Nucleotide sequence, genetic organization and expression strategy of the double-stranded RNA associated with the ‘447’ cytoplasmic male sterility trait in *Vicia faba*

Pierre Pfeiffer

Institut de Biologie Moléculaire des Plantes du CNRS, 12 Rue du Général Zimmer, 67000 Strasbourg, France

The entire nucleotide sequence of the double-stranded (ds) RNA associated with the unconventional ‘447’ cytoplasmic male sterility (CMS) trait in *Vicia faba* was determined from overlapping cDNA clones and by RT–PCR. Confirming previous observations, it was found that the negative-strand was continuous and 17635 nt long, while the positive-strand featured an interruption, probably a nick, that could potentially define two subgenomic RNAs of 2735 nt and 14900 nt, with the smaller RNA being located on the 5′ side. The entire positive-strand could encode a single in-frame ORF starting at the first AUG at position 42–44 and ending with a TGA at 17517–17519. This long potential polypeptide with a predicted molecular mass of 654109 is the largest described to date in the plant kingdom and contains conserved amino acid sequence motifs typical of viral helicases and RNA-dependent RNA polymerases (RDRP). Only limited sequence homology was detected with the ORF B encoded by the hypovirulence-associated dsRNA of chestnut blight fungus, a dsRNA replicon similarly contained in host-derived membranous vesicles and considered to share a common ancestry with potyviruses. By contrast, the helicase and RDRP domains were in the same respective arrangement and shared extensive sequence homologies with those identified in the polyprotein encoded by the dsRNA isolated from Japonica rice, another dsRNA replicon featuring a specific nick in the positive-strand. Although no proteolytic self-cleavage activity has yet been demonstrated, it appears likely that this long ORF is a polyprotein that undergoes proteolytic maturation, with one of the polypeptides derived by self-cleavage being the determinant of the CMS trait.

Introduction

The cytoplasmic male sterility (CMS) trait of the ‘447’ line of *Vicia faba* is unusual in that it does not result from rearrangements of the mitochondrial DNA that lead to the production of aberrant polypeptides as has been described in several plant species (Dewey *et al.*, 1986; Young & Hanson, 1987; Lewings, 1993), but rather correlates with the presence of a double-stranded (ds) RNA of high molecular mass and unknown origin (Grill & Garger, 1981). This dsRNA is transmitted exclusively in a vertical mode, is absent from maintainer lines and is permanently lost after restoration of male fertility by crossing with a restorer line, or as a consequence of spontaneous reversion to fertility (Scalla *et al.*, 1981). Early electron microscopic observations (Edwardson *et al.*, 1976), performed at a time when CMS was considered to be of virus aetiology, revealed that all tissues of the male-sterile plants contain cytoplasmic membranous vesicles that were dubbed ‘cytoplasmic spherical bodies’ or ‘virus-like particles’, and which disappeared after restoration of male fertility. We have previously established (Lefebvre *et al.*, 1990) that these membranous vesicles do indeed contain the dsRNA described by Scalla *et al.* (1981) and Grill & Garger (1981), and that this dsRNA is associated with a specific RDRP. These structures probably correspond to a virus lacking a capsid protein which is maintained in the form of a replicative complex contained in host-derived membranous vesicles rather than as virions, very reminiscent of the hypovirulence-associated viruses of the chestnut blight fungus (Hansen *et al.*, 1985). The structure of this dsRNA was unusual, with the negative-strand being continuous while the positive-strand was interrupted and could therefore be potentially divided in two subgenomic RNAs (Pfeiffer *et al.*, 1993). Run-on RNA synthesis performed in vitro with transcription/replication complexes isolated from the cytoplasmic vesicles confirmed that a smaller subgenomic
RNA was preferentially synthesized, but no information was available on its coding capacity or on the various proteins specified by the dsRNA in general. In addition, any evidence to support the involvement of the dsRNA in the male sterility trait remained circumstantial, since all attempts to convert fertile plants to male sterility by inoculation or by transfer of the dsRNA failed (Duc et al., 1984; Turpen et al., 1988). This precluded demonstration of a cause-and-effect relationship and prevented potential extension of this CMS to other plants of agronomical interest.

In the case of the chestnut blight fungus, it has been demonstrated that the virus-mediated changes do not result from a general debilitation of the host, but rather from the differential regulation of specific genes. Thus, ORF A of the virus was identified as the determinant of traits associated with alterations of fungal phenotype, such as reduced pigmentation, and its proteolytic processing was, for instance, demonstrated to be required for reduction of fungal conidiation (Craven et al., 1993). ORF B controls virulence attenuation, and recent progress in the elucidation of the signal transduction pathways that govern fungal pathogenesis has identified the role of G protein-mediated cAMP accumulation on the reduction of fungal virulence (for a review, see Nuss, 1996).

The only data available so far for the ‘447’ dsRNA were limited to the description of the degeneration of the microspores and tapetal cells in the male-sterile line, but provided no clues as to the mechanism involved. The elucidation of the complete sequence of the ‘447’ dsRNA should now allow us to explore these avenues and to assess how the individual viral gene products that are most probably generated by autoproteolysis of the polyprotein interfere with pollen viability.

**Methods**

**Isolation and purification of dsRNA.** Briefly, dsRNA was extracted from flower buds or young leaves homogenized in grinding buffer as described (LeFebvre et al., 1990). After phenol extraction and ethanol precipitation, the nucleic acids were solubilized in TE, adjusted to 2 M LiCl, 5 mM EDTA, and incubated overnight at 0 °C. After centrifugation, soluble nucleic acids (soluble RNA, dsRNA and DNA) were ethanol-precipitated from the supernatant, pelleted and resuspended in water. Contaminating plant DNA was digested with DNase, and dsRNA was either purified by two cycles of CF-11 chromatography (Morriss & Dodds, 1979) or by electrophoresis through a 1% agarose gel run in TAE buffer and electroelution into dialysis tubing.

**Polyadenylation of dsRNA.** Purified, undenatured dsRNA was polyadenylated in vitro using E. coli poly(A) polymerase (BRL) as described by Sippel (1973); more efficient polyadenylation was obtained at later stages using yeast poly(A) polymerase (Amersham).

**Reverse transcription of denatured dsRNA and molecular cloning of cDNA.** Several methods were used to generate cDNA from dsRNA, and full denaturation of dsRNA was found to be the limiting parameter for successful reverse transcription. Both oligo(dT)12 equipped with a cloning cassette (total length 36 nt) and random hexamers or specific sequence-derived primers were used to generate cDNA clones. Only short clones were obtained after denaturation by heat with or without DMSO (Cashdollor et al., 1982; Asanizou et al., 1985), possibly due to incomplete denaturation or to rapid snapback of the separated strands. On the contrary, denaturation with 10 mM methylmercuric hydroxide (MMH) consistently yielded the best results, provided that the denaturation mix was directly pipetted into the prewarmed reverse transcription mix without prior neutralization of the MMH by mercaptoethanol (Antoniw et al., 1986; Jelkmann et al., 1989). First-strand synthesis was directed by RNase H-minus moloney murine leukaemia virus reverse transcriptase (Promega) at 42 °C for 1.5–2 h. Second-strand synthesis was performed according to Gubler & Hoffman (1983). High molecular mass cDNA was purified by centrifugation through GlassMax columns (Gibco BRL) or Sephacryl S-400 spin columns (Pharmacia). Size-selected cDNA was then phosphorylated with T4 polynucleotide kinase, blunted with Klenow or T4 DNA polymerase, and either cloned in dephosphorylated pKS (Stratagene) cleaved at the EcoRV or Smal site, or treated with restriction enzymes and cloned in an appropriately digested vector.

**Gap bridging by RT–PCR.** To bridge the gaps between non-overlapping clones, high molecular mass first-strand cDNA was selected by precipitation with 8% PEG (Pathankark & Prasad, 1991) after alkaline hydrolysis of the RNA template. PCR (30 cycles) was performed after an initial denaturation step of 5 min at 95 °C. Under these conditions, even the RNA–DNA hybrid obtained in the first-strand synthesis reaction could be directly used as a template. Primers were designed after initial alignment of the various clones, and PCR was performed using a mixture of enzymes with proof-reading capability (Expand Long Template PCR system, Boehringer Mannheim) to prevent PCR-induced mutations. PCR products were generally blunt-end cloned after phosphorylation and polishing of the ends, but some recalcitrant PCR products had to be cloned using the T-vector system of Marchuk et al. (1991).

**Determination of the extremities of the dsRNA.** Free 3' OH termini were polyadenylated in vitro and oligo(dT)12 equipped with a cloning cassette was then used to prime cDNA synthesis. For sequence determination of those extremities of the dsRNA that did not support efficient addition of poly(A), cDNA primed with termini-proximal oligonucleotides was submitted to PCR amplification using the 3' and 5' RACE (rapid amplification of cDNA ends; Frohman et al., 1988) or SLIC (single-strand ligation to single-stranded cDNA; Dumas et al., 1991) techniques.

**DNA sequencing and homology searches.** cDNA clones were sequenced by the dideoxynucleotide chain-termination method using T7 DNA polymerase (Promega). In regions difficult to sequence, compressions and strong stops were overcome using 7-deazaGTP. Incorporation close to the primer was improved by addition of Mn2+ ions. One strand was completely sequenced from a series of overlapping Exo III deletion series, while the other strand was sequenced from internal deletions derived from the restriction map thus obtained. At later stages of this study, dye terminator cycle sequencing was performed with an Applied Biosystems 373 (Perkin-Elmer) automatic sequencer. Fig. 2 shows the position of the most significant clones on the dsRNA: at least three independent cDNA clones were analysed for sequence determinations, and sequences determined by RT–PCR (see below) were derived from at least five independent clones. The detailed list of the primers used and of the cDNA clones generated can be obtained from the author on request. Sequence analysis, nucleotide and amino acid comparisons were performed using the GCG package, or the DNA analysis software DNAid and DNA Strider. For protein homology searches, the improved BLAST algorithm of Altschul et al. (1997) was used to search the EMBL, NBRF and SWISS-PROT databases, and the PROSITE database (Bairoch & Bucher, 1994) was used to identify potential function signatures.
Results

Cloning of the small subgenomic RNA

Early observations (Grill & Garger, 1981) had revealed that the ‘447’ dsRNA probably has an unusual structure, with at least one nicked strand. This was later confirmed by the studies of Lefebvre et al. (1990) and Pfeiffer et al. (1993) who demonstrated that the coding strand was interrupted, thus defining two potential subgenomic RNAs. ‘Run-on’ experiments performed in vitro with the cytoplasmic vesicles that contain the dsRNA associated with its specific RNA-dependent RNA polymerase (RDRP) showed that only the smaller RNA was elongated, reaching a limit size estimated at 4.5 kb by electrophoresis on a formaldehyde gel. This indicated its preferential, if not exclusive, transcription by the RDRP in these complexes. The existence of this subgenomic RNA and the preferential accessibility of its 3’ OH end were further confirmed by in vitro labelling of the purified dsRNA by T4 RNA ligase-mediated addition of [32P]pCp, which then allowed a tentative determination of this 3’-terminal sequence (Pfeiffer et al., 1993). This analysis also revealed that none of the 3’ ends of the dsRNA were polyadenylated, but that the 3’ OH terminus of the small subgenomic RNA was an efficient substrate for poly(A) polymerase. In vitro polyadenylated dsRNA was therefore used as a template to prime reverse transcription with oligo(dT), and specific primers were subsequently used to extend the clones obtained towards the 5’ end of the RNA.

The quality of the cDNA synthesis primed by oligo(dT) was routinely monitored by RT–PCR using as primers oligo(dT) and the antisense oligonucleotide 141CT (nt 1985–2004) mapping at an estimated 750 bp from the 3’ end of the RNA. The cloned product was subsequently used to extend the clones obtained towards the 5’ end using oligonucleotides 141NT-Xho (nt 364–339) and L15CT (nt 16946–16971, lane d) similarly reflected the variable efficiency of in vitro polyadenylation of the 3’ OH extremities of the dsRNA, and their sizes matched exactly those in the predicted structure.

Attempts to extend cDNA clones further towards the 5’ end using oligonucleotides 141NT-Xho (nt 364–339) and 141NT-Ext (nt 105–83) as primers led to cDNA clones extending up to nt 24. Similarly, PCR products obtained by 5’ RACE (Frohman et al., 1988) or SLIC (Dumas et al., 1991) extended to nt 17. No clones longer than those corresponding
to the 2·8 kb oligo(dT)-directed self-PCR product were obtained, and all of these started at nt position 1 or 2. Finally, RT–PCR performed on oligo(dT)-primed cDNA with 141NT-Xho as antisense primer produced a PCR product corresponding to the sequence between the polyadenylated 3’ OH of the negative-strand and this primer. The length of the band was 400 nt (Fig. 1, lane c), as predicted for termination near nt 1, and sequencing confirmed that the 5’ end of the small subgenomic RNA had been reached. The intensity of this band was lower than that of the 141CT-oligo(dT) product, reflecting the lower efficiency of in vitro polyadenylation of the 17 kb negative-strand vs that of the small subgenomic RNA in the dsRNA (Fig. 1, compare lanes b and c). This indicated in turn that the ca. 2·8 kb PCR product obtained with oligo(dT) alone (Fig. 1, lane a) resulted from a ‘megaprimer’ extension of oligo(dT)-primed cDNA copies of the smaller subgenomic RNA which hybridized to antisense cDNA copies of the polyadenylated 17 kb negative-strand, and sequencing confirmed that the 5’ end of the small subgenomic RNA mapped opposite the 3’ end of the negative-strand.

Reappraisal of the arrangement of the small and large subgenomic RNAs and generation of cDNA clones representative of the complete sequence of the ‘447’ dsRNA

In keeping with the general 3’-coterminal localization of subgenomic RNAs in RNA plant viruses, the preferential accessibility of the smaller subgenomic RNA to enzymes that require a free 3’ OH to add nucleotide(s) had favoured mapping the smaller subgenomic RNA at the 3’ end of the positive-strand of the dsRNA molecule (Pfeiffer et al., 1993). This implied in turn that the 3’ end of the longer subgenomic RNA would map at the discontinuity in the positive-strand, where its lack of reactivity [e.g. to poly(A) polymerase] suggested that it was either blocked as a 3’ or a 2’–3’ cyclic phosphate (presumably produced by a nuclease cleavage event), or inaccessible to these enzymes due to steric hindrance.

The reverse arrangement of the two subgenomic RNAs was confirmed by taking advantage of a NcoI site (nt 2500–2505) to generate a clone (NcoB23) that extended 186 nt beyond the 3’-terminal nucleotide of the small subgenomic RNA. Assuming that the single-stranded interruption is limited to a nick in the phosphodiester bond, the first AUG codon encountered was at position 124–126 downstream of that nick. Sequence determination with clones extending further into the larger subgenomic RNA confirmed that it encoded a continuous ORF (tentatively designated as ORF B) that ended with a UGA stop codon at position 17517–17519 (i.e. 14782–14784 nt from the nick). All other reading frames present on both strands contained numerous stop codons and are unlikely to be translated into proteins.

Clones spanning the 3’ region of ORF B were generated with several specific oligonucleotides and by 3’ RACE (Fig. 1, lane d), and the one that extended the farthest after the stop codon of ORF B was assumed to contain the 3’ terminus of the larger subgenomic RNA, which was 14900 nt long. The 3’ untranslated region was 116 nt long and devoid of a poly(A) tail in the native dsRNA. The nucleotide sequence is deposited in the EMBL database (AJ000929) and is not duplicated here.

One or two ORFs?

The ‘447’ dsRNA shares remarkable similarities with the dsRNA associated with hypovirulence in the chestnut blight fungus [Cryphonectria (Endothia) parasitica]. This dsRNA replicon is also contained in membranous vesicles (Hansen et al., 1985), in which it is associated with a specific RDRP (Fahima et al., 1993). The virus-like genetic organization of the chestnut blight fungus dsRNA and its expression strategy (Shapira et al., 1991) suggest that it shares a common ancestry with potyviruses (Koonin et al., 1991). It was first referred to as HAV, for chestnut blight fungus hypovirulence-associated virus, and later renamed CHV1, for Cryphonectria hypovirulence-associated virus 1. This dsRNA features two ORFs that are translated in a ‘stop/restart’ mode and undergo cotranslational autoproteolysis (Choi et al., 1991b). ORF A by itself is able to suppress fungal sporulation and reduce pigment and laccase accumulation (Choi & Nuss, 1992), and this suppressive activity was mapped by Craven et al. (1993) to the autocatalytic papain-like protease p29, resembling the potyvirus-encoded protease HC-Pro (Choi et al., 1991a). ORF B encodes the ‘house-keeping genes’ responsible for the helicase, polymerase and a second protease function. Depending on the virus strain, ORF A may or may not encode a protease.

Sequencing multiple clones of cDNA and of the 141CT-oligo(dT) PCR product confirmed, however, that there was no stop codon delimiting ‘ORF A’ and that the ORF continued through the discontinuity into the other subgenomic RNA. Such in-frame continuation of ‘ORF A’ into ‘ORF B’ suggests, therefore, that translation of a single, continuous, 5825 aa long polypeptide can occur, possibly from those rare molecules of dsRNA that feature an uninterrupted positive-strand (Pfeiffer et al., 1993). Given their size and the relationship of the ‘447’ dsRNA with potyviruses (see below), it is at any rate expected that such polyprotein(s) undergo(es) proteolytic self-maturation. By analogy with CHV1 (Shapira et al., 1991) and for comparison purposes, however, the two blocks of genetic information located upstream and downstream of the nick will be referred to as ‘ORF A’ and ‘ORF B’, respectively. Fig. 2 highlights the important features of the ‘447’ CMS-associated dsRNA of V. faba and summarizes its genetic organization as deduced from its complete nucleotide sequence.

Comparison of the polypeptide(s) encoded by the V. faba dsRNA and by other RNA replicons: ‘ORF A’ is specific to the ‘447’ dsRNA

All cases of CMS elucidated to date result from the expression of abnormal or chimaeric mitochondrial protein
Fig. 2. Map of the major cDNA and RT–PCR clones (A), coding capacity (B), main sequence features (C) and structure (D) of the dsRNA associated with the '447' CMS in *V. faba*. (A) Clones pZO140–143 (Turpen et al., 1988) are indicated by hatched shading, the major cDNA clones by open boxes, 3′ RACE clones of the extremities are black, and RT–PCR clones bridging the gaps are shaded grey. The positive-strand features a 41 nt long 5′ UTR with two in-frame stop codons upstream of the AUG 42–44, and a 116 nt long 3′ UTR with no ORFs of significant length after UGA (17517–17519). The single long ORF encoded by the dsRNA is interrupted by a discontinuity in the coding strand at nt 2735 (arrow). The region ‘B’ of the 5825 aa long potential protein encoded by this ORF contains the conserved helicase and RDRP motifs typical of RNA viruses which are detailed in Fig. 4.
variants that usurp the role of their functional counterparts in the anthers, thus disrupting mitochondrial function in tissues where energy demand peaks during pollen development (Lewings, 1993). To find clues for the mechanism of ‘447’ CMS, ‘ORF A’ (suspected to be the CMS determinant) was analysed for sequence motifs that could be indicative of a specific function. The most prominent feature of this tentative polypeptide was the presence of a fumarate lyase signature spanning residues 492–501 (GSTNMFSK), which matches the short conserved sequence (GSx2Mx2KxN) centred around an essential methionine residue probably involved in catalytic activity (Bairoch & Bucher, 1994). In addition, a putative arabino-galactan protein motif SPPAP (C. Reuzeau, personal communication) was found at the N terminus (residues 13–17) of ‘ORF A’ (see Fig. 2B), suggesting that ‘ORF A’, or a derivative thereof, can be glycosylated. In this respect, it is noteworthy that most of the antigenicity of the dsRNA-containing cytoplasmic vesicles (Dulieu et al., 1988) corresponds to epitopes that contain sugars (Desvoyes & Dulieu, 1996). No further significant or extensive sequence homology could be detected between ‘ORF A’ and any other proteins of virus or other origin, strengthening the contention that the ‘ORF A’ region is a specific hallmark of this dsRNA.

‘ORF B’ encodes the replication functions

Assuming that the positive-strand is interrupted by a simple nick and that the larger subgenomic RNA actually represents a functional messenger, the 5′-proximal initiation codon for ORF B would be the AUG at nt position 124–126 from that nick. No significant ORFs were found in any of the other phases, nor was there any other AUG codon found outside the conserved sequence (GSx2Mx2KxN) centred around an essential methionine residue probably involved in catalytic activity (Bairoch & Bucher, 1994). In addition, a putative arabino-galactan protein motif SPPAP (C. Reuzeau, personal communication) was found at the N terminus (residues 13–17) of ‘ORF A’ (see Fig. 2B), suggesting that ‘ORF A’, or a derivative thereof, can be glycosylated. In this respect, it is noteworthy that most of the antigenicity of the dsRNA-containing cytoplasmic vesicles (Dulieu et al., 1988) corresponds to epitopes that contain sugars (Desvoyes & Dulieu, 1996). No further significant or extensive sequence homology could be detected between ‘ORF A’ and any other proteins of virus or other origin, strengthening the contention that the ‘ORF A’ region is a specific hallmark of this dsRNA.

Helicase domain:

**VPdsRNA:** 1949

**RDR domain:**

**VPdsRNA:** 2023

**TRVL**

**RDR:** 1566

Fig. 3. Alignment of the helicase and polymerase regions of the polypeptide encoded by the *V. faba* ‘447’ dsRNA (VFdsRNA) and by the rice dsRNA (RDR). The highest degree of sequence conservation is found around the invariant amino acid residues of helicase motifs I (GXGKS) and II (DE), and around the RDR motifs V (TGQXXTXXXN) and VI (GDD) indicated in bold.

The sequencing of the dsRNA associated with hypovirulence in *Cryphonectria* (Endothia) *parasitica* (Shapira et al., 1991) and of the symptomless rice dsRNA (Moriyama et al., 1995) has established without ambiguity that such high molecular mass dsRNAs, which are associated with a specific polymerase and are contained within membranous vesicles (Lefebvre et al., 1990; Fahima et al., 1993), have a virus-like genetic organization and synthesize polypeptide precursors that are subsequently processed into functional polypeptides by the dsRNA-encoded protease function(s). In the case of *Cryphonectria* hypovirus 1-713 (CHV1), co-translational processing has been demonstrated to occur for both ORF A and ORF B (Shapira & Nuss, 1991) and mutational analysis has further confirmed the existence of two virus-encoded papain-like proteases (Choi et al., 1991b; Shapira & Nuss, 1991).

Like ORF B of CHV1 and the single ORF of the rice dsRNA (Fukuhara et al., 1993), ORF B of the ‘447’ dsRNA contained the conserved domains typical of RNA-helicase and RDRP functions (Koonin et al., 1991) that are required for the replication and transcription of these virus-like genetic elements. Computer-assisted sequence comparisons revealed that the ‘ORF B’ region of the ‘447’ dsRNA was indeed distantly related to ORF B of CHV1, but a much stronger relationship was found with the single ORF encoded by the dsRNA from japonica rice (Moriyama et al., 1995; Fig. 3).

The matrix comparison of these two proteins (not shown) revealed three regions of high homology. The highest incidence and length of clusters of identical amino acids was found around the conserved motifs of the polymerase region located towards the C terminus of ORF B (aa 5250–5650, with 41% amino acid identity and 58% similarities with only three
gaps introduced), followed by the motifs in the helicase region (aa 1950–1996 with 47% amino acid identity and 56% similarities, and aa 2023–2043 with 35% amino acid identity and 56% similarities), with a BLASTP score of 392. A more detailed analysis of the phylogeny of the dsRNA replicons and of their classification based on these sequence similarities will be published elsewhere (M. J. Gibbs, R. Koga, H. Moriyama, P. Pfeiffer & T. Fukuhara, unpublished data), but it is interesting to note that the next closest protein in the databases was the replicase of hepatitis A virus, with a BLASTP score of 118. In this virus, however, the helicase region precedes immediately the RDRP region, while these are separated by some 3000 aa residues in the ‘447’ dsRNA.

In addition to the helicase and RDRP motifs which we anticipated to find in the polyprotein, searches using the PROSITE base identified a potential β1–3 glucanase signature (VLMGLENEWL) where the underlined conserved glutamic acid is an active site residue in glycosyl hydrolases (Py et al., 1991). It remains to be established, however, whether any specific enzymatic activities correspond to the sequence homologies detected.

Discussion

High molecular mass dsRNAs have been found in several non-virus-inoculated plant species where they have long remained undetected because of the lack of obvious symptoms. Their presence is generally restricted to a single line or cultivar in which they are transmitted efficiently in a vertical manner through both egg and pollen. Zabalgogeazcoa et al. (1993), for example, traced the presence of a 13-2 kb dsRNA in the ‘Barsoy’ cultivar of barley to an ancestral progenitor imported from Japan at the beginning of this century, and retrieved this dsRNA from the various intermediate crosses performed. The presence of the dsRNA seemed therefore to confer some desirable agronomic trait(s) during the selection scheme rather than cause adverse effects.

Such unique, vertically transmitted 12–14 kb linear dsRNAs are also present in specific cultivars of pepper (Valverde et al., 1990), in the Black Turtle Soup bean (Wakarchuk & Hamilton, 1985) and in cultivated rice Oryza sativa ssp. japonica (Wang et al., 1990). In the latter two species, their presence was thought to be associated with CMS, but this hypothesis was later disproved (Mackenzie et al., 1988; Fukuhara et al., 1993). In addition, elucidation of the molecular mechanisms of CMS (for a review, see Lewings, 1993) subsequently confirmed that this phenotype does not result from a virus infection, but rather is associated with mitochondrial DNA rearrangements that lead to the production of chimaeric, non-functional polypeptides.

The dsRNA found in the ‘447’ male-sterile line of V. faba is the only apparent exception to this mechanism, since male-sterile and restored F1 plants have indistinguishable mitochondrial DNA patterns (Turpen et al., 1988). By contrast, plants that have spontaneously reverted to fertility no longer contain the dsRNA, and both the dsRNA and the male-sterility trait are lost irreversibly in the progeny obtained after fertilization of the male-sterile line with restorer pollen. Determination of the entire sequence of this dsRNA has now demonstrated that it belongs to a family of RNA replicons encoding long polypeptides that are subsequently processed into functional peptides by an endogenous protease, a strategy used by many RNA viruses (Spall et al., 1997).

The size, genome organization and sequence similarities found in the helicase, polymerase and protease domains of CHV1-713, the type member of Hypoviridae (Hillman et al., 1995), led Koonin et al. (1991) to suggest that this dsRNA replicon shares a common ancestry with potyviruses. The rice dsRNA and the ‘447’ dsRNA differ, however, from CHV1 in the order of the helicase and polymerase domain, and in the absence of a long non-coding sequence and poly(A) tail; in addition, they differ in possessing a single ORF spanning the entire length of the dsRNA instead of two ORFs being read in a stop/restart mode in CHV1 and related hypoviruses. Such differences, together with the presence of a nick in the coding strand, suggest that these two dsRNAs may be classified into a family other than Hypoviridae. They are at any rate more closely related to positive-stranded than to dsRNA viruses, and computer-assisted analysis showed a clear-cut affinity of the helicase and RDRP regions with hepatitis E virus, a virus that also encodes a papain-like cysteine protease. Due to their low degree of conservation, protease functions are difficult to identify in viruses (Koonin et al., 1992; Rozanov et al., 1995) and are generally inferred from functional studies and site-directed mutagenesis that allow assignment of catalytic residues and target sequences. Preliminary analysis of the long ORF encoded by the ‘447’ dsRNA has identified an ILHAY sequence (residues 572–576) similar to the IVHAY sequence of the X domain (Koonin et al., 1992) found upstream of the papain-like protease domain of Ross River virus, a member of the ‘alpha-like’ supergroup; the invariant proline located two residues downstream in the X domain was, however, missing in the polypeptide encoded by the ‘447’ dsRNA and in vitro proteolysis studies will therefore be required to detect and identify eventual protease activity associated with the expression of the ‘447’ dsRNA genome.

Unencapsidated dsRNA replicons have a plasmid-like lifecycle; their copy number is controlled by the host but can vary with the differentiation and developmental stage. The rice dsRNA, for instance, is kept at a constant concentration estimated at 100 copies per cell, but this concentration increases 10-fold in suspension cultures and 100-fold in pollen without affecting its viability (H. Moriyama, personal communication). Likewise, the dsRNA content of the ‘447’ line of V. faba was found to depend on the nuclear background (G. Duc, personal communication).

Sequence comparisons have revealed that the ‘447’ dsRNA is most closely related to the rice dsRNA, with stretches of
protein sequence being highly conserved around the helicase and polymerase domains. An additional, intriguing feature specific to these two RNAs is the presence of a nick in the positive-strand, interrupting the long ORF they encode (Fukuhara et al., 1995; and this work). Although it is not known at present how this interruption is generated and whether it is relevant to the translation of these dsRNAs, it is tempting to speculate that this could constitute an alternative means of expressing the genetic information in two separate blocks instead of using the stop/restart arrangement of the hypovirulence-associated virus CHV1, or a bipartite genome as in the bymoviruses. In this respect, it is striking that the most conserved part of these RNA replicons or viruses corresponds to the domain of the polyprotein endowed with the functions related to RNA replication and proteolytic self-maturation, while ORF A or its equivalent is much more variable and probably reflects the acquisition of specific properties, such as transmission by specialized vectors or phenotypic effects on their host. Fig. 4 summarizes the similarities in the genetic organization of these dsRNA replicons compared to that of potyviruses, bymoviruses and alpha-like viruses.

It is not known at present whether the mRNA(s) derived from the ‘447’ dsRNA feature a 5’-terminal structure such as a cap or a VPg, and, whether a ‘leaky scanning’ translation strategy may be used as in potyviruses (Simón-Buela et al., 1997). Similarly, a polyprotein self-processing strategy can only be inferred from comparisons with CHV1, potyviruses and alpha-like viruses where this mechanism has been investigated in detail. Finally, it is hoped that transformation with ‘ORF A’ or with specific domains of ORF A+B will result in male-sterile plants and will allow a better understanding of the mechanism by which the ‘447’ dsRNA induces male sterility. Indeed, several signatures – fumarate lyase, thioredoxin and β1–3 glucanase – have been identified in the polyprotein(s) and suggest various potential mechanisms can cause male sterility. In petunia for instance, male sterility has been attributed to mis-timing of callose wall degradation by callase, a β1–3 glucanase normally expressed at a specific stage in the anther locule (Izhar & Frankel, 1971). This situation could be reproduced experimentally in transgenic tobacco (Worall et al., 1992) expressing a modified β1–3 glucanase, and in which premature breakdown of the callose deposits resulted in male sterility. It is tempting to speculate that a similar mechanism may operate in the ‘447’ CMS, and expression studies should help resolve this point.

I wish to thank Philippe Hammann for excellent assistance with automated sequencing, Dr Rodrigo Valverde for help with some of the experiments and useful discussions, Drs Hiromitsu Moriyama, Christophe Reuzeau and Gérard Duc for communication of unpublished results, and Drs Vivek Prasad and Ken Richards for help with the manuscript.

References


CMS-associated dsRNA in *V. faba*


Received 18 February 1998; Accepted 17 June 1998