Geminivirus-related extrachromosomal DNAs of the X-clade phytoplasmas share high sequence similarity

Djaouida Rekab, Luigi Carraro, Bernd Schneider, Erich Seemüller, Jianchi Chen, Chung-Jang Chang, Romano Locci and Giuseppe Firrao

Author for correspondence: Giuseppe Firrao. Tel: +39 432 558503. Fax: +39 432 558501. e-mail: firrao@pldef.uniud.it

INTRODUCTION

In the last few years, the taxonomy of non-helical plant-pathogenic mollicutes, previously named mycoplasma-like organisms, has been extensively revised on the basis of molecular data. As a result of the analysis of the 16S rRNA gene, the trivial name phytoplasmas has been given to this monophyletic group of micro-organisms. As a side effect of the increasing sequence information available, methods based on PCR amplification and RFLP analysis of the 16S rRNA gene (16S rDNA) have been developed for the rapid assignment of phytoplasma isolates to major phylogenetic clusters (Lee et al., 1993; Schneider et al., 1993).

As more phytoplasmas are being classified, it is apparent that some biological properties, such as plant/animal vector specificity and symptoms, are not phylogenetically significant as was supposed before molecular methods had been introduced. The limited value of the overall genetic similarity among phylogenetically different but pathologically similar phytoplasmas, confirmed by RFLP analysis with random chromosomal probes, supports the hypothesis that the specificities of the interactions with the host may be determined by a relatively small set of genes, or even by plasmid-carried genes (Kuboyama et al., 1998).

The phytoplasmas are at present classified in 14 different clades (being described as Candidatus species). The most represented is the aster yellows (AY)-clade, which is composed of phytoplasmas which infect predomin-
ently herbaceous plants, are very widespread and have a genome size of 850–1200 kbp. After the AY-clade, one of the largest phytoplasma groups is the X-clade, which is composed of phytoplasmas related to the agent of Western X-disease of cherry and peach. X-clade phytoplasmas, which typically have genome sizes of 600–700 kbp, are mainly found in North America, where they infect many plant species. Until recently, only the witches’ broom disease of blueberry (Siller et al., 1986) was reported to occur in Europe.

A few years ago, a geographically restricted area in north-eastern Italy, Cavazzo Carnico, was found to have different plant species showing symptoms of phytoplasma infection. Most herbaceous plants were infected with phytoplasmas with a restriction fragment pattern of PCR-amplified 16S rDNA fragments (16S rDNA-RFLP pattern) similar to the X-clade. In the remaining herbaceous plants the phytoplasmas showed a 16S rDNA-RFLP pattern similar to the AY-clade phytoplasmas (Osler et al., 1994).

The X-clade phytoplasmas were further analysed by Southern blot hybridization using four cloned chromosomal fragments of the Western X-disease phytoplasma as probes. No differences in the RFLP patterns were obtained although three restriction enzymes were used (Firrao et al., 1996, Recab, 1998). The isolates from different plant species were therefore thought to represent strains of a single phytoplasma, which was named the Italian clover phyllody phytoplasma (ICPh), since it was first detected in *Trifolium repens* (Firrao et al., 1996). In this paper, we report the analysis of extrachromosomal DNAs (EC-DNAs) which were found to be associated with the ICPh strains collected in Cavazzo Carnico and with the phylogenetically related *Vaccinium* witches’ broom (VAC) and walnut witches’ broom (WWB) phytoplasmas.

**METHODS**

**Origin and transmission of phytoplasma isolates.** Surveys for the identification of phytoplasma-infected herbaceous plants in a plot in Cavazzo Carnico, Udine, Italy, started in 1992. The site was visited every 15 d from April to October for 4 years. Plants showing symptoms of leaf yellowing or reddening, virescence, phyllody, witches’ broom and stunting were collected. Symptomatic plants were uprooted, planted in pots and maintained inside a greenhouse. Phytoplasmas were transferred to periwinkle (*Catharanthus roseus*) by dodder bridge using a clone of *Cuscuta campestris* and a previously described inverse bridge technique (Carraro et al., 1991). They were named according to the plant species of origin as follows: TRR (*Trifolium repens*), TA (*Taraxacum officinale*), MA (*Chrysanthemum leucanthemum*), ER (*Erigeron annuus*), CR (*Crepis biennis*), LEO (*Leontodon hispidus*) and VE (*Veronica arvensis*). Two periwinkle isolates were obtained by feeding naturally infected insects on periwinkle plants and were named API (from the leafhopper *Euscelis variigatus*) and CP (from *Euscelidius plebeius*; this strain was collected near Bergamo, Italy, and not in Cavazzo Carnico).

The VAC phytoplasma was isolated in periwinkle by dodder bridge from a *Vaccinium myrtillus* (blueberry) plant showing witches’ broom symptoms as described earlier by Siller et al. (1986), and then maintained by grafting.

For the insect-mediated transmission of phytoplasmas, about 30 *Euscelidius variigatus* leafhoppers, grown in a controlled environment and fed on barley, were placed in cages on field-collected *Trifolium repens* or *Chrysanthemum leucanthemum*. After 3 d, the insects were separated into two groups which were transferred to *Trifolium repens* and *Chrysanthemum carinatum*, respectively. In subsequent experiments the insects were placed on an infected *Chrysanthemum carinatum* plant and then transferred to *Trifolium repens*, *Chrysanthemum indicum*, *Zinnia elegans*, *Apium graveolens*, *Gomphrena globosa* and different ecotypes of *Arabidopsis thaliana* (Co-2, Ms-0, Loh-0 and BR-0). Three months after inoculation the nucleic acids were extracted from the plants showing symptoms.

**Nucleic acid extraction and hybridization.** Nucleic acids were extracted using the phytoplasma enrichment procedure described by Ahrens & Seemüller (1992). For Southern blots an RNase treatment was performed followed by phenol/chloroform extraction and ethanol precipitation of the DNA. For each sample, 5 μg DNA was loaded on a 0.8% (w/v) agarose gel which was run for 16 h at 2 V cm⁻¹ in Tris/acetate buffer. After staining and photographing under UV light the DNA was capillary-transferred to nylon membrane using a standard protocol (Ausubel et al., 1992). The hybridizations were performed using the DIG-system, according to the manufacturer’s protocol (Boehringer Mannheim). A mixture of pVAC79 and pVAC131 (Schneider et al., 1992), or individual fragments obtained from these plasmids by PCR amplification as detailed in Results, were used as probes. An EC-DNA fragment (pSA45) from the severe strain of the Western AY phytoplasma cloned in pUC19 (Kuske et al., 1991) was kindly provided by B. C. Kirkpatrick (University of California, Davis, CA, USA).

**DNA cloning and sequencing.** Cloning of EC-DNA from the X-clade VAC (plasmids pVAC79 and pVAC131) and WWB (plasmid pWWB1) phytoplasmas has been described previously (Schneider et al., 1992; Chen et al., 1992a,b). Cloning of EC-DNA from the DNA of ICPh isolates was carried out following the method described by Schneider et al. (1992). Briefly, EC-DNA bands, which could be clearly visualized under UV light on ethidium-bromide-stained agarose gels after running undigested DNA, were excised from the gel, purified, digested with HindIII and ligated to a HindIII-linearized, dephosphorylated pUC19 plasmid vector. The ligation reactions were used to transform *Escherichia coli* (strain XL-1 Blue) competent cells by the calcium chloride method (Ausubel et al., 1992).

**PCR amplification.** The nucleic acids extracted from periwinkle were checked for the presence of phytoplasmas by PCR using the universal primers for phytoplasmas rU3/fU5 described by Lorenz et al. (1995).

Several primers were developed for the amplification of EC-DNA from the nucleic acids extracted from phytoplasma-infected plants. The sequences of the primers used are: DjD1, 5'-TTGTGCCGAAATAGATAAAGC-3'; DjU1, 5'-GATTGAGTCCCATTA−

AACCAGCAGAT−3'; DjD2, 5'-AAAAATATGTA−

ACCGACGT−3'; DjU2, 5'-GATTGAGTCCCATTA−

TTGTACC−3'; DjD3, 5'-TTATATGTAATGGACATTTTT−

CTT−3'; DjU3, 5'-AGAAGATGTCTCATTACATAAA−3'. The standard M13 universal (M13U) and M13 reverse (M13R) sequencing primers (Ausubel et al., 1992) were also used in amplifications. Amplification of EC-DNA was carried out in a
40 μl volume of PCR buffer (Boehringer Mannheim) in the presence of 300 ng of each primer, 0.2 mM dNTPs, 1 unit Taq polymerase and about 200 ng template DNA. Reactions were cycled 35 times with the following parameters: 98 °C (30 s), 55 °C (30 s) and 72 °C (30–120 s depending on the size of the expected product).

Sequence analysis. The similarity between the putative translation products of the ORFs identified in the sequence of pVAC79 and database entries was calculated using the program BESTFIT of the Wisconsin Package (Genetics Computer Group, 1994). In the maximum-parsimony analysis, carried out with the program PAUP (Swofford, 1993), the accession numbers P21947 (abutilon mosaic virus), P27258 (potato yellow mosaic virus), Q65418 (beet curly top virus), Q88888 (tomato pseudo-curl top virus), JQ1553 (Panicum streak virus), JQ1359 (Miscanthus streak virus), AF039530 (Egyptian sugarcane streak virus), C42452 (tobacco yellow dwarf virus), P29048 (squash leaf curl virus) and M88179 (bean dwarf mosaic virus) were used together with the ORFB sequence.

RESULTS

Investigation of the biological role of EC-DNAs

As a result of the surveys in the area of Cavazzo Carnico, the AY-clade phytoplasmas VE and LEO and the X-clade phytoplasmas ICPH strains TRR, TA, MA, ER, CR and API were isolated on periwinkle. The DNAs of the various ICPH strains, which were indistinguishable by chromosomal RFLP (Firrao et al., 1996; Rekab, 1998), showed different banding patterns when hybridized to the EC-DNA probes pVAC79 and pVAC131 in Southern blots (Fig. 1; and data not shown). A subsequent field survey indicated that variability of the EC-DNA banding pattern also occurred among samples from the same plant species, but it was more marked between samples from different plant species: typically, two to three differences were found in transmission efficiency, disease severity or EC-DNA patterns of the recipient plants. In addition, A. H. Purcell (unpublished), while all patterns detected in samples from different plant species differed in size and number of the EC-DNA bands. This finding raised the question of whether the changes in the EC-DNA banding patterns were related to the original host plant species of the ICPH infection. Therefore, leafhoppers were used to transmit the phytoplasmas from each of the field-collected ICPH-infected Chrysanthemum leucanthemum and Trifolium repens plants to both Chrysanthemum carinatum and Trifolium repens plants. No differences were found in transmission efficiency, disease severity or EC-DNA patterns of the recipient plants. In addition, the Chrysanthemum carinatum plant infected with strain TRR was used as a donor for further insect-mediated transmissions of the phytoplasma to several different plant species. The undigested DNA extracted from the plants which showed symptoms after 3 months (i.e. Trifolium repens, Apium graveolens, Chrysanthemum carinatum and four ecotypes of Arabidopsis thaliana) was analysed by Southern blot hybridization using EC-DNA probes. As shown in Fig. 2, no variation in the hybridization pattern was detected, except for the back-transmission to Trifolium repens, where two EC-DNA bands were missing. This was the only variation in the EC-DNA hybridization pattern which was observed during this work on the plants kept in the greenhouse: none of the strains maintained by grafting onto periwinkle for over 4 years showed a change in the EC-DNA banding pattern.

The wide distribution of EC-DNAs in all ICPH isolates prompted us to determine the presence of homologous sequences in the AY-clade phytoplasma isolates from the same location as the ICPH strains. At this site AY-
clade and ICPh phytoplasmas may infect the same plant (Osler et al., 1994; Firrao et al., 1996) and insect host (S. Palmano & G. Firrao, unpublished) and multiple infections occur (Firrao et al., 1996). Hybridizations with probe pSA45 showed that the AY-clade phytoplasmas VE and LEO contained EC-DNA molecules, but they did not hybridize with EC-DNA from the VAC phytoplasma. Conversely, DNA extracted from plants infected with the ICPh strains did not react with pSA45 (not shown).

Sequence similarity among X-clade EC-DNAs

The insert of pVAC79 was sequenced (1489 bp, accession no. AJ012625) and primers were devised for the amplification of subfragments, as depicted in Fig. 3. Independent Southern blot hybridizations with the labelled DNA fragment amplified by the primer pairs M13U/DjD1, DjD2/DjU1, DjU2/DjD3 and M13R/DjU3 to DNA extracted from periwinkles infected with different ICPh strains showed that homologous regions span the entire sequence cloned in pVAC79 (data not shown). Partial sequences were also obtained from the cloned fragment of pWWB1 (212 bp, accession no. U29882; and 180 bp, accession no. U29883). A comparison between sequences U29882 and AJ012625 revealed a region of high homology (Fig. 3). Consensus primers devised from those sequences (DjU1 and DjD2) were used to amplify by PCR the homologous regions from the ICPh strains (Fig. 4a). The amplification products of strains TA and TRR were sequenced (accession no. AJ012624) and were found to be identical to each other and highly similar to those of the VAC (99.5%) and WWB (89.5%) phytoplasmas. Notably, the variation among nucleotide sequences produced non-familial amino acid changes in only two cases.

Sequence analysis

Primers DjU1 and DjD3, derived from the 3' and 5' ends of the fragment cloned into pVAC79, amplified a DNA fragment of about 1100 bp from the DNA of the VAC phytoplasma and from the ICPh strain API (Fig. 4b). Using the reverse and complement primers (DjD1/DjU3), a DNA fragment of about 3000 bp was amplified from DNA of the ICPh strain API (Fig. 4c). This DNA fragment, purified by gel electrophoresis and labelled, hybridized with the phytoplasmal EC-DNA bands when used as probe in Southern blot hybridization (not shown). These results supported the hypothesis that the single hybridization signal of the HindIII-cut API DNA (Fig. 1, lane 7), corresponding to a DNA fragment estimated to be 4300 bp, might represent the full-length EC-DNA molecule. Attempts were made to clone this DNA fragment after gel purification for sequencing, since this approach should produce more reliable results than the direct sequencing or cloning of the PCR products, given the very high mol% A+T content and the abundance of short repeated sequences. However, attempts to clone the HindIII-restricted EC-DNA from several ICPh strains failed and the sequence analysis was focused on pVAC79.

The cloned insert of pVAC79 consisted of a partial ORF (ORFB) and a complete ORF (ORFA) separated by a region rich in relatively short direct and inverted repeats (Fig. 3). Database searching revealed similarity (29%) of the ORFA product to a putative product (accession no. Q84486) of the DNA sequence of the Paramecium bursaria endoparasite Chlorella virus 1 (PBCV1), a virus of Chlorella strain NC64A. The potential translation product of ORFB showed highly significant homology with...
to the replication initiation (Rep) proteins of geminiviruses, with similarity values ranging from 61% (with tobacco yellow dwarf virus) to 57% (with Panicum streak virus). Fig. 5 shows the alignment of the putative translation product of ORFB with the C-terminal domain (helicase) of the Rep protein of geminiviruses belonging to the Begomovirus and Mastrevirus genera. A maximum-parsimony analysis with 10 representative strains of the family Geminiviridae indicates that ORFB is related to, but evolved independently from, the Begomovirus and Curtovirus genera of geminiviruses (not shown).

Analysis of the ORFB–ORFA intergenic region DNA indicated putative eukaryotic signals, such as a TATA box upstream of ORFA and a polyadenylation signal 6 nt downstream of the stop codon of ORFB. However, prokaryotic features were also present, i.e. a sequence starting 13 nt before the beginning of ORFA which is complementary to the Shine-Dalgarno sequence of the phytosplasmal 16S rRNA and —35 and —10 signals very similar to those of the promoter of the 16S rRNA gene (Kuske & Kirkpatrick, 1992; see the feature table of accession no. AJ012625 for further detail).

**DISCUSSION**

Since the beginning of molecular studies on phytoplasmas, it has been noted that they have associated EC-DNAs (Davis et al., 1988; Kuske & Kirkpatrick, 1990; Harrison et al., 1991; Davis et al., 1990; Denes & Sinha, 1991; Nakashima et al., 1991). It is now clear that there are at least two distinct groups of EC-DNAs associated with phytoplasmas. One group includes the EC-DNAs of the AY-clade phytoplasmas. Goodwin et al. (1994) detected very high similarities of the EC-DNAs associated with the AY phytoplasma and the elm yellows phytoplasma, which belongs to a phylogenetic group not closely related to the AY-clade. A second group, which was the subject of the present study, appears to have a narrower host range, being associated only with phytoplasmas belonging to the X-clade. Again some sequences are highly conserved, even among isolates collected from different parts of the world, with different plant and presumably insect hosts.

Examination of the phytoplasmal EC-DNA in the restricted area of Cavazzo Carnico, where phylogenetically different phytoplasmas occur in the same plant and insect host (Firrao et al., 1996; S. Palmano & G. Firrao, unpublished), indicated that the two different EC-DNA groups present there correlate with the phylogeny of their associated phytoplasma, with no inter-clade horizontal transfer. The high sequence similarity observed among EC-DNAs of phylogenetically related phytoplasmas of very different geographical origin and different hosts suggests an ancient common origin. Similar observations have been made on the EC-DNAs of AY-clade phytoplasmas: the maize bushy stunt phytoplasma and the Western AY phytoplasma have been reported to share EC-DNA sequences without having plant or insect hosts in common (Davis et al., 1988; Kuske et al., 1991).

In both groups of EC-DNAs their sizes and number vary significantly. The sequence information and the transmission experiments carried out in this work support the conclusion that EC-DNA size variations are not related to the interaction with different host plant species. The presence of different EC-DNA banding patterns in field-collected samples may be more easily explained by hypothesizing recombination events involving the regions of the EC-DNA rich in repeated sequences. Recombinant EC-DNA molecules may then be efficiently spread from plant to plant by insect-mediated transmission.

Despite the size variability, their striking sequence conservation and almost universal occurrence in all strains suggest a function for these molecules or at least for the part of the molecule that does not appear to be involved in recombination events which lead to size variability. Kuboyama et al. (1998) reported the association of an EC-DNA banding pattern variation with the occurrence of milder pathogenic characteristics. Denes & Sinha (1991) reported the loss of insect transmissibility of phytoplasmas when EC-DNA rearrangements occurred after *in vitro* cultivation. In this work, no differences were found in the insect-mediated transmission efficiency and in the infected plant's symptoms expression for phytoplasmas associated with EC-DNAs of different size. In conclusion, the involvement of the X-clade-specific EC-DNAs in host–phytoplasma interactions, as postulated by Kuboyama et al. (1998) and Denes & Sinha (1991) for other EC-DNAs, could not be ruled out by our results and remains to be determined. It should also be noted that several insect-transmitted, highly phytopathogenic phytoplasmas do...
not appear to contain EC-DNA molecules (Schneider et al., 1992; G. Firrao & B. C. Kirkpatrick, unpublished).

EC-DNA molecules associated with phytoplasmas have been postulated to be plasmids because virus particles were not observed by electron microscopy (Kuske & Kirkpatrick, 1990; Kuske et al., 1991; Goodwin et al., 1994). Rolling circle replication (RCR)-type plasmids are widespread among Gram-positive bacteria and a plasmid of the pMV158 family (pKMK1) has been reported to occur in Mycoplasma mycoides subsp. mycoides (King & Dybvig, 1992). Recently, Kuboyama et al. (1998) characterized a RCR plasmid (named pOYW1) from the AY-clade onion yellows phytoplasma and detected motifs characteristic of the Rep proteins of the pMV158 plasmid family in the putative product of one of the identified ORFs. Conversely, the results reported in the present paper showed that the EC-DNA associated with the X-clade VAC phytoplasma has the potential to encode a protein similar to the geminivirus Rep protein. The EC-DNA of the VAC phytoplasma showed prokaryotic features, such as Shine-Dalgarno and promoter sequences, in addition to eukaryotic polyadenylation signals and TATA boxes, as previously reported for the geminivirus abutilon mosaic virus (Frischmuth et al., 1990). Koonin & Ilyina (1992) have hypothesized that the geminiviruses may have originated from prokaryotic ssDNA plasmids, on the basis of the occurrence of moderately similar motifs in the RCR initiator domain (N-terminal region) of the Rep proteins of the geminivirus and the RCR initiator protein of plasmids of the pUB110 and pMV158 families. However, no similarity was reported between the geminivirus and ssDNA plasmids with reference to the helicase domain (C-terminal region) of the Rep protein. Due to the similarity of the ORFB product to the helicase domain of the geminivirus Rep protein, the X-clade EC-DNAs appear to be only distantly related to pOYW1 and their plasmid nature remains to be demonstrated.

Despite several attempts, full-length EC-DNAs of the ICPH strains could not be cloned. Since similar difficulties have been encountered in cloning other full-length phytoplasmal EC-DNAs (B. C. Kirkpatrick, personal communication), we hypothesize that the expression in E. coli of a replication-associated protein may interfere with vector or host chromosome replication.

ACKNOWLEDGEMENTS

We are grateful to Bruce C. Kirkpatrick (University of California, Davis, CA) for providing many of the probes used in this work.

REFERENCES


Received 27 November 1998; revised 23 February 1999; accepted 4 March 1999.