Molecular characterization of the largest mycoviral-like double-stranded RNAs associated with Amasya cherry disease, a disease of presumed fungal aetiology

Z. Kozlakidis,1 L. Covelli,2,3 F. Di Serio,4 A. Citir,5 S. Açıkgoz,6 C. Hernández,2 A. Ragozzino,3 R. Flores2 and R. H. A. Coutts1

Correspondence
R. H. A. Coutts
r.coutts@imperial.ac.uk

1Division of Biology, Faculty of Natural Sciences, Sir Alexander Fleming Building, Imperial College London, Imperial College Road, London SW7 2AZ, UK
2Instituto de Biología Molecular y Celular de Plantas (UPV-CSIC), Universidad Politécnica de Valencia, 46022 Valencia, Spain
3Dipartimento di Arboricoltura, Botanica e Patologia Vegetale, Università di Napoli, 80055 Portici, Italy
4Istituto di Virologia Vegetale del CNR, Sezione di Bari, 70126 Bari, Italy
5Tekirdag Ziraat Fakültesi, Trakya Universitesi, 59030 Tekirdag, Turkey
6Adnan Menderes University, Agricultural Faculty, Plant Pathology Department, 09100 Aydin, Turkey

The sequence of the four large (L) double-stranded RNAs (dsRNAs) associated with Amasya cherry disease (ACD), which has a presumed fungal aetiology, is reported. ACD L dsRNAs 1 (5121 bp) and 2 (5047 bp) potentially encode proteins of 1628 and 1620 aa, respectively, that are 37% identical and of unknown function. ACD L dsRNAs 3 (4458 bp) and 4 (4303 bp) potentially encode proteins that are 68% identical and contain the eight motifs conserved in RNA-dependent RNA polymerases (RdRp) of dsRNA mycoviruses, having highest similarity with those of members of the family Totiviridae. Both terminal regions share extensive conservation in all four RNAs, suggesting a functional relationship between them. As ACD L dsRNAs 1 and 2 do not encode RdRps, both are probably replicated by those from either ACD L dsRNA 3 or 4. Partial characterization of the equivalent L dsRNAs 3 and 4 associated with cherry chlorotic rusty spot revealed essentially identical sequences.

In previous publications, we illustrated that two cherry diseases, cherry chlorotic rusty spot (CCRS) from Italy and Amasya cherry disease (ACD) from Turkey, are two disorders that are likely to be of fungal aetiology, have very similar symptoms and are consistently associated with a complex profile of 10–12 double-stranded RNAs (dsRNAs), presumably of viral origin, that are absent from healthy-looking trees (Açıkgoz et al., 1994; Di Serio et al., 1996, 1998; Coutts et al., 2004; Covelli et al., 2004). We reported that four of them comprised the genome of a novel species in the genus Chrysovirus (Covelli et al., 2004; Ghabrial et al., 2005a), whilst another two comprise the genome of a novel species in the genus Partitivirus (Coutts et al., 2004; Ghabrial et al., 2005b), two genera in which all species infect fungi. In support of the mycoviral character of these six dsRNAs, mycelial structures have been observed in tissues affected by both diseases (Alioto et al., 2003). The precise nature of the putative fungus remains unknown because attempts to culture it have failed (A. Ragozzino, unpublished results); alternative approaches to tackle this question, based on the characterization of internal transcribed spacer regions, are in progress (G. Geraci, personal communication).

To verify that other ACD- and CCRS-associated dsRNAs also have properties consistent with those expected for mycoviruses and to get an insight into the taxonomic group that they belong to, here we report the complete sequence for the four largest ACD-associated dsRNAs and the partial sequence for two of the equivalent CCRS-associated dsRNAs. By using a range of randomly primed cDNA clones generated from individual gel-purified dsRNAs and sequencing, we were able to show that the two largest dsRNAs (Fig. 1)
are in fact both pairs of similarly sized dsRNAs that we have nominated large (L) 1–4 in this investigation.

The L dsRNAs 1–4 were isolated from sweet cherry leaves collected in Turkey and Italy exhibiting typical symptoms of ACD and CCRS, respectively; this material was the same as used in previous studies (Coutts et al., 2004; Covelli et al., 2004). After separation by electrophoresis on polyacrylamide or agarose gels, they were used either collectively or individually for reverse transcription, PCR amplification, cloning and sequencing as described previously (Coutts et al., 2004; Covelli et al., 2004). The cDNA clones of ACD L dsRNAs 1–4 were obtained by random priming of methyl mercuric hydroxide-denatured dsRNA, with further DNA manipulations being performed according to standard protocols (Sambrook & Russell, 2001). For the synthesis of additional cDNAs covering their complete sequence, purified ACD L dsRNAs 1–4 were denatured with methyl mercuric hydroxide and subjected to a single-primer, genome-walking RT-PCR protocol developed previously to sequence plant virus RNAs (Livieratos et al., 1999). An RNA ligase-mediated (RLM)-RACE PCR procedure was used to determine the 5'- and 3'-terminal sequences of the dsRNAs (Coutts & Livieratos, 2003). The sequences of the equivalent CCRS-associated dsRNAs were obtained by the single-primer method as described previously (Covelli et al., 2004). Nucleotide and amino acid sequences were manipulated (Devereux et al., 1984), aligned by using CLUSTAL_X (Thompson et al., 1997) and drawn by using the MEGA program v. 3 (Kumar et al., 2004).

The complete sequences of ACD-associated L dsRNAs 1–4 were respectively 5121, 5047, 4458 and 4303 bp in length and each included one open reading frame (ORF) potentially encoding proteins of 1628, 1620, 1363 and 1294 aa with molecular masses of approximately 178 920, 179 218, 151 860 and 143 120 Da, respectively. For ACD L dsRNAs 3 and 4, the C-terminal regions of the predicted proteins (residues 720–1100 and 720–1099, respectively) contained the eight conserved motifs characteristic of RNA-dependent RNA polymerases (RdRps) of dsRNA viruses infecting simple eukaryotes (Bruenn, 1993) (Fig. 2a). Multiple alignment of these regions (Fig. 2a) and assembly of a phylogenetic tree (Fig. 2b) revealed that ACD L dsRNAs 3 and 4 were most similar to the RdRp regions of the totiviruses Saccharomyces cerevisiae virus L-A (ScV-L1) and Ustilago maydis virus H1 (UmV-H1). Further analysis of the putative RdRp regions encoded by ACD L dsRNAs 3 and 4 with the program BLASTP (http://www.ncbi.nlm.nih.gov) revealed similarities between them and the ankyrin-repeat proteins of viral origin (data not shown), but the significance of this observation is unclear.

**Fig. 1.** Agarose-gel electrophoresis (1% in TAE) of dsRNA-rich preparations from ACD-affected leaves (lane 2) and control preparations obtained under the same conditions from an asymptomatic cherry tree (lane 3), compared with DNAs of known molecular size indicated to the left in kbp (lane 1). The four L dsRNAs were derived from the upper two dsRNA species seen on the gel in lane 2. The four genomic dsRNAs of a chrysoivirus and the two genomic dsRNAs of a partitivirus, characterized previously (Coutts et al., 2004; Covelli et al., 2004), are also indicated.
Fig. 2. Relationships between the ACD-associated large (L) dsRNAs 3 and 4 and some representative species of three families of dsRNA viruses. (a) Multiple alignment of the region containing the eight motifs conserved in RdRps of dsRNA viruses. Abbreviations precede virus names and GenBank accession numbers are also shown. Family Totiviridae: genus Totivirus: UmV-H1 (Ustilago maydis virus H1), ScV-L1 (Saccharomyces cerevisiae virus L-A), ACD-L 4 (Amasya cherry disease large dsRNA 4), ACD-L 3 (Amasya cherry disease large dsRNA 3), Hv190SV (Helminthosporium victoriae virus 190S); genus Leishmaniavirus: LRV1-1 (Leishmania RNA virus 1-1); genus Giardiavirus: TVV (Trichomonas vaginalis virus), GLV (Giardia lamblia virus). Family Chrysoviridae: genus Chrysovirus: PcV (Penicillium chrysogenum virus), Hv145SV (Helminthosporium victoriae virus 145SV), ACD-CV (Amasya cherry disease-associated chrysovirus). Family Partitiviridae: genus Partitivirus: Ahv-2 (Atkinsonella hypoxylon virus 2H), RhsV 717 (Rhizoctonia solani virus 717), FpV1 (Fusarium poae virus 1), ACD-PV (Amasya cherry disease-associated partitivirus dsRNA 1); genus Alphacryptovirus: BCV-3 (Beet cryptic virus 3). Numbers at the top refer to the eight motifs conserved in RdRps of dsRNA viruses of lower eukaryotes (Bruenn, 1993). Numbers of residues between the conserved motifs are indicated. Asterisks in the consensus denote identical amino acids and capital letters denote those found in at least five sequences. (b) Unrooted phylogenetic tree based on the neighbour-joining method with a 10,000 replicate bootstrap search. Bootstrap values (expressed as percentages) are indicated at the branch points.

### Table: Multiple Alignment of RdRp Conserved Motifs

<table>
<thead>
<tr>
<th>Virus</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>UmV-H1</td>
<td>LYGRG</td>
<td>WLVGSSAG</td>
<td>GKAAY</td>
<td>DVDDENPMH</td>
<td>GLYSDROTTNNTNLIAYA</td>
<td>H</td>
<td>GDDIT</td>
<td>EYLRI</td>
</tr>
<tr>
<td>ScV-L1</td>
<td>IMNRG</td>
<td>WVPAGSSVHS</td>
<td>GKRQAIY</td>
<td>DVDDENQSHS</td>
<td>TLSSNLRTFFSPNTNLWAYN</td>
<td>HNGDDVMI</td>
<td>ELFRL</td>
<td>16</td>
</tr>
<tr>
<td>ACD-L 4</td>
<td>LNHRA</td>
<td>TTFAGSAAFT</td>
<td>SKRLAF</td>
<td>DKAPFNNLIT</td>
<td>GMFSTGRTFLNLYNLNNRAT</td>
<td>HSSGDDYY</td>
<td>ELFRL</td>
<td>12</td>
</tr>
<tr>
<td>ACD-L 3</td>
<td>LNHRA</td>
<td>ATFAAGSAFT</td>
<td>SKRLAF</td>
<td>DARNENLHRT</td>
<td>GMFSTGRTFLNLYNLNNRAT</td>
<td>HSSGDDVY</td>
<td>ELFRL</td>
<td>12</td>
</tr>
<tr>
<td>Hv190SV</td>
<td>LQGRY</td>
<td>WCVRNGSQA</td>
<td>GKDRAIF</td>
<td>DVDDENQSHS</td>
<td>TLSSNLRTFFSPNTNLWAYN</td>
<td>HAGDDVY</td>
<td>ELFRL</td>
<td>12</td>
</tr>
<tr>
<td>LRV1-1</td>
<td>LSSGR</td>
<td>WAANSGHS</td>
<td>GKRTRLL</td>
<td>DVDDENQSHT</td>
<td>TLSSNLRTFFSPNTNLWAYN</td>
<td>HAGDDVYL</td>
<td>ELFRL</td>
<td>12</td>
</tr>
<tr>
<td>TVV</td>
<td>LSSGR</td>
<td>WSSGS</td>
<td>GKEFITY</td>
<td>DVDDENQSHT</td>
<td>TLSSNLRTFFSPNTNLWAYN</td>
<td>HAGDDVIL</td>
<td>ELFRL</td>
<td>12</td>
</tr>
<tr>
<td>GLV</td>
<td>LSSGR</td>
<td>WGGY</td>
<td>GQSNFQPD</td>
<td>DVDDENQSHT</td>
<td>TLSSNLRTFFSPNTNLWAYN</td>
<td>HAGDDVIL</td>
<td>ELFRL</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consensus</th>
<th>L</th>
<th>G</th>
<th>R</th>
<th>G*</th>
<th>A*</th>
<th>S*</th>
<th>Q*</th>
<th>V*</th>
</tr>
</thead>
<tbody>
<tr>
<td>LGRG</td>
<td>WTKG</td>
<td>GLV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WTGS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>G</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>M</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>W</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://vir.sgmjournals.org 3115
As both ACD L dsRNAs 3 and 4 potentially encode proteins that show similarities to the conserved motifs of RdRps of several totiviruses and share nucleotide and amino acid identities of 67 and 68 %, respectively, both can therefore be regarded tentatively as two novel species of the family Totiviridae. However, in contrast with most members of this family, whose genomic dsRNAs contain two ORFs — the first, ORF A, encoding the coat protein (CP) and the second, ORF B, the RdRp — those from ACD L dsRNAs 3 and 4 potentially encode a single protein, in the C-terminal region of which are found the eight motifs conserved in RdRps. In the genus Totivirus, UmV-H1 is related phylogenetically to L dsRNAs 3 and 4 (Fig. 2b), but whether ACD L dsRNAs 3 and 4 have similar properties, e.g. UmV-H1 dsRNA is translated into a single polypeptide that is then processed to the CP and the RdRp by a (viral) papain-like protease (Kang et al., 2001), is not known. Whilst ACD L dsRNAs 3 and 4 might have a similar expression strategy, a search of amino acid motifs similar to those present in viral proteases failed to identify any in the proteins predicted from their sequences. BLASTP analysis of the N-terminal regions of proteins predicted from the sequences of ACD L dsRNAs 3 and 4 did not reveal any significant similarity with CPs of other totiviruses. Additional comparative analysis, using the programs LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html) and SCANPROSITE (http://www.expasy.org/prosite), of these N-terminal regions with the N-terminal regions of the single ORF of UmV-H1 and ScV-L1 fusion protein (Wickner et al., 2005) and the global database, respectively, revealed scattered amino acid identity of between 19 and 32 % for the totiviruses and the pattern M—X(6)—P—D—X(3)—T between aa 637 and 649, also present in the CPs of some other viruses, but absent from totivirus CPs characterized thus far (data not shown). Whether the N-terminal regions of proteins predicted from the sequences of ACD L dsRNAs 3 and 4 encode CPs is thus far unknown.

For ACD L dsRNAs 1 and 2, the predicted proteins were 37 % identical to one another, contained no recognizable motifs and had no similarity to any other proteins found in global databases. However, the two dsRNAs appear to be related to ACD L dsRNAs 3 and 4 (see below).

The lengths of the 5′-untranslated regions (5′ UTRs) flanking the single ORFs of the ACD-associated L dsRNAs 1–4 were 35, 35, 59 and 73 nt, respectively. The lengths of the 3′ UTRs of the same molecules were 199, 149, 307 and 345 nt, respectively. The complementary strands did not contain ORFs of a minimal size compatible with a functional protein (data not shown). The 5′-terminal 10 nt were almost identical in all four dsRNAs and there was also significant, although lower, conservation in the sequences of the 3′ UTRs (Fig. 3). As there is significant identity in the 5′-terminal sequences and the 3′ UTRs of ACD L dsRNAs 1–4 (Fig. 3), and ACD L dsRNAs 1 and 2 do not apparently encode RdRps, both may be replicated by the RdRp of either ACD L dsRNAs 3 or 4 and be part of a multipartite genome. The possibility that ACD L dsRNAs 1 or 2 may be satellite viruses or satellite dsRNAs, dependent functionally on either ACD L dsRNAs 3 or 4, appears less likely. The absence of any similarity of the proteins predicted from the sequences of either ACD L dsRNAs 1 or 2 to any other protein makes any further speculation very difficult. However, it is unlikely that ACD L dsRNAs 1 and 2 are associated functionally with other viruses found in ACD-diseased tissue because the UTRs of both dsRNAs have no sequence similarity to the dsRNAs of the two previously characterized viruses, both of which are complete in genomic RNA content (Coutts et al., 2004; Covelli et al., 2004).

Only partially complete nucleotide sequences for two of the CCRS-associated L dsRNAs, 3 and 4 (1840 and 3557 nt, respectively), were determined and visual inspection of the dsRNAs by PAGE revealed that L dsRNA 2 might not always be present in Italian samples (results not shown). However, as the partial sequences of CCRS-associated L dsRNAs 3 and 4 were 98 and 97 % identical to the equivalent ACD-associated L dsRNAs, respectively, it is reasonable to assume that the CCRS-associated L dsRNAs 3 and 4 are variants of ACD L dsRNAs 3 and 4, described fully here.

It is now generally accepted and verified by sequencing that the incidence of multiple viruses in single fungal isolates is a relatively common occurrence. This has been reported for mitoviruses in Ophiostoma novo-ulmi, infecting Dutch elm (Hong et al., 1998), for two viruses belonging to the families Totiviridae and Chrysoviridae found in a single isolate of Helminthosporium victoriae that infects oats (Ghabrial et al., 2002), for several viruses belonging to different families in Helicobasidium mompa, which infects a wide range of plants (Nomura et al., 2003; Osaki et al., 2004, 2005), and for three unrelated viruses in Gremmeniella abietina, infecting coniferous trees (Tuomivirta & Hantula, 2005). However, the complexity of viruses associated with the fungus or fungi presumed responsible for ACD and CCRS is unprecedented and appears to include representatives of at least three fungal

![Fig. 3. Comparison of the 5′- and 3′-untranslated terminal regions of the coding strands of the four ACD large dsRNAs. Identical nucleotides have a black background, whereas those conserved in two or three of the dsRNAs are shaded.](image-url)
dsRNA virus families (Coutts et al., 2004; Covelli et al., 2004; this study). Whilst we can report the presence of two new viruses, ACD L dsRNAs 3 and 4, that are related to the family Totiviridae, until their mode of replication and their relationships with ACD L dsRNAs 1 and 2 are known, we consider it prudent to nominate them as tentative novel members of this family.

Acknowledgements

Work in the laboratories of R.F. and A.R. was partially supported by the Generalidad Valenciana (Spain) and the Ministero delle Politiche Agricole e Forestali (Italy), respectively. Z. K. was supported by a bursary from the Faculty of Natural Sciences, Imperial College London, UK. We thank Dr D. Alioto for detailed information about the fungus-like organism found in CCRS and ACD symptomatic leaves. During this work, L.C. was the recipient of predoctoral fellowships from the Marie Curie Program (European Union) and from the Università ‘Federico II’ di Napoli (Italy).

References


