distribution of the $\omega$ values along ORFs 0 and 1 (Fig. 3) showed that the $\omega$ values are equally distributed on both sequences and no region with a particularly high or low value can be identified. These results represent evidence for greater negative selection at the amino acid level for ORF0.

**Table 3. Nucleotide and amino acid identities in ORFs 3 and 4 for PLRV, the beet polerovirus complex and CYDV-RPV**

Nucleotide identities are indicated in roman numerals; amino acid identities are indicated in italic numerals.

<table>
<thead>
<tr>
<th>Virus (no. of sequences compared)</th>
<th>ORF3</th>
<th>ORF4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Minimum</td>
</tr>
<tr>
<td>PLRV (19)</td>
<td>0·9818</td>
<td>0·9650</td>
</tr>
<tr>
<td></td>
<td>0·9770</td>
<td>0·9530</td>
</tr>
<tr>
<td>Beet polerovirus complex (23)*</td>
<td>0·9479</td>
<td>0·9160</td>
</tr>
<tr>
<td></td>
<td>0·9304</td>
<td>0·8930</td>
</tr>
<tr>
<td>CYDV-RPV (3)</td>
<td>0·9283</td>
<td>0·9190</td>
</tr>
<tr>
<td></td>
<td>0·9243</td>
<td>0·9100</td>
</tr>
</tbody>
</table>

* Beet polerovirus complex, including the species BMYV, BChV and BrYV, as proposed in Hauser et al. (2000).
Table 4. Comparison of the different ML models used in ORF0 and partial ORF1 analysis for positive selection

<table>
<thead>
<tr>
<th>ORF</th>
<th>∆lnL for model*</th>
<th>Parameter estimates for positively selected sites†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M2 versus M1</td>
<td>M3 versus M1</td>
</tr>
<tr>
<td>0</td>
<td>2.79</td>
<td>4.07</td>
</tr>
<tr>
<td>1</td>
<td>14.88ᵃ</td>
<td>21.20ᵇ</td>
</tr>
</tbody>
</table>

* The statistics ∆lnL (twice the difference in log-likelihood between the two models tested) is compared with a χ² table according to Yang et al. (2000). If the value ∆lnL is significantly greater than the tabulated χ², the model accounting for positively selected sites (ω > 1) is the best fit to the data, indicating the action of positive selection. a, P < 0.001; b, P < 0.05.
† The proportion of sites under positive selection (pₛ) and their ω values are given for model M8.

Evidence for major evolutionary events

Examination of the alignment of the complete sequences revealed that several major events have occurred during the evolutionary history of PLRV. Firstly, the comparison of the sequence of the isolate Noir with those of the other isolates revealed a 27 nt long deletion in the overlapping region of ORFs 1 and 2, extending from nt 1709 to 1735 (numbering based on the sequence of the isolate PLRV-S) in the region where Mayo et al. (1989) found three tandem repeats of a nearly conserved 27 nt motif. As this deletion is located upstream of the pseudo-knot involved in the frameshifting that leads to the expression of ORF2 (Kim et al., 2000), this corresponds to a deletion of 9 aa in the P1 protein.

Secondly, we found a substitution at the stop codon of ORF0 of the isolate 14.2: UGA → CGA (Arg). The abolition of this stop codon would allow the translation of 22 extra amino acids at the C terminus of P0 before the next in-frame stop codon, i.e. an extension of 9% in the size of this protein.

Finally, we noticed that, whereas group 2 contains only Australian isolates, one isolate from Australia, Au16, clustered with group 3 during ORF0 phylogenetic tree reconstruction. By checking the sequence alignment, this isolate appeared to share some molecular signatures typical of the Australian isolates in the 5' part of the ORF0 sequence but to be more closely related to group 3 isolates in the 3' part. As this observation could be the result of a recombination or gene conversion event between a group 2 (Australian) and a group 3 isolate, this eventuality was tested by two methods. The siscan profile (Fig. 4) comparing the putative recombinant
isolate Au16 with two putative parents from group 2 and group 3 (Au40b and L18, respectively) in the region of 881 nt starting from ORF0. This clearly showed that Au16 is significantly closer to the first parent from about nt 300 to 640 and to the second from about nt 100 to 250 and from nt 660 to the end of the sequence. The recombinant status of Au16 was also supported by the maximum $\chi^2$ test with isolates Au40b and L18 taken as parents, locating a putative recombination site after nt 706 starting from the beginning of ORF0 ($P = 0.01$).

**Discussion**

In this paper, we have shown that the genetic diversity of PLRV, as assessed by analysing genomic sequences of isolates originating from different countries and continents, is low. However, a restricted analysis conducted on ORF0 sequences revealed that some isolates could be distinguished, such as the Peruvian isolate CIP01 and the Australian isolates. Interestingly, CIP01 was initially collected and maintained on S. tuberosum ssp. andigena, which is a different potato subspecies than that cropped in most temperate countries. The sequence divergence of CIP01 in ORF0 might result from an adaptation to the andigena subspecies. Nevertheless, three European isolates collected from the tuberosum subspecies clustered with CIP01. Australian isolates, including one isolate of PLRV-TYT, appeared to be more diversified than others and formed a second group. The geographical isolation of Australia could be the cause of this diversification, since tuber exchanges are presumably reduced with the other continents.

The apparent low mutation fixation rate seems to be peculiar to PLRV among the genus Polerovirus. Indeed, when the variability levels of ORF3 within PLRV species were compared to those within beet poleroviruses, we found significantly more sequence conservation in the former species (two-tailed, unpaired, Student $t$-test on arcsinus-squared, pairwise, identities; $t = 19.9589, df = 422, P < 0.001$). The few ORF3 sequences of the other polerovirus CYDV-RPV confirm this tendency. With regard to the Luteoviridae family as a whole, the comparison of the variability among PLRV isolates to that of BYDV-PAV isolates from even a single country like Morocco (Bencharki et al., 1999) indicates that low genetic variability is not the rule for members of this family.

The genetic homogeneity of PLRV may be related to the genetic homogeneity of both the host plant and the aphid vector. On the one hand, as opposed to the beet polerovirus complex and to BYDV-PAV, which need to infect alternative hosts to complete their epidemiological cycle, most of the PLRV isolates studied originated from a single plant species, the potato. Several studies have revealed the narrowness of the genetic base of the main potato cultivars (Demeke et al., 1996; Mendoza & Haynes, 1974) and, in particular, the monomorphism of chloroplast DNA (Powell et al., 1993), which is probably a consequence of the use of a limited number of germplasm clones during breeding programs. Also, due to the essentially clonal propagation of potato, a lower PLRV diversity can be expected. Such a correlation has already been reported for Tobacco leaf curl virus (TLCV) (family Geminiviridae) in Eupatorium sp. TLCV populations were found to be less diverse in clonal populations of Eupatorium sp. than in sexual ones (Ooi & Yahara, 1999). On the other hand, M. persicae is the main aphid species to colonize potato crops and, as it is the most efficient aphid in the transmission of PLRV, it is considered to be its main vector. This contrasts with BYDV-PAV for which different aphid species play an important role in virus transmission between different hosts. Moreover, in a study using a combination of molecular tools to characterize different clones of the M. persicae ‘complex’ collected worldwide, Terradot et al. (1999) have demonstrated very close relationships among these clones. The restricted range of plant hosts and aphid vectors, associated with their lack of genetic polymorphism, could therefore impose constraints on the genetic variability of PLRV, similar to those proposed by Nichol et al. (1993) to account for the evolutionary stasis of VSV isolates in stable environmental conditions and to maximize the fitness of PLRV in both the host and the vector.

Despite the low overall level of genetic variability, major events were shown to have occurred in the evolutionary history of PLRV. Firstly, the 27 nt deletion found in ORF1 of isolate Noir is located in a region with a concentration of active sites: the cis-acting signals involved in the −1 ribosomal frameshifting (Kim et al., 2000), the VPg-encoding sequence (van der Wilk, 1997) and the P1 proteolytic cleavage sites (Prüfer et al., 1999; Sadowy et al., 2001a). Secondly, the single substitution in isolate 14.2, which abolishes the consensus ORF0 stop codon, extends the coding capacity of this ORF by 22 aa; this should increase the size of P0 (by about 9%), a protein whose role in the accumulation of viral RNA has recently been demonstrated (Sadowy et al., 2001b). Thirdly, the Australian isolate Au16 is supposed to issue from a recombination event between putative parents from two distinct phylogenetic groups, one of which contains isolates able to cause tomato yellow top disease (group 2). Such evolutionary events may therefore confer particular characteristics (replication rate and host range) to some PLRV isolates and affect their fitness in either a positive or a negative manner.

Overlapping reading frames are often viewed as a means for RNA viruses to maximize the genetic information in smaller genomes (Jordan et al., 2000) and as a constraint on the variability of these concurrent genes (Bilse et al., 1990; Ina et al., 1994). However, we found that, in the PLRV genome, regions where two ORFs overlap do not tend to be less variable than non-overlapping ones. It seems that this constraint depends on the presence of important active sites encoded in both ORFs. We have shown that in two overlapping regions of the PLRV genome (ORF0/ORF1 and ORF3/ORF4), one ORF is preferentially conserved over the other. Such a preference was also found for Sendai virus overlapping reading frames (Fuji et al., 2001). This finding...
suggests that P0 contains putative important active sites, in agreement with the recently demonstrated indispensability of this protein for PLRV replication (Sadowy et al., 2001b). In contrast, the region of ORF1 that overlaps ORF0 is less conserved in terms of amino acid sequence. The reason why it tolerates more mutations is that it probably does not contain important active sites. In fact, the known P1 catalytic sites of the serine proteinase activity (Sadowy et al., 2001a) and the VPG domain (van der Wilk et al., 1997) both map downstream of this region. Thus, it seems that the use of overlapping reading frames is a strategy that evolved rather for its capacity to condense genetic information than to limit the accumulation of mutations.

Our finding indicates that PLRV is subject to very strong selective constraints imposed not only by its genetic structure (31% of the genome consists of overlapping coding regions) but also by both the host plant and the aphid vector, acting on the whole genome. Together, these constraints may confine PLRV to a region of the adaptive landscape (a projection of the relative fitness over the sequence space), where only few steep fitness peaks exist. This restriction in the exploitable sequence space is expected to limit the number of allowed point mutations, as most of them are deleterious and will bring the variant down from its fitness peak, resulting in its elimination by selection. However, some major evolutionary events, like insertions/deletions (examples of isolates Noir and 14.2), may allow PLRV to reach new fitness peaks without having to go across fitness valleys. This makes PLRV an interesting model to study the nature and the influence of selective constraints on virus evolution.

We are indebted to Drs José A. Caram de Souza Dias, Danuta Hulanicka, Mike A. Mayo, Luis F. Salazar and John E. Thomas for the kind supply of PLRV isolates. We thank Dr Ziheng Yang for helpful discussion about the use of models in PAML, and Frédérique Pasquer for help in sequencing of the isolates 14.2 and Cu87. Finally, we are grateful to Drs Emmanuel Jacquot, Noël Tordo, Denis Fargette and Manuel Plantegenest for their comments during the work and preparation of the manuscript. S.G. is supported by a fellowship from the French Ministry of Research.

References


López, L., Muller, R., Balmori, E., De La Riva, G., Ramírez, N., Doreste,
Genetic diversity of Potato leafroll virus


Received 21 December 2001; Accepted 15 February 2002