FINAL REPORT

Defra Project CPEC 9

Peter Palukaitis & Julie Squires

Scottish Crop Research Institute

26 Aug 2007

Contents

Glossary

Cauliflower mosaic virus (CaMV) – a DNA virus with an 8000 basepair circular genome encoding six genes (designated gene I to gene VI) and containing two promoters of RNA transcription: the 19S promoter and the 35S promoter.

CaMV 19S RNA promoter – a sequence in CaMV that acts as the binding of DNA-dependent RNA polymerase II and promotes transcription of an RNA species designated as 19S RNA, based on its sedimentation coefficient (relative mobility) in a sucrose gradient. The 19S RNA allows expression of gene VI of CaMV.

CaMV 35S RNA promoter – a sequence in CaMV that acts as the binding of DNA-dependent RNA polymerase II and promotes transcription of an RNA species designated as 35S RNA, based on its sedimentation coefficient (relative mobility) in a sucrose gradient. The 35S RNA allows the expression of CaMV genes I to V. The 35S RNA also is used as a template for reverse transcription to DNA during the replication of CaMV.

Coat protein – one or more proteins encoded by a virus, which assemble together with the virus genetic material to form virus particles (also known as virions) as a means of protecting the genetic material against the environment and allowing the virus to move through the plant and from plant to plant. Coat protein is also called capsid protein.

Cucumber mosaic virus (CMV) – an RNA virus with an 8700 nucleotide genome consisting of three single-stranded RNA molecules encoding five genes. CMV is unrelated to CaMV, but has a very broad host range and is present in many vegetable crops and weeds in the UK.

Encapsidation – the process by which virus genetic material (DNA for CaMV) is packaged by molecules of the coat protein of a virus to form virus particles.

Episome – a genetic element that can multiply independently of the chromosomal DNA, such as a virus or a plasmid.

Episomal replication – multiplication of a virus or a plasmid inside a living cell, independent of the multiplication of the chromosomal DNA.

Germline – those cells that are in the line of cells (germ cells) that will be passed on to the progeny. In plants, the germline cells are in the flowers and pollen and will form the seeds.

Hemizygous – GM plants generated through transformation have integration at one or more sites in the plant genome. Thus, there is no corresponding gene located at the same site of the sister chromosome and hence such plants are genetically referred to as hemizygous, rather than homozygous (the same gene located at the corresponding site in both sister chromosomes) or heterozygous (two variants of the same gene located at the corresponding sites of sister chromosomes).

Integration – the insertion of foreign sequences into the chromosomal DNA of an organism.

Promoters – nucleic acid sequences that promote transcription of RNA by acting as binding sites for enzymes that synthesise RNA (so-called polymerases or transcriptases).

Reverse transcription – the copying of an RNA template into DNA by an enzyme designated reverse transcriptase. Gene V of CaMV encodes a reverse transcriptase.

Somatic – cells or tissues that normally do not contribute to progeny plants, except when plants are generated by making cuttings (clones) or are grown in tissue culture as explants to form de-differentiated tissue that after treatment with hormones will grow into new plants.

T-DNA – Transferred DNA, the single-stranded DNA that is transferred from Agrobacterium into the plant nucleus and into the plant genome. The T-DNA is bordered by two, 25-basepair, imperfect, direct repeat sequences referred to as the right and left borders.

Transcription – the copying of a stable DNA template into the more labile RNA. This is the manner in which genes are expressed, with some RNAs (messenger RNAs or mRNAs) being translated to produce the encoded protein products.

Transmission – transfer of an organism or a trait. Transmission can be horizontal, such as virus transmission from plant to plant in a field, or vertical, such as from parent to progeny, through seed or pollen.

Virulence – infection caused by a pathogen becoming established in a susceptible host. A pathogen causing strong symptoms is referred to as either severe or highly or strongly virulent, while one causing only mild symptoms is referred to as either mild or poorly virulent.

Executive Summary

The 35S RNA promoter of cauliflower mosaic virus (CaMV) has been used to facilitate gene expression in genetically modified (GM) plants. This promoter was the most active promoter available for dicotyledonous plants. The promoter has been used in many GM crop species released under licence in numerous countries. Although there are no reports to suggest that the promoter itself has posed a problem to crops, the environment, or human health, the use of a plant virus promoter has proven to be controversial within the European Union. The potential ability of the CaMV 35S RNA promoter to act as a hotspot for recombination with infectious CaMV, leading to virus integration and/or promoter inactivation, has been raised as an environmental risk issue. Questions also have been raised concerning the potential problems caused by CaMV infection silencing the activity of the CaMV 35S RNA promoter-driven transgenes, which was demonstrated for some strains of CaMV infecting oilseed rape plants containing a CaMV 35S RNA promoter [Al-Kaff *et al*., 1998; 2000]. Therefore, this project was undertaken to respond to those specific concerns and to determine experimentally to what extent integration of the CaMV genome occurred during natural virus infection of non-GM plant species as well as GM plants of the same species expressing genes controlled by a CaMV 35S RNA promoter. In addition the project sought to examine what the consequence of such an integration event would be for those plants. Finally, the effects of CaMV infection also were assessed on a CaMV 35S RNA promoter-driven transgene with regard to gene expression and gene stability, in three host and one non-host species of CaMV, as well as in the progeny of these plants.

Whether integration of the CaMV genome into non-GM plants occurred was examined in this study for four plants species: *Arabidopsis thaliana* (thale cress), *Brassica napus* (oilseed rape), *Nicotiana benthamiana* (an experimental host species), and *Nicotiana tabacum* (tobacco). The first three plants are hosts for CaMV, while the last is not, but GM tobacco has been used in many field trial experiments abroad, as well as in glasshouse trials in the EU.

An adaptor-linker primer and nested polymerase chain reaction (PCR) assay of the plant DNA was used to detect any evidence of integration of CaMV into the genomes of these plants and the progeny of plants infected by CaMV under laboratory conditions. A sensitivity limit of the equivalent of one cell containing integrated CaMV sequences in 2000 cells not containing integrated sequences (or the equivalent of one copy per 4000 haploid genome copies) was set using this assay.

In non-GM plants, the results of the assay showed only one sample with integration of small pieces of the CaMV genome in 780 non-GM plant samples tested. This was detected in one out of 160 infected leaves of *A. thaliana* that were sampled, with none detected in 560 samples taken from the infected plants of the other three species. Integration was not observed in 60 samples taken from the progeny plants of the various infected non-GM plants. The integration events occurring into the genome of *A. thaliana* appeared to be due to homologous recombination between the virus and the plant DNA. These occurred in regions of the plant genome containing little sequence similarity with the virus [5 or 10 nucleotides (nt)], rather than in any of the numerous regions of the *A. thaliana* genome where there are higher levels (18-25 nt) of contiguous sequence identity with the CaMV genome. Therefore, it is concluded that integration of part of the CaMV genome can occur in a non-GM *A. thaliana* test plant under laboratory conditions. It occurs at an extremely low frequency and does not occur in the progeny of CaMV – infected *A. thaliana*. The

actual frequency of integration cannot be estimated, since only a single sample showed integration, and in only one of the four plant species examined. Nevertheless, if integration occurred in other non-GM plant genomes, then it would have been at a level below the detection limit set here (the equivalent of one in 4000 haploid genomes).

Whether integration of the CaMV genome also occurred in GM plants and their progeny was examined for the same four plants species, each expressing the green fluorescent protein (GFP) driven by the CaMV 35S RNA promoter (35S-GFP). In this case, the infected plant samples and their progeny seedlings were examined for integration into the 35S RNA promoter. No integration of CaMV could be detected into the 35S RNA promoter sequence of 3856 samples tested. Therefore, it is concluded that despite the several hundred contiguous nucleotides in common between the CaMV genome and the 35S RNA promoter present in the plant genome, recombination between CaMV and the 35S RNA promoter leading to integration of the viral genome is either an extremely rare event that cannot be measured, or it does not occur.

GM plants of the above four species expressing the 35S-GFP transgene that were infected by CaMV were examined for the effects of infection on the expression and the stability of the transgene. Generalised silencing of transgene expression was not observed in any of the 382 plants infected with CaMV, nor were there any effects on transgene expression in 10 seedlings obtained from each of these infected plants. Therefore, it is concluded that infection by CaMV does not affect the stability of the transgene in subsequent generations, and that previous reports of silencing of transgenes driven by CaMV 35S RNA promoters, after infection by CaMV, were both virus strain and host species dependent. In comparison with the previous investigations of this phenomenon [Al-Kaff *et al*., 1998; 2000; Dale & Al-Kaff, 2006], it is apparent that the silencing of transgenes in GM plants following infection by CaMV occurs in only limited and defined circumstances.

To test for potential effects of CaMV integration, should it occur, on the growth and fecundity of plants as well as on the stability and expression of a transgene, *A. thaliana* plants were transformed with the entire CaMV genome (producing CaMV-GM plants) and these plants were crossed with 35S-GFP-GM plants. These doubly transgenic CaMV-GFP-GM plants did not exhibit any differences in the expression of the GFP transgene, nor did the progeny seedlings show any effects on the expression of the transgene. Therefore, genetically-engineered viral integration into the plant genome in itself did not lead to effects on the expression or stability of a transgene. The singly transgenic CaMV-GM plants showed no significant differences from non-GM plants in plant or root biomass, and only a variable and slight effect on plant height and days to flowering. This could have been due to the transformation event itself, since the plants transformed with an empty vector lacking the CaMV genome had similar effects. Similar variable effects on the number of seed set, and the percent seed germination also were observed between transformed lines. Therefore, overall, the effects observed were due mostly to either the engineered integration of CaMV (or the empty vector), or the location of insertion, and not due to the presence of the CaMV genome per se. Similarly, a microarray analysis of gene expression in these CaMV-GM plants indicated that the integration of the T-DNA had a greater effect on gene expression than the presence of the integrated CaMV genome. By contrast, infection by CaMV had a much greater effect on plant gene expression than integration of either the CaMV genome or an empty vector, as observed previously for other viruses.

In the literature, several cases of episomal replication of virus genomes have been found, although horizontal transmission does not occur for most of these viruses. Therefore, the ability of the CaMV genome in the CaMV-GM plants to transcribe CaMV genes and recombine to excise and produce a fully replicating virus outside the nucleus (episomal replication) was assessed. Most of these GM plants could express CaMV gene VI, as was expected. However, episomal replication also was observed to occur in many of the engineered plants tested, whether or not they were subjected to various environmental stresses. This was shown by the expression of the CaMV gene IV product, the viral coat protein, as well as by the PCR for the replicating virus and electron microscopy visualisation of virus particles formed in such plants. The episomal viral genomes all showed alterations in the non-essential gene II, the site of integration into the plant genome, and in some cases, in the essential gene III. These experiments show that in plants genetically engineered to contain one copy of the entire CaMV genome, CaMV can excise itself from the plant nuclear genome but only with modification to the viral sequence borders at which integration and excision occurred, and therefore the recombinant virus will, in most cases, be debilitated. However, since integration of a fragment of the CaMV genome occurred at a very low frequency in non-GM infected plants in this study and does not occur in the germline, the ability of integrated CaMV to excise its genome from the plant genome and again replicate episomally has little consequence. This is because the plants are already infected by CaMV.

In conclusion, integration of CaMV sequences can occur in non-GM plants, but at such a low level that the consequences for human health and the environment are minimal. Moreover, since the virus does not invade germline tissue, it is unable to integrate into such tissues and is therefore unable to pass integrated virus on as a new trait. The presence of the CaMV 35S RNA promoter being used in GM plants has not been found to facilitate homologous integration of the CaMV genome or CaMV sequences in the 35S RNA promoter, despite the extensive sequence identity between the virus and the promoter. Moreover, the CaMV 35S RNA promoter being used in GM plants does not facilitate the seed transmission of CaMV into progeny seedlings.

Scientific Report

The Insertion of Cauliflower Mosaic Virus into Host Genomes During Natural Viral Infections

General Introduction

Cauliflower mosaic virus (CaMV) is a plant virus that infects a broad range of cruciferous (brassica) plant species, such as cabbage, cauliflower, oilseed rape and turnips. CaMV is transmitted in nature by more than 27 species of aphids. Experimentally, only a few strains of CaMV can infect a limited selection of solanaceous plants such as *Nicotiana benthamiana*, *N. bigelovii*, *N. clevelandii*, and *N. edwardsonii*. Therefore, this virus is not a pathogen of solanaceous crops such as tobacco or tomato.

CaMV has a circular, double-stranded DNA genome containing six genes, designated genes I to VI. [See Fig. 1.] The genes of CaMV encode proteins that are important for the natural infection cycle of the virus. There include an enzyme called reverse-transcriptase, which is required for the replication of the virus and is encoded by gene V, as well as proteins required for the packaging (encapsidation) of the viral DNA (gene IV), movement of the virus from one cell to another (genes I and III), transmission of the virus by aphids from one plant to another (genes II and III) and for expression of genes I-V from the 35S RNA plus function in the formation of a matrix in the cytoplasm of infected cells to facilitate CaMV replication (gene VI).

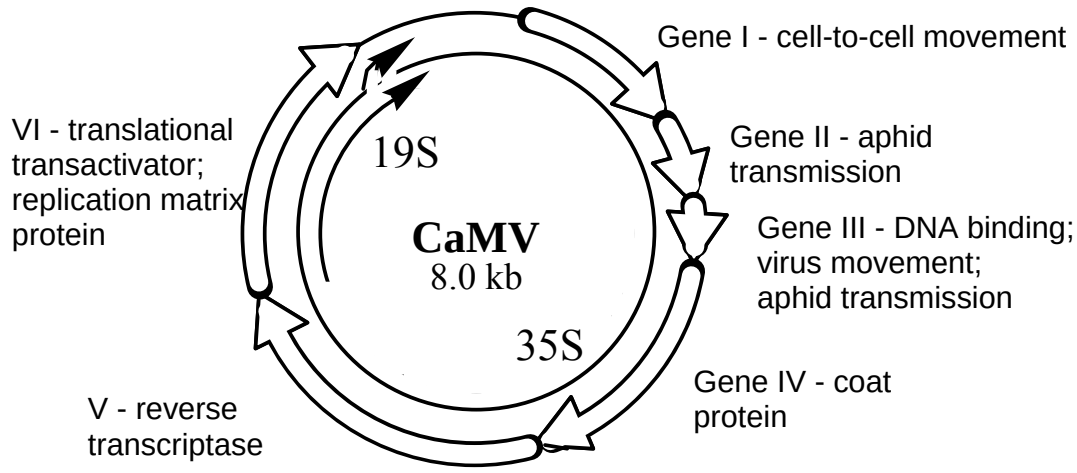

Fig. 1. Genome organization and expression strategy of CaMV. The circular DNA 8000 basepair (8 kb) genome of CaMV showing the six characterised genes of CaMV (genes I to VI) along with the major functions of the encoded proteins. In addition, two mRNAs are transcribed from the CaMV genome: the 19S RNA encoding the gene VI product, which is needed together with the 35S RNA to allow expression of genes I to V from the 35S RNA. The 35S RNA also is the template for the reverse transcriptase during the replication of CaMV.

After CaMV enters a cell, the virus particle is disassembled and the viral DNA is replicated. [See Fig. 2]. Virus replication involves several steps: First, the DNA enters the nucleus, where the nicks in the viral DNA are joined creating a supercoiled DNA. This DNA is then used as a template for RNA transcription by the plant enzyme normally involved in synthesising mRNA, DNA-dependent RNA polymerase II. This enzyme binds to the 35S RNA promoter resulting in synthesis of the 35S RNA, and also to the 19S RNA promoter for synthesis of the 19S RNA. Both viral RNAs are exported from the nucleus into the cytoplasm. The 19S RNA is translated to produce the gene VI product which forms a matrix in the cytoplasm that acts as a nest for replication of CaMV from the 35S RNA. The 35S RNA is translated, with the help of the gene VI product, to produce the other CaMV gene products including the reverse transcriptase. The reverse transcriptase then copies the 35S RNA into DNA, which is then packaged into CaMV particles by the coat protein of CaMV. [See Fig. 2.]

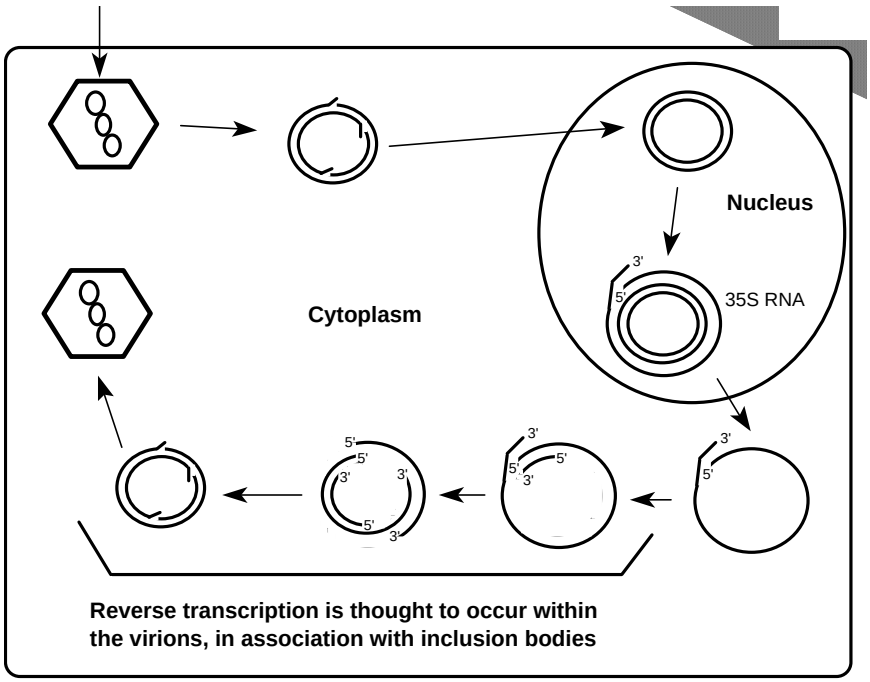

Fig. 2. The replication cycle of CaMV. Virus particles are disassembled in the cytoplasm, the viral DNA enters the nucleus where the 35S RNA is synthesised and then exported back into the cytoplasm. Replication of CaMV and encapsidation of the viral DNA into new virus particles occurs in the cytoplasm, in a matrix formed from the gene VI product.

The CaMV DNA sequences containing the 35S RNA promoter strongly promote RNA transcription in plants and therefore this promoter has been used to facilitate high levels of gene expression in genetically-engineered transgenic plants. While promoters of plant origin that show either tissue-specific or developmental stage-specific regulation of gene expression are being tested, most of the genetically modified (GM) plants generated to date have used the CaMV 35S RNA promoter, predominantly to obtain high levels of gene expression in a wide variety of plant tissues. Therefore, the vast majority of the GM plants that are grown agriculturally in a number of countries contain the CaMV 35S RNA promoter.

As summarised below, there has been some concern expressed about the safety of using the CaMV promoter. Experimental assessment of these environmental and genetic issues is the basis of this research project.

Scope of the Project

The issues of whether CaMV inserts in to plant genomes that are GM to express genes under the control of the CaMV 35S RNA promoter, the consequences of such insertion events for the use of GM plants and how this compares to the situation in non-GM plants that are infected with CaMV have not been addressed to date by experimental analysis. The approach described below investigated whether insertion of CaMV DNA into the host plant genome occurs and the effects of insertion on the stability and expression of transgenes in GM plants containing the CaMV 35S promoter.

We describe a series of [12] objectives to fulfil the Aims of the project. These objectives were either specific objectives of the Aims, or the development of tools or systems required to fulfil the objectives. The objective numbers are given in brackets next to the Aims written below.

Aims

To conduct a literature review on the integration of plant viral sequences into plant genomes and their effects on gene expression [Objective 1].

To conduct experimental research on the insertion of CaMV DNA into a natural host and the effect on the stability and expression of transgenes in the same species of plant genetically engineered to contain the CaMV 35S RNA promoter sequence. Specifically:

(i) To determine whether and to what extent CaMV DNA integrates into wild type and GM host plant genomes as a consequence of natural viral infections." [Objectives 2-4, 6 & 7]

(ii) If natural infection results in DNA integration into wild type and GM host plants, to investigate the propensity to:

- affect expression and stability compared to viral DNA that is not integrated, [Objectives 8 & 10]
- destabilise the resulting genome, e.g. activate/inactivate a vital physiological function, [Objective 9]
- cause episomal virus disease under certain environmental conditions; i.e., are integrated DNA sequences replicated? (can the sequences be transcribed and reverse transcribed) and if so, does episomal disease affect transgene expression? [Objectives 11 & 12]

(iii) To investigate whether *in vitro* insertion of CaMV DNA sequences into plant genomes using genetic modification promotes or reduces the chance of integration of viral DNA during natural infections. [Objectives 4 & 5].

Objective 1. Literature review on integration and the effects on gene expression.

A literature review was done to establish the current knowledge base concerning the integration of CaMV DNA into plant genomes, including literature on the presence of caulimo-like virus (other presumptive pararetroviruses related in genome organisation to CaMV) sequences present in plant genomes and the extent to which transcription, reverse transcription and episomal replication of these sequences occur. An interrogation of the genomic sequence of *Arabidopsis thaliana* (thale cress), a pasture weed, as well as the *Brassica* DNA database, known hosts of CaMV, was done to establish whether CaMV sequences are present already in the genome of these plants. Finally, the extent to which it is known if CaMV infection can inhibit the expression of transgenes regulated by the CaMV 35S RNA promoter was assessed.

Summary

The issue of whether the use of sequences derived from CaMV to drive the expression of transgenes in GM plants poses a risk to the environment, following natural infection of such GM plants with CaMV, is assessed. The hypothetical risks associated with CaMV infection of such genetically modified plants include increasing the frequency of integration of CaMV into plant genomes, with hypothetical subsequent vertical transmission (i.e., through seed) of the CaMV sequences, subsequent episomal replication of the CaMV genome under stress conditions, and alteration in expression or the stability of the transgene. The literature concerning the extent to which plant viral sequences integrate into and are expressed episomally from plant genomes is reviewed. Such natural integration has been shown to have occurred in progenitor lines of some plant species, in the case of a few viruses with DNA genomes, but not to date with CaMV. The assessment of whether CaMV sequences are already present in the genome of the experimental host plant species *Arabidopsis thaliana* was made by interrogating this plant genome sequence. Large regions of CaMV sequences were not present in either the *A. thaliana* genome, or in the partial genome sequences of various other host *Brassica* species. However, short regions of 18-25 nt were observed to occur commonly. The potential consequences of integration of CaMV sequences were evaluated from three published studies involving genetic manipulation to insert CaMV genomes into plant genomes. These data showed that limited CaMV gene expression could occur in such plants, but episomal expression did not occur when only a single copy of the CaMV genome was integrated in such as way as to split the reverse transcriptase gene, while integration of more than one copy led to recombination and episomal replication. The effects on suppression of transgene expression following CaMV infection of GM plants expressing transgenes driven by the CaMV 35S RNA promoter also were evaluated based on data available from the literature. Three studies showed that 35S RNA promoter-mediated expression could be suppressed in one host species, itself showing recovery from CaMV infection. However, whether or not suppression occurred was dependent on both the strain of CaMV and the specific host species.

Introduction

The issue of whether transgenic or GM plants expressing genes regulated by the 35S RNA promoter sequence derived from the virus CaMV pose unique risks has been raised (Ho *et al*., 1999), refuted (Hull *et al*., 2000a; Morel and Tepfer, 2000) and responded to with renewed criticism (Ho *et al.*, 2000). These debates have led to additional experiments to determine whether the CaMV 35S RNA promoter can function in mammalian cells (Vlasák *et al.*, 2003; Tepfer *et al.*, 2004), as had previously been shown to be the case for promoting gene expression in bacteria (Assaad and Signer, 1990), yeast (Pobjecky *et al*., 1990; Rüth et al., 1992; 1994), fungi (Sun *et al.*, 2002) or toad eggs (Ballas *et al.*, 1989). However, there is still no evidence supporting the speculative chain of events envisioned by Ho *et al*. (1999) for the risks to human or animal health by using the CaMV 35S RNA promoter in GM plants. Moreover, questions concerning whether CaMV (or other plant pararetroviruses) can integrate into plant genomes, whether the presence of the CaMV 35S RNA promoter in a GM plant could increase the frequency of CaMV integration and if it occurs, whether such integrated sequences could replicate episomally, and what effect infecting or integrated CaMV will have on the expression and stability of transgenes regulated by the 35S RNA promoter have received limited direct attention, experimentally. Thus, Defra has commissioned this project to examine these issues.

This literature review provides an analysis of the current state of knowledge concerning the integration and expression of viral-like sequences in plant genomes and the effects of viral infection on the expression of transgenes containing viral-derived promoters. Over the last 10 years, a number of examples of viral sequences and viral-like sequences integrated into plant genomes have been identified (reviewed in detail by Harper *et al*., 2002). With one exception, these sequences are related to varying extents to known viruses with DNA genomes. The extent to which these DNA sequences can be expressed episomally, as well as conditions giving rise to such, will be described. In addition, assessments will be made of whether the sequences of one DNA virus (CaMV) are present in the complete genome of *Arabidopsis thaliana* and in partial sequences from various *Brassica* species. The effects of artificial integration of the CaMV genome into a plant genome on gene expression of CaMV or episomal expression of CaMV also will be reviewed. Finally, since the expression of most plant transgenes is regulated by a transcription promoter derived from CaMV, the extent to which infection by CaMV affects transgene expression has been analysed in several instances. These data also will be evaluated.

Approximately 20 % of almost 1,000 known plant viruses contain DNA genomes (Lovisolo *et al*., 2003) and are grouped into three taxonomic families. *Geminiviridae*, *Circoviridae* and *Caulimoviridae*. The *Geminiviridae* and the *Circoviridae* replicate in the nucleus, through DNA-DNA intermediates, whereas viruses in the *Caulimoviridae* are defined by the presence of a DNA genome that is replicated by reverse transcription of an RNA intermediate (Hull *et al*., 2000b), a property shared with viruses of vertebrates in the family *Hepadnaviridae*. The caulimoviridae and hepadnaviridae families constitute the taxonomic pararetrovirus supergroup. Animal retroviruses contain an RNA genome and a DNA intermediate with a requirement for integration into the host genome for replication. However, pararetroviruses do not integrate into the host genome as part of their normal replication cycle, but rather replicate episomally. Indeed, until recently, it was believed that plant viruses, unlike viruses of animals and bacteria, never integrated into the host genome. Plant genomes contain other retroelements called

retrotransposons, which have sequence features indicating replication via reverse transcription. These retroelements differ from pararetroviruses in that retroelements are transcribed from sequences integrated in the plant DNA. However, the boundary between pararetroviruses and these various retroelements has become less clear, since some of these various retrotransposons assemble into virus-like particles, and one, from insects, appears to be infectious. Since much of this work has been reviewed recently (Harper *et al*., 2002), we will focus here on what clearly are integrated viral-like sequences, and will not deal with retrotransposons.

Because the current generation of GM crops for the most part express transgenes using the CaMV 35S RNA promoter sequence, it is important to assess the effects of such viral sequences, vis-à-vis infection by the cognate virus and on transgene expression. This will allow a scientific evidence-based decision to be made concerning the safety or risks associated with the environmental release of such GM plants.

Integrated viral sequences in plant genomes

The great expansion of plant genome sequencing in recent years, together with studies aimed directly at the identification of viral sequence in plant host genomes, has led to the discovery of a number of stably integrated plant viral sequences in plant genomes. Despite the fact that the majority of plant viruses have RNA genomes, in all but one case the integrated sequences are derived from DNA viruses. The sole example of integration of a plant viral RNA sequence is the observation that sequences of the coat protein gene of potato virus Y (PVY) were found integrated into the genomes of several grapevine cultivars. Similarly, such sequences, although in a rearranged form, also were found in the tobacco genome (Tanne and Sela, 2005).

The first report of viral sequence in plant genomes was of geminivirus-related DNA in tobacco (Bejarano *et al*., 1996), which was found subsequently to be present in four closely related *Nicotiana* spp. (Ashby *et al*., 1997). Only a portion of the viral genome is present and there is no evidence of transcription of the sequences or of viral infection resulting from them.

No examples of sequences derived from nanoviruses (in the family *Circoviridae*) have been described yet as present in any plant genomes. In contrast, three members of the family *Caulimoviridae* have integrated forms that are complete and can be activated by stresses to give episomal infections: banana streak virus (BSV), tobacco vein clearing virus (TVCV) and petunia vein clearing virus (PVCV). These are derived taxonomically from three of the six genera (*Badnavirus*, Cassava vein mosaic virus-like, and Petunia vein clearing virus-like, respectively) in the family *Caulimoviridae*. There are no examples of sequences derived from CaMV identified as being present in plant genomes.

Banana streak virus

The causal agent of viral leaf streak in *Musa* spp., BSV is transmitted by mealybug vectors. The observation that BSV infection could arise in healthy plants during tissue culture led to a search for integrated sequences capable of giving rise to episomal viral infection (Harper *et al*., 1999; Ndowoara *et al.*, 1999). Two types of integrants have been found: partial BSV sequences that appear unlikely to contribute to episomal infection and a second class that consists of multiple, tandemly arranged copies of BSV sequence that could yield the viral genome via homologous

recombination. Other work also implicates the process of genetic hybridisation of the host in triggering episomal expression of the BSV integrants and identifies host genetic markers linked to disease expression in hybrids (Lheureux *et al*., 2003). Considerable variation in the sequences of such integrated badnavirus sequences was observed in different *Musa* spp. These indicated that integration may have occurred several times and that the sequences have evolved differently, but have retained coding potential, suggesting that such sequences may give rise to episomally replicating badnaviruses (Geering *et al*., 2005a,b). It is not known if the virus produced in such plants can be transmitted horizontally in nature, although mealybug transmission was demonstrated experimentally (Geering *et al*., 2005b).

Tobacco vein clearing virus

TVCV occurs only in the hybrid species *Nicotiana edwardsonii* and is transmitted only vertically, through seed to progeny plants. As for BSV, TVCV sequences have been found to be integrated in the host genome (Lockhart *et al*., 2000). The mechanism by which episomal infection is generated is unclear, but may involve recombination between separate, partial TVCV integrants to form functional full-length genome. As in the case of BSV, a hybrid host genome appears to be required for episomal virus replication and disease development. The TVCV sequences are present in the *N. glutinosa* parent of *N. edwardsonii*, but not in the *N. clevelandii* parent. Moreover, similar sequences are present also in the genomes of other *Nicotiana* spp. with origins in South America.

Endogenous pararetroviruses

Host genome sequencing also has revealed the presence of endogenous pararetroviral-like (EPRV-L) sequences in *Nicotiana tabacum*, similar in sequence (88%-94 %) to TVCV and integrated to such high copy numbers as to form a major constituent of the host genome (Jakowitsch *et al*., 1999). Further research has revealed the existence of distinct EPRV-L families in the parental *Nicotiana* spp. of tobacco [*N. sylvestris* (Jakowitsch *et al*., 1999) and *N. tomemtosiformis* (Gregor *et al*., 2004)]. EPRV-L sequences have been detected in other solanaceous plant species (Jakowitsch *et al*., 1999; Staginnus *et al*., 2007) and in species from a diverse range of plant phyla (reported in the thesis of C. N. Hansen, 2003; referred to in Lovisolo *et al*., 2003 and recently reviewed by Staginnus and Richer-Pöggeler, 2006). However, there is no evidence yet for episomal infections from these EPRV-L sequences.

Petunia vein clearing virus

PVCV is transmitted vertically, as well as by grafting, and its sequence is present in the petunia genome (Richert-Pöggler and Shepherd, 1997). However, it now appears that PVCV differs from BSV and TVCV in that its entire genome exists arranged in tandem arrays in the genome of *Petunia hybrida*, allowing direct release of virus by transcription (Richert-Pöggler *et al*., 2003). PVCV may therefore represent an intermediate, in terms of replication strategy, between pararetroviruses and true retroviruses. PVCV sequences are present to different extents in both of the parents of hybrid petunia plants, but are not expressed episomally in those species. Moreover, wounding and heat stress can induce episomal expression from hybrid petunia plants (Richert-Pöggler *et al*., 2003; Noreen *et al*., 2007).

In all of the above instances where episomal expression occurs in hybrid plants (petunia hybrids or *N. edwardsonii*), the viral sequences are already present in at least

one parent, but some factor is required from the other parent, as well as particular forms of stress to induce episomal replication. In the cases where episomal expression has not been found to occur, it is conceivable that the presence of these sequences in plant genomes acts to prevent infection by related viruses, by homology-dependent gene silencing (Mette *et al*., 2002).

Interrogation of sequence databases with the CaMV sequence

CaMV is among the most extensively studied plant viruses and in recent years has been in the spotlight due to the extensive use of its 35S RNA promoter in plant biotechnology. CaMV is the type member of the genus *Caulimovirus*, in the family *Caulimoviridae*. We have investigated the sequence databases to establish whether CaMV sequences are present in the completely sequenced genome of *Arabidopsis thaliana* or the partially sequenced genomes of other members of the genus *Brassica*, which are the natural hosts of most strains of CaMV, and consider the possible effects if CaMV should integrate into host genomes.

The complete genome of CaMV (Franck *et al*., 1980; Genbank accession number NC_001497) was used to query either the whole *A. thaliana* genome (BAC clones) or the "Brassica DNA" database, using the WU-BLAST2 programme (Altschul *et al*., 1990) at The Arabidopsis Information Resource (TAIR) webpage:
http://arabidopsis.org/wublast/index2.jsp
Further BLAST analysis was done using the non-redundant nucleotide database of the U.S. National Center for Biotechnology Information:
http://www.ncbi.nlm.nih.gov/BLAST/

The greatest similarity between CaMV sequence and the *A. thaliana* genome was to a region within the chromosome 4 BAC clone F7K2, with 56 % nucleotide identity over a 758-nt section (p=9.6e^{-8}). In this region, the longest contiguous sequences in common are 13 nt, with several other regions containing no more than 6 or 7 contiguous nucleotides in common. Sequence annotation of this region indicates it to be a Large Terminal Repeat (LTR) retrotransposon, while the matching CaMV sequence is a part of gene V, encoding the viral reverse transcriptase. It is therefore likely that the sequence similarity is due to shared function and it is unlikely to be evidence of past integration of CaMV sequence in the *A. thaliana* genome.

A second example of a region with similar percentage sequence identity, but encompassing a smaller region was observed with another BAC clone region within the chromosome 1 BAC clone T7P1, with 56 % nucleotide identity over a 530-nt section (p=96.5e^{-5}). In this region, the longest contiguous sequences in common are 12 nt, with several other regions containing no more than 7 contiguous nucleotides in common. Sequence annotation of this region indicates it to nucleotides 47658-49225 of subject annotated as similar to gibberellin 20-oxidase, and the other sequences are not annotated, while the matching CaMV sequences are derived from gene II (1371-1828).

A third example is located in the same BAC clone with 52 % nucleotide identity over a 994-nt section (p=6.5e^{-5}). In this region, the longest contiguous sequences in common are 15 nt, with several other regions containing no more than 6 contiguous nucleotides in common. The *A. thaliana* sequences for this region are not annotated, while nucleotides 535-1347 is derived from gene I, followed by gene II (1348 beyond 1491).

A large number of hits in the BLAST search of the *A. thaliana* database showed limited sequence identity to small regions of the CaMV genome; however, the longest contiguous sequence was 18 nt. In the case of CaMV gene VI and the *A. thaliana* genome, there were 15 clusters with 18-19 nt identical out of 19 nt and one cluster of 25 nt identical out of 27 nt. [See Table 1.] Although the number of hits between sequences of the CaMV genome and the genome of the non-host species rice (*Oryzae sativa*) was less than for experimental host species *A. thaliana*, the sizes of the clusters of (near) identical sequences were similar [Table 1]. On the other hand, no such clusters were identified in the sequences available from the barley and wheat databases. Based on the work undertaken in this project, it does not appear that this limited sequence similarity either supports or prevents integration of CaMV sequences into the *A. thaliana* genome.

Table 1. Longest contiguous sequence identities between CaMV genes (I-VI) and the database of the *A. thaliana* and *Oryzae sativa* genomes, as well as barley and wheat sequences

Host	Gene I	Gene II	Gene III	Gene IV	Gene V	Gene VI	IGR
A. thaliana							
No. of hits	50	80	43	137	73	16	43
No. of nt	18-22	19-25	17-20	18-23	18-22	18-27	18-22
O. sativa							
No. of hits	23	41	11	33	40	21	17
No. of nt	19-23	18-24	19-23	19-24	19-22	19-25	19-22
Barley	None	None	None	None	None	None	None
Wheat	None	None	None	None	None	None	None

IGR = intergenic region between genes VI and I, containing the core of the 35S RNA promoter. No. of hits = number of clusters of sequence identity identified between the particular CaMV gene and the given host genome sequences. No. of nt = range of contiguous nucleotides found in the clusters, with one or two mismatches within a cluster.

Querying the *Brassica* DNA database yielded a number of hits. Some of these are clearly due to the inclusion in the database of sequences from the CaMV 35S RNA promoter present in the T-DNA tags used in the production of clones for sequencing. When these are excluded, the greatest similarity is to a *Brassica oleracea* sequence from a shotgun genome sequencing project, accession number BH950756, which shows 61% identity over 696 nucleotides (p=7.8e^{-32}), although with no more than 12 contiguous nucleotides in common. However, when the *B. oleracea* sequence was itself used to query the non-redundant nucleotide database, the top matches were to sequences from soybean chlorotic mottle virus and blueberry red ringspot virus, both members of the family *Caulimoviridae*, and not to CaMV. Thus, again, the sequence in question is unlikely to be derived from CaMV integration. It is interesting to note that neither of these viruses is known to infect brassica plants and it may be that the presence of this sequence in the *B. oleracea* genome confers homology-dependent resistance, in the manner suggested to account for the conservation of tobacco endogenous pararetroviruses by Mette *et al*. (2002), or to resistance of grapevines to PVY (Tanne and Sela, 2005).

A large number of hits in the BLAST search of the *Brassica* spp. database showed limited sequence identity to small regions of the CaMV genome; however, the longest contiguous sequences were 16-21 nt. [See Table 2.] There were no sequences of similar length related to sequences of CaMV genes IV and V, but sequences of this size range were found in common with each of the other CaMV

genes as well as with the 35S RNA promoter region in the large intergenic region (IGR; between genes VI and I).

Table 2. Longest contiguous sequence identity between CaMV genes (I-VI) and the database of *Brassica* species sequences

Gene I	Gene II	Gene III	Gene IV	Gene V	Gene VI	IGR
Br 21nt Bo 16 nt	Br 21 nt Bn 20 nt Bo 16 nt	Bo 16 nt Br 16 nt	None	None	Br 18 nt; 17 nt	Br 18 nt Bo 18 nt

IGR = intergenic region between genes VI and I, containing the core of the 35S RNA promoter. Br = *Brassica rapa*; Bo = *B. oleraceae*; Bn = *B. napus*.

Effects of integrated CaMV sequence

Although neither the literature nor our analysis of genome databases provide evidence of CaMV integration in host genomes, the possible effects if such integration were to occur are worthy of consideration. Shewmaker *et al*. (1985) introduced a full-length copy of CaMV into plant chromosomes by *Agrobacterium tumefaciens*-mediated transformation and demonstrated transcription of CaMV genes from the integrated DNA in the crown galls. Further study (Young *et al.,* 1987) showed that at least some viral proteins were expressed in the transformed tissue. Potential effects on expression of plant proteins were not examined. The studies noted above on BSV, TVCV and PVCV demonstrated that under certain circumstances integrated sequences of some DNA viruses were capable of leading to episomal infection. However, in neither of the two examples involving integration of a single copy of the CaMV genome was episomal infection of CaMV observed. By contrast, transformation of *Brassica napus* with a partially duplicated copy of the CaMV genome did lead to episomal replication of the CaMV genome, following recombination (Gal *et al.*, 1991). Thus, it was thought to be necessary for some of the viral genome to be duplicated in order to allow transcription of the entire viral genome or recombination to occur, leading to formation of a complete, episomally-replicating genome.

Recombination between CaMV viral sequence present as a transgene in plants and infecting virus also has been shown to occur under laboratory conditions (Gal *et al*., 1992; Schoelz and Wintermantel, 1993) and may even occur under conditions of only moderate selection pressure (Wintermantel and Schoelz, 1996). Considerable attention has been focused on the possibility that the CaMV 35S RNA promoter may contain a recombination hotspot (Kohli *et al.*, 1999), although earlier work that looked at recombination sites in CaMV found these more evenly distributed throughout the whole CaMV genome (Vaden and Melcher, 1990). A different study that examined episomal replication as a consequence of recombination of a partial dimeric CaMV genome transgene found recombination could occur in several regions, although those authors also suggested hotspots for recombination (Gal *et al.*, 1991). However, it is questionable whether any of these examples has any relevance to the situation with a 35S RNA promoter already integrated into a plant genome in the absence of the virus. As has been pointed out by Ho *et al.* (2000), there is great difference between the virus replicating in the cell and the 35S RNA promoter being used as a transgene. In the case of the data obtained by Kohli *et al.* (1999), the authors saw recombination between duplicated CaMV 35S RNA promoters present in

the same plasmid. Recombination between duplicated sequences in viruses or plasmids is well known and that is why such vectors are often designed to avoid having duplicated sequences. The other examples of recombination occurred between defective CaMV genomes co-infecting the same cells (Vaden and Melcher, 1990), CaMV infection of a transgenic plant expressing gene VI of CaMV (Schoelz and Wintermantel, 1993; Wintermantel and Schoelz, 1996), or a partially duplicated CaMV sequences, expressed transiently and episomally (Grimsley *et al.*, 1986) or as a transgene (Gal *et al.*, 1991). Most of the examples of recombinant progeny viruses were generated by template-switching during transcription and so do not represent DNA recombination by strand breakage-ligation reactions, as envisioned for recombination between the CaMV 35S RNA promoter and some other target sequence in the chromosome (Ho *et al.*, 1999; 2000). Some of the examples of recombination involving a permuted partial dimer insert of CaMV might have occurred by the latter mechanism, but these were also the recombinants where the crossing-over occurred furthest from the promoter region (Gal *et al.*, 1991). Therefore, for the other recombinants the explanation that recombination occurred during transcription is the most plausible explanation, since the sequences designated as possible hotspots are those near the initiation or termination sites for transcription, rather than within the 35S RNA promoter sequence itself.

Transgene suppression after virus infection

The question has arisen as to whether transgene expression driven by an RNA transcription promoter derived from a DNA virus would be affected by infection of the transgenic plants with the same DNA virus. This has been evaluated in a few cases. A study by Al-Kaff *et al.* (1998) showed that in oilseed rape (*B. napus*) transgenic for expression of either a *glucuronidase* (*GUS*) gene or the *nptII* gene, both under the control of a CaMV 35S RNA promoter, infection by CaMV resulted in silencing of the transgene expression. In those cases where the transgene cassette also included the 35S RNA transcription terminator sequences, the suppression was post-transcriptional, while in the absence of the CaMV 35S RNA transcriptional terminator sequences, the suppression was due to an inhibition of transcription (Al-Kaff *et al*., 1998). A similar analysis was done with oilseed rape plants expressing the *BAR* gene for tolerance to the herbicide Harvest (ammonium glufosinate). Various levels of silencing of expression of the *BAR* gene also were observed, although the extent of suppression depended on the strain of CaMV and the time of infection (Al-Kaff *et al*., 2000). However, as oilseed rape also showed recovery from CaMV infection (Al-Kaff *et al*., 1998), it is important to know to what extent this effect can be extrapolated to other plants that are hosts for CaMV and do not show recovery from infection, let alone to plant species that are not hosts of CaMV. In fact, Dale and Al-Kaff (2006) found that neither recovery from infection, nor transgene silencing was observed by their strain of CaMV in pakchoi (*B. rapa*). Therefore those studies showed that there is both a CaMV strain and host genotype component determining whether a given GM plant species can be subject to silencing of the transgene.

Although promoters from geminiviruses or nanoviruses have not been evaluated extensively for the expression of transgenes, in one study, the six promoters of the geminivirus tomato leaf curl virus (TLCV) were used to assess whether GUS expression driven by these promoters could be suppressed after infection by TLCV (Seemanpillai *et al*., 2003). Infection by TLCV resulted in the transcriptional

silencing of all six transgene promoters. In addition, this suppression (of one TLCV promoter) occurred after infection by two strains of TLCV (with 100 % and 84 % sequence identity to the transgene promoter), but not after infection by two other geminiviruses (with c. 55 % sequence similarity to the transgene promoter) (Seemanpillai *et al*., 2003).

Conclusions

Under laboratory conditions, evidence has been obtained for the integration of partial or complete genome sequences of several viruses into plant genomes. All but one of these are viruses containing DNA genomes. In the literature, several cases of episomal replication of the virus genome have been found, although horizontal transmission does not occur for most of these viruses. In other cases, the fragmented viral genome sequences do not appear to support episomal replication, but rather may inhibit virus infection. Episomal expression is associated usually with genetic hybridisation of the host and stress. There is no previous evidence for integration of the CaMV genome or large sequences derived from this genome into its host species. The analysis of the *A. thaliana* genome shows that CaMV had not integrated into the genome of the ecotypes for which the genome was partially or completely sequenced. CaMV infection of *B. napus* containing transgenes driven by CaMV 35S RNA promoters led to a recovery from infection and to suppression of transgene expression. However, such suppression of transgene expression did not occur in *B. rapa*, another host of CaMV, but one that did not show recovery from CaMV infection.

Thus, there is a need to assess experimentally the following: (1) whether CaMV integrates into its host genomes, and whether this is at a high or low frequency; (2) whether the CaMV genome integrates into the germline and is transmitted vertically; (3) whether integrated CaMV genomes can be expressed episomally and under what (stress) conditions; (4) whether the presence of a CaMV 35S RNA promoter in the genome of a transgenic plant increases the frequency of CaMV integration in plant genomes; (5) whether integrated CaMV alters plant gene expression, vital functions, or major characteristics of the plant; and (6) whether integrated CaMV expresses its own genes and replicates episomally, and the effects of various stresses on CaMV gene expression and episomal replication. These issues are addressed in this project.

Objective 2. Preparation and validation of methods and materials.

Two modified versions of CaMV were prepared for use in determining whether CaMV sequences integrate into plant genomes of non-GM and GM host plants of CaMV. These were the genetically engineered strain of CaMV designated H7 (Schoelz and Shepherd, 1988) and the natural strain designated W260 (Gracia and Shepherd, 1985). Strain H7 is derived from the aphid-nontransmissible strain CM1841, which had a naturally mutated gene II, but contains gene VI from the mildly virulent strain D4 [See Fig. 1].

The H7 construct was used in most of the experiments and was designed to prevent severe virus infection from generally causing plant death or preventing flowering, and also to prevent aphid transmission from systemically infected and GM plants.

CaMV W260, unlike other CaMV strains, which show no movement in tobacco (*N. tabacum*), will replicate and move cell to cell in tobacco, but will not move long distance in this host species. This allowed assessment of effects of CaMV infection in a plant species (tobacco) that has expression of a transgene driven by the CaMV 35S RNA promoter, but is not considered a natural host of CaMV. CaMV W260 was then used in the GM tobacco here to determine whether gene silencing could be propagated systemically in the absence of systemic virus infection, as potentially could occur in the majority of plant species not susceptible to systemic infection by CaMV. The CaMV strain H7 was used to inoculate the three other plant species to be tested in this study: *A. thaliana* (thale cress), *Brassica napus* (oilseed rape) and *Nicotiana benthamiana*, and CaMV strain W260 was used to inoculate *N. tabacum* in Objective 8 to assess effects of CaMV infection on transgene silencing. The same strains and species were used to examine integration of CaMV into non-GM plants (Objective 3) and into the CaMV 35S RNA promoter of GM plants (Objective 4). The test plants were inoculated with buffered sap extracts made from recently infected turnip (*B. rapa*) plants. The plants were then grown in a greenhouse maintained at 22 / 18 °C (day / night) with 16 hr of daylight supplemented with artificial lighting at 250 μmol m^{-2} s^{-1}.

GM plants expressing the jellyfish green fluorescent protein (GFP) under the control of the CaMV 35S RNA promoter and the *nos* gene transcription terminator were obtained for four species: *A. thaliana* ecotype C24 (Haseloff *et al.*, 1997), *B. napus* cv. Westar (Halfhill, *et al.*, 2001), *N. benthamiana* (Ruiz *et al*., 1998), and *N. tabacum* cv. Xanthi (Harper and Stewart, 2000). All of these plants showed green fluorescence when viewed under a UV lamp or under a fluorescence microscope. Some of the GM *A. thaliana* plants did not show green fluorescence (due to segregation) and so were removed prior to inoculation.

Symptoms of infection were observed in the oilseed rape plants, both GM and non-GM, with vein-clearing present in leaves, [Fig. 3A] or necrotic patches appearing in the leaves [Fig. 3B]. One of the GM oilseed rape plants also exhibited a change in morphology of the flowers [Figs. 4A vs. 4B], from which it later recovered. Back inoculation of sap from a selection of inoculated oilseed rape plants gave rise to symptoms in turnip plants, consistent with CaMV infection. Most of the *N. benthamiana* plants exhibited leaf symptoms following inoculation with CaMV strain H7 [Fig. 5]. Flower production seemed to be unaffected. CaMV infected tobacco produced no distinct symptoms as was expected (not shown).

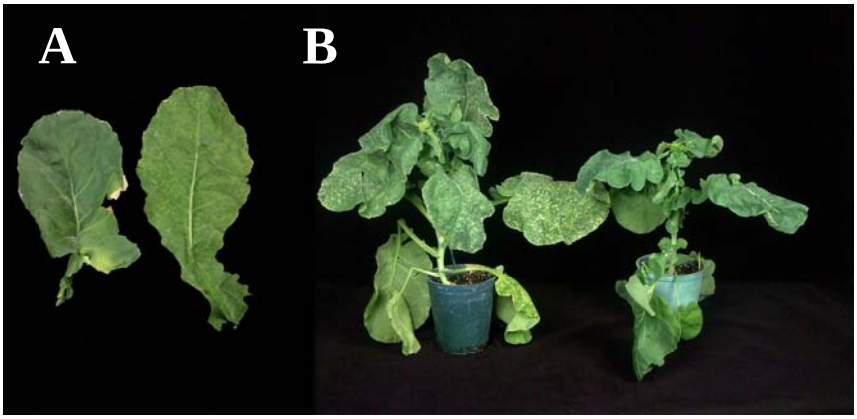

Fig. 3. Oilseed rape plants. (A) Upper leaves of 35S-GFP plants, noninoculated (left) or infected with CaMV strain H7 (right). The photo was taken at 60 days post inoculation. (B) Oilseed rape plants inoculated with CaMV strain H7. 35S-GFP transgenic plant (left), nontransformed plant (right). The photo was taken at 29 days post inoculation.

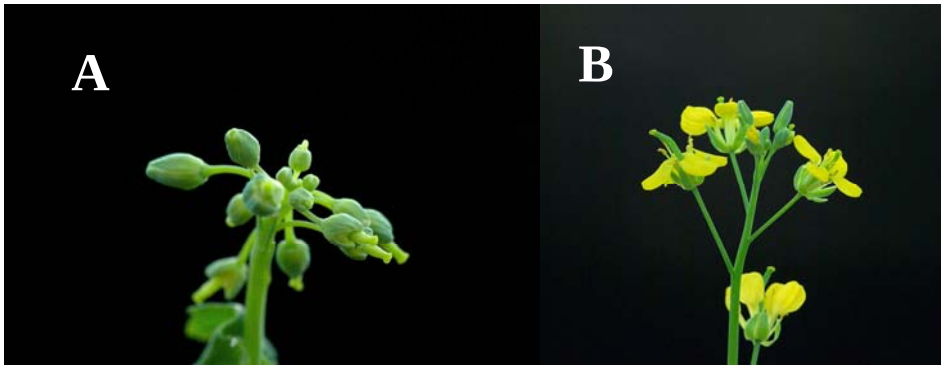

Fig. 4. Flowers of 35S-GFP oilseed rape plants. (A) Distorted flower production in 35S-GFP oilseed rape inoculated with CaMV strain H7. [Note the absence of petals.] (B) Normal flower development in 35S-GFP oilseed rape not inoculated with CaMV.

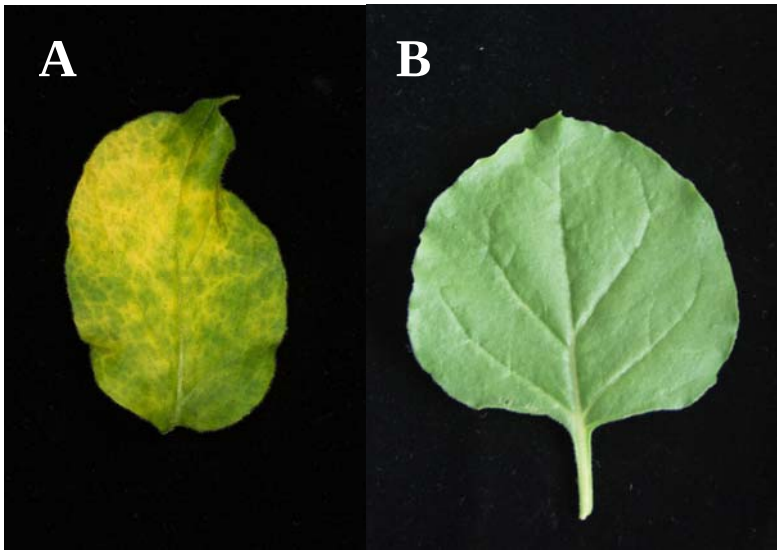

Fig. 5. Leaf of CaMV infected *N. benthamiana*. (A) The *N. benthamiana* leaf exhibits a yellow-green mosaic symptom. The photo was taken 90 days post inoculation with CaMV strain H7. (B) Normal leaf of noninoculated plant.

The detection method involves the use of adaptor-primers and nested PCR to detect the presence of either CaMV integrated into non-GM plants or CaMV integrated in to the CaMV 35S RNA promoter of GM plants. The plant DNA was isolated as described by Cullen *et al.* (2001) and the DNA was digested with *Kpn*I, a restriction endonuclease that does not cut the CaMV genome of strains H7, D4 or W260. The strategy of using adaptor-primers containing *Kpn*I compatible ends and nested PCR to detect integration into the two types of genomes is described in Annex 1. The PCR products for each sample tested were then visualised by agarose gel electrophoresis, and the PCR products were ligated into a plasmid designed to accept PCR products and used to transform *Escherichia coli*. Plasmids isolated from the transformed bacteria were then subjected to sequence analysis, to determine the DNA sequence present in the plasmid. In Objective 3, this strategy used primers specific to CaMV and an adaptor-linker, to amplify DNA that contained CaMV sequences at one end and plant DNA sequences at the other end. In Objective 4, this strategy used primers specific to integrated sequences at both ends: one to the GFP region of the 35S-GFP transgene and another to gene VI of CaMV present as integrated DNA.

To test the sensitivity of the extraction method used, and the efficiency of primers designed to the CaMV 35S RNA promoter region of a GFP transgene, decreasing amounts of 35S-GFP oilseed rape leaf material either alone, or mixed with a constant amount of non-GM oilseed rape material were prepared and extracted. Our extraction method was found to be sensitive enough to detect 0.1 mg of material containing the CaMV 35S RNA promoter in a total of 200 mg of plant material, using primers designed to the CaMV 35S RNA promoter [Fig. 6]. That is, we would have the capacity to detect the equivalent of one cell containing integrated sequences in 2000 cells (or one copy per 4000 haploid genome copies).

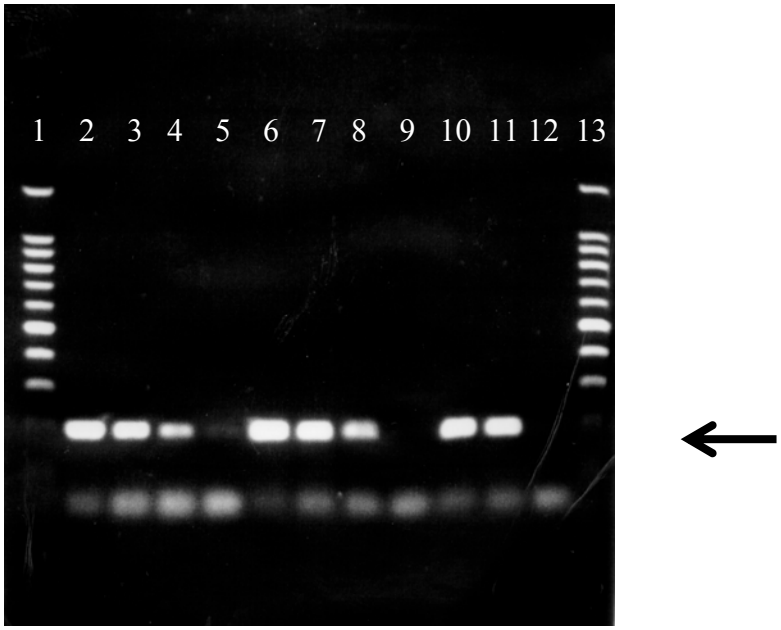

Fig. 6. Analysis of dilution endpoint for detection of CaMV 35S RNA promoter sequences. A 2 % agarose gel was run at 70 V for 30 min. 35S RNA promoter PCR products (5 µl) were analysed after 40 cycles. Samples : (1) M = molecular weight markers; (2-11) DNAs obtained from the following tissues: (2) 100 mg (NonGFP) plus 100 mg (GFP); (3) 190 mg (NonGFP) : 10 mg (GFP); (4) 199 mg (NonGFP) : 1 mg (GFP) (5) 199.9 mg (NonGFP) : .1 mg (GFP); (6) 100 mg (GFP) (7) 10 mg (GFP); (8) 1 mg (GFP); (9) 0.1 mg (GFP); (10) chloroform only extraction (GFP); (11) phenol/chloroform extraction (GFP); (12) No template control; (13) M = Promega 100 bp Markers. Arrow = expected band of 194 bp.

Objective 3. Integration of CaMV into non-GM plants (hosts or non-hosts of CaMV).

CaMV strain H7 was used to inoculate the three plant species *A. thaliana*, *B. napus* and *N. benthamiana*, and CaMV strain W260 was used to inoculate *N. tabacum*. This allowed for an assessment in three different host species and one non-host species (tobacco), respectively. The first two species are in the family *Brassicae* (formerly *Cruciferae*) and the latter two in the family *Solanaceae* (in which many GM plants have been produced with the transgene driven by the CaMV 35S RNA promoter).

Establishing whether CaMV integrates into the genome of its host plants depends on the extent to which integration occurs. In principle, there are four possible levels of integration: (i) Integration may occur at a high frequency and in the germline. In this case, a high proportion of the progeny plants will contain CaMV sequences in all of the cells. (ii) Integration may occur in the germline, but only at a very low frequency. (iii) Integration may occur with a high frequency, but only in somatic cells, since most viruses including CaMV do not invade germline cells. (iv) Integration may occur only in somatic cells, but with a very low frequency.

To determine whether integration occurred in the somatic tissue, leaves from infected non-GM plants of each species were harvested 30 days after inoculation for *A. thaliana* (10 plants), and *N. tabacum* (10 plants), and at 30, 60 and 90 days after inoculation for *B. napus* (10 plants) and *N. benthamiana* (10 plants). Except for the smaller leaves of *A. thaliana*, the leaves of the other species were sectored to allow the best chance of detection of integration of viral genome within the plant DNA; in *A. thaliana*, up to eight leaves were sampled per plant without sectoring. [See Table 3.] These combined approaches allowed us to determine whether CaMV sequences could integrate into the genome of their hosts and if integration occurred more than once, then with what frequency.

To determine whether integration occurred within the germline required the analysis of plants generated from seeds obtained from those plants initially infected with CaMV. Thus, one or two progeny plants were obtained from each of the above CaMV-infected non-GM plants and these were also assessed for CaMV integration.

Table 3. Samples examined for potential integration of CaMV into non-GM plant genomes.

Species	Inoculated Leaf	Systemic Leaf 30day	Systemic Leaf 60day	Systemic Leaf 90day	Seedlings
A. thaliana	80 Leaves [80]	80 Leaves [64]	NS	NS	10 [6]
B. napus	80 Sectors [51]	80 Sectors [44]	80 Sectors [20]	80 Sectors [30]	20 [20]
N. benthamiana	40 Sectors [41]	40 Sectors [40]	40 Sectors [21]	40 Sectors [22]	20 [20]
N. tabacum	40 Sectors [42]	40 Sectors [16]	NS	NS	10 [10]

[No.] = Number of DNA clones sequenced; total number of clones sequenced = 531. NS = Tissues not sampled. Total number of samples tested = 780 = 720 inoculated samples + 60 progeny samples. The progeny samples represent one or two seeds from each of ten parents.

For the DNA extracted from plants inoculated with CaMV, PCR was used with a series of primer pairs as outlined above to determine whether integration of CaMV occurred. In many samples, PCR products were obtained, but usually as smears rather than discreet bands. The nature of the PCR products was established by ligating the PCR DNAs into a plasmid, transforming the plasmids into *Escherichia coli*, isolating the amplified plasmids and determining the nucleotide sequence of the cloned PCR products. The samples were assessed for the presence of the CaMV 35S RNA primer at one end of the PCR product and the adaptor-primer sequence at the other end of the product. Only such products could represent integrated sequences. The sequences of all but one sample (described separately below) indicated PCR amplification of only plant DNA sequences with adaptor-primers at both ends of the sequence, part of the CaMV genome amplified from presumably fragmented CaMV DNA, or unknown sequences [See Table 4].

In only one sample obtained from directly inoculated leaves of non-GM *A. thaliana* was there evidence suggesting CaMV integration into the genome. This is described below [See Figs. 7-9]. In no other sample from any of the four plant species was there any evidence for integration of CaMV into the plant genome. However, if integration occurred with a high frequency, then many of the sectors should have resulted in the positive detection of CaMV sequences. Therefore, these data indicate that if integration of CaMV occurs in leaves of non-GM plants under laboratory conditions, it does so at an extremely low frequency, potentially below the detection limit of our system (i.e., integration into less than one in two thousand cells or less than one per 4000 haploid genomes).

To determine whether integration occurred in the germline, seeds from the above plants were collected and germinated. Leaves were processed as for the leaves from the parent plants. None of the samples tested positive for integration of CaMV. It is interesting to note that discrete PCR bands that were sometimes seen in samples taken from inoculated plants were never seen in samples taken from the progeny of the infected plants. This is consistent with the fact that CaMV does not invade the germline tissue and therefore is not seed transmissible (Brunt *et al*., 1996; Hull *et al*., 1984). Therefore, even in the single non-GM plant where only one of eight directly inoculated leaves and no systemically infected leaves showed evidence of integration of CaMV sequences, there was also no insertion of CaMV to the germline tissue and thus no vertical transfer to the next generation.

The results of this study provided the baseline for the events assessed in Objective 4.

Table 4. Sequence analysis of PCR products from CaMV infected non-GM plants for determination of potential integration of CaMV into plant genomes.

Species	Plant ID No.	Tissue Type	Leaf Sector	Clone ID No.	Adaptor Primer	35S Primer	Search Length	BLAST Search Result
A. thaliana	168	Inoculated	4	1	+	—	565	A. thaliana DNA chromosome 3
A. thaliana				2	+	—	656	A. thaliana chromosome 2
A. thaliana	173	Inoculated	3	3	+	—	402	No Match
A. thaliana				4	+	+	400	No Match
A. thaliana	174	Inoculated	4	5	+	—	977	A. thaliana unknown protein AT3G58940 mRNA
A. thaliana				6	+	—	552	No Match
A. thaliana	170	Inoculated	2	7	+	+	327	CaMV gene six protein gene
A. thaliana				8	+	+	406	A. thaliana unknown protein AT2G30880 transcript variant
A. thaliana	168	Systemic	4	9	+	+	533	CaMV gene six protein gene
A. thaliana				10	+	+	550	CaMV gene six protein gene
A. thaliana	169	Systemic	3	11	+	+	550	CaMV gene six protein gene
A. thaliana				12	+	—	1375	No Match
A. thaliana	169	Inoculated	3	13	+	+	389	A. thaliana DNA chromosome 4
A. thaliana				14	+	+	320	CaMV gene six protein gene
A. thaliana	169	Inoculated	2	15	+	+	128	A. thaliana BAC T8M17
A. thaliana				16	+	—	653	A. thaliana DNA, chromosome 4 centromere region
A. thaliana	173	Inoculated	2	17	+	—	1091	No Match
A. thaliana				18	+	—	1759	No Match
B. napus	24	Inoculated	3	19	+	+	157	A. thaliana DNA chromosome 4
B. napus				20	+	+	258	Brassica rapa BAC clone, KBrH001P13
B. napus	25	Inoculated	2	21	+	+	232	A. thaliana genomic DNA, chromosome 5,
B. napus				22	+	+	85	CaMV P6 transactivator protein gene
B. napus	18	Systemic	2	23	+	+	189	CaMV gene six protein gene,
B. napus				24	—	+	944	CaMV gene six protein gene approx bases 460 to end

Samples were from either *A. thaliana* or *B napus* and taken from either directly inoculated leaves or systemically infected leaves (at 30 dpi). For each sample that yielded a DNA clone for which a sequence could be obtained, the number of the plant and the leaf number (*A. thaliana*) or leaf sector number (*B. napus*) are given, along with the DNA clone ID number, whether the adaptor primer or the 35S primer sequence was present in the PCR DNA clone (+/—), the length of the sequence (in bp), and nature of sequence, as determined by a computerised search using the programme BLAST.

The single plant sample that showed putative integration of CaMV sequences yielded several clones that when sequenced showed possible integration into several different sites of the *A. thaliana* genome. The first clone obtained from one PCR reaction showed possible integration in *A. thaliana* DNA sequences in a region devoid of genes from Chromosome 2 [Fig. 7]. The plant DNA sequences [red, Fig. 7] were fused directly to part of the extended CaMV 35S RNA promoter sequence [dark green, Fig. 7], upstream of the core promoter region, with five nucleotides being present in both the CaMV gene VI sequence and in the *A. thaliana* DNA [yellow, Fig. 7], indicating that homologous recombination had occurred. The adaptor-primer was present at the end of the plant DNA [blue, Fig. 7] and the 35S RNA promoter

sequence was present at the end of the CaMV gene VI sequence (light green, Fig. 7), confirming the nature of the PCR product. Additional PCR reactions from the same sample using different primers indicated that only a small region of the CaMV genome had potentially integrated into the *A. thaliana* genome, although it was not possible to establish the amount of the CaMV genome that may have integrated.

A The nucleotide sequence of the PCR product.

CACGCGNNGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCA
CTAGTGATTGCTTCCGAATTCCAGTTACACGGCCGGTTTAAAAAGAAAACAA
ATAATCTTTTGAGTTAGGTTCTTTAGTTTTTTTCTCTGAGAAACAATTATCCAA
AACTCCAAAGGTTTGGTTACATGTCATATTGGGGGATATGTCAATCGAGACG
GAGACGTTTTAGTTCTACAAAGGGACATTAAGCTCATTGCCATGGTGCTACTC
ATTACTTTTTGTAATAGGACTCATTACTTGAATTGTTTACAAAAGGTTTCTTA
AAAGTCACTTCAATTATTTTAGTAAACAATAATATCTGTCACTTCATCGAAAG
GACAGTAGAAAAGGAAGGTGGCTCCTACAAATGCCATCATCAAATCGAATTC
CCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCCAATTCGCCCT
ATAGTGAGTCGTATTACAATTCAC

B The partial sequence of *A. thaliana* ecotype Columbia, chromosome 2, clone F27L4.

33061 atcttcagtt tatagtaaaa tagttaagaa aaacgtgaca tatttgttta
aaaagaaaac aaataatctt ttgagttagg ttctttagtt tttttctctg agaaacaatt
atccaaaact ccaaaggttt ggttacatgt catattgggg gatatgtcaa tcgagacgga
gacgttttag ttctacaaag ggacattaag ctcattgcca tggtgctact cattactttt
tgtaatagga ctcattactt gaattgttta caaagggttt tcttaaaagt cacttcaatt
attttagtaa acaataatat ctacgtttca cctaattcct tacaactaat aaatttctat
tcgaagtatt ttatatactc tattaggaga gtatgatgac actataattg gcaatataaa
tcttacatga **33480**

C The sequence of CaMV gene VI from nt 7121 to 7200.

7121 aacctcctcg gattccattg cccagctatc tgtcacttca tcgaaaggac agtagaaaag
gaaggtggca cctacaaatg 7200

Fig. 7. Potential integration of CaMV into the *A. thaliana* genome. (A) The sequence of the PCR product: White = cloning vector, Adaptor primer, Remainder of adaptor, Plant sequence, Change over from plant to virus, Viral sequence, 35S primer; (B) the partial sequence of the *A. thaliana*, Ecotype "Columbia", chromosome 2 clone F27L4 map CIC06C07, Accession number AC004482. Sequence matching in clone, Crossover site; (C) the sequence of CaMV gene VI from nucleotide 7121 to 7200. Crossover site, Sequence matching in clone

The sequences from two other cDNA clones of PCR products derived from a separate amplification reaction from that in Fig. 7, but using the same DNA sample, showed putative insertion into two other regions of the *A. thaliana* genome [Figs. 8 and 9]. In one case, possible insertion occurred into a ribosomal RNA (rRNA) gene on chromosome 2 [Fig. 8], while in the other case, potential insertion occurred within an auxin response transcription factor (ARF2) gene on chromosome 5 [Fig. 9]. In the former case, there were only 5 nt in common between the CaMV sequence and the rRNA gene sequence [yellow, Fig. 8], while in the latter case there were 10 nt in common between the CaMV sequence and the ARF2 gene sequence [yellow, Fig. 9].

A The nucleotide sequence of the PCR product for >220A4.6, plant 220 Inoculated leaf A4, clone 6.

GCTTCCGAATTCCAGTTACACGGCCGGTACTATCGCGACGCGGGCATCAGTAG
GGTAAAACTAACCTGTCTCACGACGGTCTAAACCCAGCTCACGTTCCCTATTGG
TGGGTGAACAATCCAACACTTGGTGGATTCTGCTTCACAATGATAGGAAGAGC
CGACATCGAAGGATCAAAAAGCAACGTCGCTATGAACGCTTGGCTGCCACAAG
CCAGTTATCCCTGTGGTAACTTTTCTGACACCTCTAGCTTCAAATTCCGAAGGT
CTAAAGGATCGATAGGCCACGCTTTCACGGTTCGTATTCGTACTGAAAATCAG
AATCAAACGAGCTTTTCAACAAAGGGTAATATCGGGAAACCTCCTCGGATTCC
ATTGCCCAGCTATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGAAGGTGGC
TCCTACAAATGCCATCA

B Part of the rDNA 10.5 kbp repeat unit sequence, BAC clone F23H14, *A. thaliana* chromosome 2.

```
      7681 gtttttggtc ccaaggctcg cttcggcggg tcgatccggg cggaggacat
tgtcaggtgg ggagtttggc tggggcggca catctgttaa aagataacgc aggtgtccta
agatgagctc aacgagaaca gaaatctcgt gtggaacaaa agggtaaaag cttgtttgat
tctgattttc agtacgaata cgaaccgtga aagcgtggcc tatcgatcct ttagaccttc
ggaatttgaa gctagaggtg tcagaaaagt taccacaggg ataactggct tgtggcagcc
aagcgttcat agcgacgttg ctttttgatc cttcgatgtc ggctcttcct atcattgtga
agcagaattc accaagtgtt ggattgttca cccaccaata gggaacgtga gctgggttta
gaccgtcgtg agacaggtta gttttaccct actgatgccc gcgtcgcgat agtaattcaa
cctagtacga gaggaaccgt tgattcgcac aattggtcat cgcgcttggt tgaaaagcca
gtggcgcgaa gctaccgtgc gctggattat gactgaacgc ctctaagtca gaatccgggc
tagaagcgac   8280
```

Fig. 8. Integration of CaMV into the *A. thaliana* genome. (A) The sequence of the PCR product: Adaptor primer, Remainder of adaptor, Plant sequence, Change over from plant to virus, Viral sequence, 35S primer; (B) the sequence of the *A. thaliana*, Ecotype "Columbia", chromosome 2 BAC F23H14 genomic sequence, Accession Number AC006837. Crossover site; Sequence matching in clone. The orientation of the sequence is complementary to that shown in A.

A The nucleotide sequence of the PCR product for 220A4.12, plant 220 Inoculated leaf A4, clone 12

GCTTCCGAATTCCAGTTACACGGCCGGTACCTCTTCGCCTTCAAATCTCATTTT
GAATCTCATGCCAATAGAGTAGTTATTCTTAACAGACTCCATATACTGATCGA
ACGGACAGTAGAAAAGGAAGGTGGCTCCTACAAATGCCATCA

B Partial *Arabidopsis thaliana* ARF2 mRNA sequence

```
      1261 acagggacta tgtttacagt ctactacaaa cccaggacga gcccatctga
gtttattgtt ccgttcgatc agtatatgga gtctgttaag aataactact ctattggcat
gagattcaaa atgagatttg aaggcgaaga ggctcctgag cagaggttta ctggcacaat
cgttgggatt gaagagtctg atcctactag gtggccaaaa tcaaagtgga gatccctcaa
ggtgagatgg   1500
```

Fig. 9. Integration of CaMV into the *A. thaliana* genome. (A) The sequence of the PCR product: Adaptor primer, Remainder of adaptor, Plant sequence, Change over from plant to virus, Viral sequence, 35S primer; (B) the sequence of the *A. thaliana*, chromosome 5, ARF2 (AUXIN RESPONSE FACTOR 2); transcription factor (ARF2) mRNA, Accession Number NM_125593. Crossover site; Sequence matching in clone. The orientation of the sequence is complementary to that shown in A.

The regions of sequence overlap (the sites of recombination) are shorter than the 18-25 nt of contiguous sequences identified in the interrogation of the *A. thaliana* database for contiguous sequence identity between the CaMV and *A. thaliana* genomes [Table 1]. It is not considered likely that these sequences were detected only in *A. thaliana* because this species has a smaller genome that the other plant species examined, since our detection limit as established for oilseed rape [Fig. 6] should have detected less than one copy per genome.

Objective 4. **Integration of CaMV into GM plants (hosts or non-hosts of CaMV).**

Establishing whether CaMV can integrate into the genome of GM plants (hosts or non-hosts of CaMV) involved virtually the same experimental analysis as in Objective 3. In this case, the GM plants were *A. thaliana*, *B. napus, N. benthamiana* and *N. tabacum*, all expressing a gene encoding the GFP under the control of the CaMV 35S RNA promoter. All of these plants were grown in a contained glasshouse. The plants were inoculated with virus and leaf material was taken and extracted according to the following strategy. Leaves from infected plants of each species were harvested 30 days after inoculation for *A. thaliana* (10 plants), and *N. tabacum* (10 plants), and at 30, 60 and 90 days after inoculation for *B. napus* (10 plants) and *N. benthamiana* (10 plants). Except for the smaller leaves of *A. thaliana*, the leaves of the other species were sectored into eights to allow the best chance of detection of integration of viral genome within the plant DNA; in *A. thaliana*, eight leaves were sampled per plant without sectoring. These combined approaches allowed us to determine whether CaMV sequences could integrate into the 35S RNA promoter. If integration occurred, these approaches would allow the frequency and nature of integration of CaMV DNA to be assessed, both in the directly infected GM plants and in progeny seedlings of the infected GM plants.

PCR analyses of the samples from the four GM species *A. thaliana*, *N. tabacum*, *B. napus*, *and N. benthamiana* were completed for the sample numbers examined and given in Table 5. PCR was conducted using primers designed to gene VI of CaMV (as the forward primer), and to a site within the GFP gene in the plant DNA (as the reverse primer). The GFP gene is regulated by the CaMV 35S RNA promoter in the GM plants in this study. No integration was detected by this approach in the 3856 samples examined indicating that either integration had not occurred or that it occurred at a level below the sensitivity of our system (i.e., equivalent to one in 2000 cells; one copy in 4000 haploid genomes). This assessment included inoculated GM plants as well as in the progeny seedlings generated from the infected GM plants. Therefore, despite the sequence identity of over several hundred nucleotides between the CaMV genome and the CaMV 35S RNA promoter present in each infected cell, integration, if it does occur, is at such a low frequency as to be undetectable. The absence of detectable integration is unlikely to have been due to substantial differences in sequence between the CaMV 35S RNA promoters present in the transgenes and the inoculated viruses, since the strains of CaMV under consideration here differed by less than 5 % in nucleotide sequence.

Table 5. Number of samples examined for potential integration of CaMV into the genome of 35S RNA promoted GFP plants.

Species	Inoculated Leaf	Systemic Leaf, 30 dpi	Systemic Leaf, 60 dpi	Systemic Leaf, 90 dpi	Seedlings
A. thaliana	80	80	NS	NS	800
B. napus	80	80	80	80	592
N. benthamiana	80	80	72	72	720
N. tabacum	80	80	NS	NS	800

NS = not sampled. In the case of *A. thaliana* the samples were complete leaves, while for the other plants, the samples were leaf sectors consisting of eighths of a leaf. Sampled: Ten parents of each species, up to 10 seedlings per parent taken at various days post inoculation (dpi). The exceptions being *B. napus* which had lesser seedlings for some: one with only 3 seedlings, one with only one seedling and one plant that did not survive to set viable seed. With *N. benth*amiana one plant died due to rotting at the base at *c.* one month. The total number of samples screened therefore is 3856.

To ensure that the samples analysed were amplifiable, and to rule out poor quality template as the reason for not detecting any integration, a subset of samples (344) covering all four plant species, different sampling time-points and the progeny seedlings were amplified for the presence of the integrated CaMV 35S RNA promoter and GFP gene. These data showed that template was of amplifiable quality and quantity for all samples tested (data not presented).

Objective 5. Effects of the CaMV 35S RNA promoter on CaMV integration in plants.

The results from Objective 3 demonstrated that natural integration of CaMV sequences into plant genomes was a rare event, being found in only one leaf from one plant species of 780 samples tested from four plant species. That is, we did not detect such integration in *B. napus*, *N. benthamiana* or *N. tabacum*, and only detected one example of integration in one of 160 CaMV-infected *A. thaliana* samples analysed. The results from Objective 4 showed no integration of CaMV viral DNA into the CaMV 35S RNA promoter sequences of four GM plant species (944 samples), or into their progeny (2912 samples). Therefore, while this does not completely rule out the possibility that the CaMV could integrate into the CaMV 35S RNA promoter of non-germline tissues at some very low level below the detection limit of 0.05 %, the data do show that the presence of CaMV 35S RNA promoter sequences in themselves do not lead to integration of CaMV DNA into the GM plants above a level that already occurs in non-GM plants. Moreover, if integration occurred below a frequency of one in 2000 cells (0.05 %), then integration is unlikely to have any measurable consequence regarding effects on gene expression. Furthermore, as demonstrated in Objective 9, it is infection by the virus (CaMV) itself that has more of an effect on the plant.

The progeny plants of all four species examined in both Objective 3 and Objective 4 also were not infected by CaMV. The absence of infection in the 2912 progeny of the infected parental GM plants in Objective 4 also indicates that the presence of the CaMV 35S RNA promoter in the GM plants does not facilitate

transmission of CaMV through the germline. This had not been established previously.

Interestingly, Schoelz and colleagues (2005) recently have reported integration of the CaMV genome into GM *A. thaliana* expressing gene VI of CaMV, but driven by a 35S RNA promoter (rather than the native 19S RNA promoter as in the virus). These examples of integration were quite common, with an integration frequency of one virus genome insertion for every 128-256 gene VI transgenes (i.e., 0.78 – 0.39 %). This high level of integration may have been due to the more extensive region of sequence identity between the virus and the gene VI transgene (c. 1560 bp) vs. the c. 650 bp of homology between the CaMV 35S RNA promoter and the virus. It is also conceivable that there are hot spots for recombination in gene VI, depending on the mechanism of recombination. The latter is suggested by the locations of the sites of insertion of CaMV within the gene VI transgene (Schoelz *et al.*, 2005). Nevertheless, even if integration of CaMV occurred into the 35S RNA promoter with a frequency as high as that observed by Schoelz and colleagues for integration into the CaMV gene VI transgene, the effects on (reporter) transgene expression at the whole plant level due to integration would be marginal. That is, if the complete genome of the virus is integrated, then the transgene would still have a functional 35S RNA promoter (albeit one derived from two CaMV strains). On the other hand, if only part of the CaMV genome is integrated and the 35S RNA promoter driving transgene expression is inactivated, this will have little consequence, since at most only 0.39 – 0.78 % of the cells would be affected. Moreover, integration into the germline should not occur as the virus is not seed transmissible (Brunt *et al.*, 1996; Hull *et al.*, 1984), again indicating that there would be little-to-no consequence even if such integration occurred. The fact that GM plants expressing gene VI are not being contemplated for use in field situations provides a further level of reassurance.

Objective 6. Transformation of Arabidopsis with the CaMV genome.

Determination of whether CaMV integration affects transgene expression and stability (Objective 8) or vital plant functions (Objective 9) requires having plants in which integration has been shown to occur to a high enough level (or percentage of cells) to measure these effects. Since integration of CaMV into the CaMV 35S RNA promoter of GM plants was not observed in the laboratory studies, integration would not have been detectable in the germline of the plants inoculated above. However, given that hundreds of thousands to millions of GM-modified plants would be expected to be present in field situations, if some plants did show integration of CaMV into their genomes, then there could be a low number of plants in which some effects could occur. Therefore, it is necessary to establish the consequence of germline integration, should such somehow occur. To this end, we have generated GM *A. thaliana* plants containing a single copy of an inserted CaMV genome (CaMV-GM plants). These plants allowed assessment of the consequences of germline integration, since such plants were not detected in our experimental assessment of four species after infection by CaMV. *A thaliana* was used for two reasons: (1) this species can be transformed and regenerated much faster than *B. napus* plants; and (2) the nucleotide sequence of the entire genome of this species has been determined and annotated, allowing a much clearer assessment of the effects of integration on the gene expression and development of the plant.

A. thaliana plants were transformed using standard protocols at the SCRI GM plant transformation facility. A full-length single-copy CaMV clone was generated by recombinant DNA techniques from a newly constructed version of the CMV hybrid virus H7 called H7-XhoI. In this construct the circular viral genome was linearised at the *Xho*I site in gene II, rather than the traditional *Sal*I site in gene V. Gene II is dispensable for replication and movement of the virus, while all of the other CaMV genes are essential and cannot tolerate insertions that may lead to frameshifts. This has also left the 35S RNA promoter region intact, and does not split the reverse transcriptase gene (gene V) as was done in the earlier transformations (Shewmaker *et al*., 1985; Young *et al*., 1987).

The viral sequences were transformed into the *Agrobacterium tumefaciens* vector pGreen 0229, and then transformed into the *A. tumefaciens* vector pSoup c58 in strain GV3101. This was then used to transform flowering *A. thaliana* plants by the flower dip method (Clough and Bent, 1998). The transformation vector without an insert (empty vector) also was used to generate control GM plants. Seeds were collected from these parents and were screened against the selectable marker Basta herbicide, both in soil and *in vitro* on MS10 media supplemented with Basta. The plants showing resistance to Basta were tested by PCR to detect the presence of the promoter from the pGreen (NOS) through to the polylinker region. The presence or absence of transformants was confirmed by the size of the PCR product. This resulted in the generation of three CaMV-GM lines containing the CaMV genome [316B, 317-2, and 318-1] and one GM line [320B] containing an empty vector.

Objective 7. Introgression of the CaMV genome into GM Arabidopsis plants expressing a marker transgene.

To produce plants that would allow the determination of the effects of the integration of the CaMV genome on the expression and stability of a transgene (Objective 8), pollen from GM *A. thaliana* plants expressing the gene encoding the GFP under the control of the 35S RNA promoter was used to fertilize the three GM lines containing the CaMV genome [316B, 317-2 and 318-1] and one GM line containing the empty vector [320B]. Prior to pollination, the recipient plants were first emasculated, which resulted in less seed set. Seeds from selected pods resulting from these crosses (three lines of CaMV x 35S-GFP and one line of empty vector x 35S-GFP) were obtained and germinated. Not all seeds germinated, but in those that did, the young seedlings were screened by PCR for the presence of both the GFP transgene and the inserted CaMV genome. Those plants containing both genes were examined further in Objective 8. It should be noted that the T2 progeny used in subsequent analyses should show segregation of the two transgenes, since the T1 parents would be hemizygous. Seeds used for experiments below were first selected on plates containing Basta herbicide, to eliminate those progeny seedlings that did not contain the CaMV transgene, and if one copy of the GFP transgene was present, then half of the remaining progeny would be expected to show green fluorescence. However, since most of the progeny seedlings showed green fluorescence, this suggests that the 35S-GFP transgenic *A. thaliana* plants had more than one integrated copy of the GFP transgene.

Objective 8. Determine whether integrated CaMV affects expression and stability of a transgene.

To determine whether integration of CaMV affects expression and stability of the transgene, GM *A. thaliana* plants containing an integrated copy of the CaMV genome (CaMV-GM, generated in Objective 6) were crossed with GM *A. thaliana* plants expressing the GFP under the control of the CaMV 35S RNA promoter (35S-GFP-GM), and the doubly transgenic progeny plants (produced in Objective 7) were screened for their effects on GFP expression. Besides screening for the presence of both the CaMV genome sequences and the GFP transgene, the plants were assessed for fluorescence indicating GFP expression in those plants. The plants were tested for fluorescence using a UV hand lamp, and subsequently by confocal microscopy (not shown). This yielded six seedlings of line 316B, two of line 317-2 and five of line 318-1, all with full length CaMV H7 insert and positive for GFP expression. In addition, nine seedlings of line 320B transformed with empty vector pGreen 0229 and positive for GFP expression were produced.

In the T1 plants listed above, no differences were detected between the fluorescence observed by screening any of the plants with a UV hand lamp. The seedlings that tested positive for both GFP and either virus insertion or empty vector insertions were studied over time up to seed set. There was no change in the expression of GFP in the plant material observed by UV lamp, with the exception that a general reduction in fluorescence was observed with the onset of senescence in the plants (see below). This was expected as it also occurred in 35S-GFP-GM *A. thaliana* and is due to the effects of aging on the plant's metabolism.

Seeds were collected from the above individuals and sown, and T2 progeny seedlings were screened for GFP fluorescence. No differences were observed within the doubly transgenic plant lines examined or between them and singly transgenic GFP-expressing seedlings used as a positive test in the experiment [Fig. 10 and Table 6]. Thus, it is concluded that integration of the entire CaMV genome into the plant genome (by transformation) had no observable effect on the stability or the expression of the GFP transgene regulated by the CaMV 35S RNA promoter.

The data in Fig. 10 showed that the GM plants were segregating for expression of the GFP transgene. These progeny non-fluorescent plants did not contain the GFP transgene, as determined by PCR [Table 6] and showed a level of autofluorescence similar to that of non-GM plants (ecotype C24). Since this was also seen in the T2 progeny of cross between the 35S-GFP and empty vector line 320B, it is not due to some sort of aberrant effect by the CaMV genome on the GFP transgene. The reason why all tested T2 progeny of two T1 plants of line 318-1 x 35S-GFP had lost the GFP transgene [Table 6] was not investigated further.

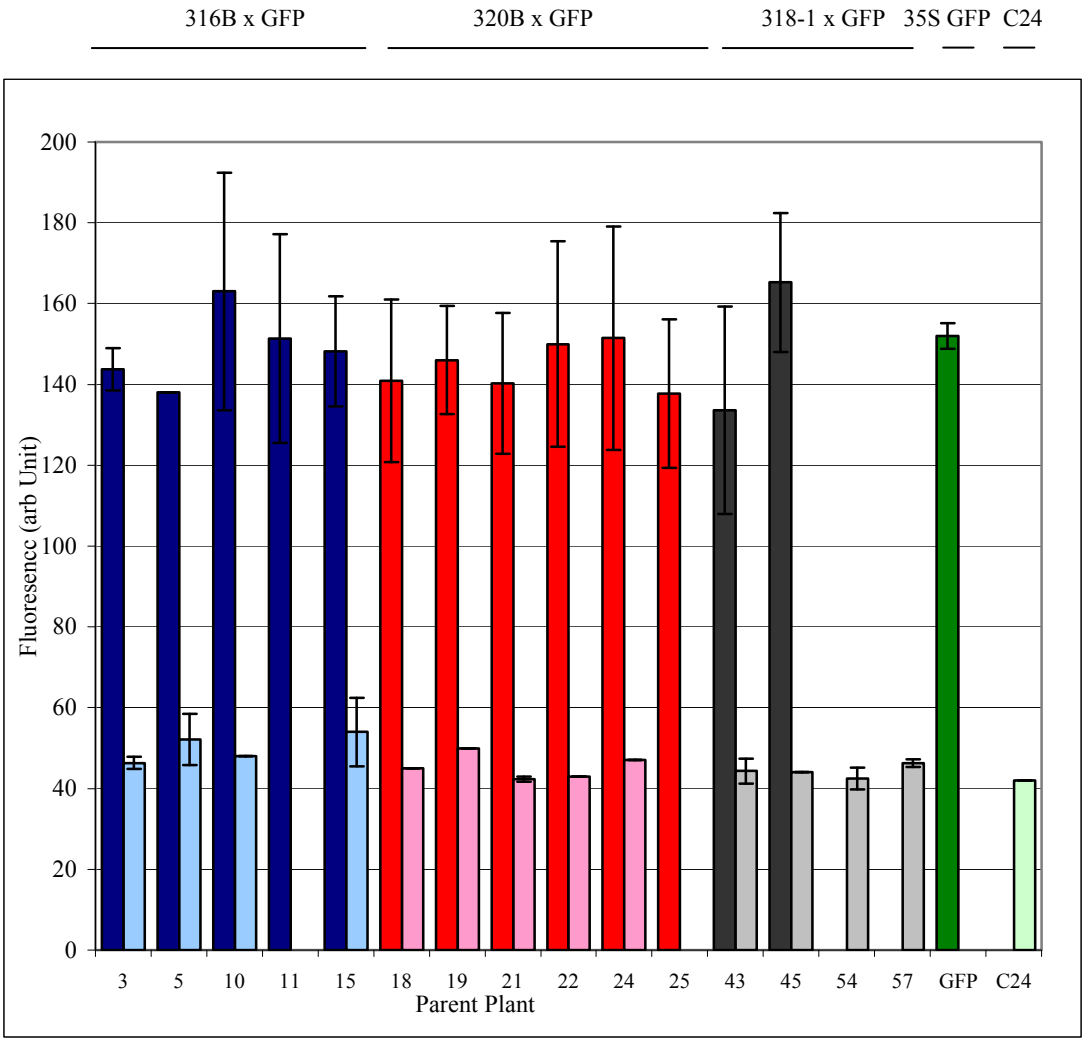

Fig. 10. GFP expression in single and double transgenic *A. thaliana* plants. Mean fluorescence of T2 GM plants of 35S-GFP x CaMV-GM, measured at 25 days after germination, using a Nikon Optiphot Microscope, under 20 x magnification, fitted with GFP3 filter, under UV light and a Coolview camera. Images were analysed with ImagePro, using chip integration of 15 frames, and a gain setting of 18. Line 316B x GFP (dark blue), line 318-1 x GFP (dark grey) (both containing full length CaMV inserts) and line 320B x GFP (red) (empty vector). The fluorescence seen in segregating non-GFP plants is shown in the lighter tint bars. Error bars are +/- standard deviation. Numbers on the *x*-axis refer to the T1 plant numbers of the T2 progeny seedlings [see Table 6]; GFP = 35S-GFP single transgenic plants; C24 = wt C24 non-GM plants, showing the basal level of autofluorescence.

Table 6. Numbers of seedlings sampled for GFP fluorescence.

GM parental plant line	T1 plant no.	No. of T2 plants. Non-GFP	No. of T2 plants. GFP
316B x GFP	3	3	7
316B x GFP	5	6	1
316B x GFP	10	1	8
316B x GFP	11	0	4
316B x GFP	15	2	4
320B x GFP	18	1	8
320B x GFP	19	1	8
320B x GFP	21	3	4
320B x GFP	22	1	3
320B x GFP	24	1	7
320B x GFP	25	0	9
318-1 x GFP	43	3	5
318-1 x GFP	45	1	9
318-1 x GFP	54	6	0
318-1 x GFP	57	4	0

Seedlings were selected on Basta (50 µg/ml) for only those that contained vector (line 320B) or virus (line 316B or 318-1) insert. Non-GFP = negative by microscopy and PCR for the 35S-GFP transgene.

Objective 9. Determine whether CaMV integration affects vital functions.

CaMV-GM plants, containing an integrated copy of the CaMV genome, were tested for effect of CaMV integration into the genome on various properties including growth, flowering, seed set and germination. Assessment of vital function was completed for eight different lines of *A. thaliana*. These were wild type ecotype C24, C24 rub inoculated with CaMV strain H7, GM line 320B (empty vector), GM lines 318-1 and 316B (both C24 transformed with full length CaMV H7), 35S-GFP (C24 transformed with 35S-driven GFP), wild type ecotype Col-0, and D4-2 (Col-0 expressing CaMV gene VI protein). These lines were examined for the number of days to flowering, the total height of the plant including inflorescence, the total fresh weight of the plant (including, root, rosette, inflorescence and seeds), the wet weight of the root mass, the wet weight of the aerial parts of the plant, the number of seed set per individual, the percentage germination of these seed (although the extent of non viability in the non geminating seed could not be determined, as non germinating seed may require longer vernalisation) and the multiplication rate of the line (i.e., the number of seed set x the germination percentage) to give the estimated number of individuals passing into the next generation after this round of screening. These results are presented in Figs. 11-13.

Integration of the CaMV genome showed some effect on plant height [Fig. 11] and on plant (but not root) biomass in one line (316B) [Fig. 12], but not in the other line (318-1) [Figs. 11 and 12]. A slight effect was also observed in line 316B on days to flowering, which was not the case in line 318-1 [Fig. 11]. There was also a slight decrease observed for days to flowering and plant height in the line expressing CaMV gene VI (D4-2) vs. its control (Col-0), but this could have been due to the transformation event itself in the case of time to flowering, since the plants transformed with an empty vector had the shortest time to flowering [Fig. 11]. One line containing an integrated CaMV genome (316B) showed a significant effect on the

number of seed set, but only a marginal effect on percent seed germination, leading to a detectable effect on the multiplication rate [Fig. 13]. The other line containing an integrated CaMV genome (318-1) had no effect on seed set, but a very strong effect on the percent germination, leading to a similar overall effect on the multiplication rate [Fig. 13]. However, the line transformed with the empty vector itself while showing no effect on the number of seed set, showed the strongest effect on seed germination, leading to the most pronounced effect on the multiplication rate [Fig. 13].

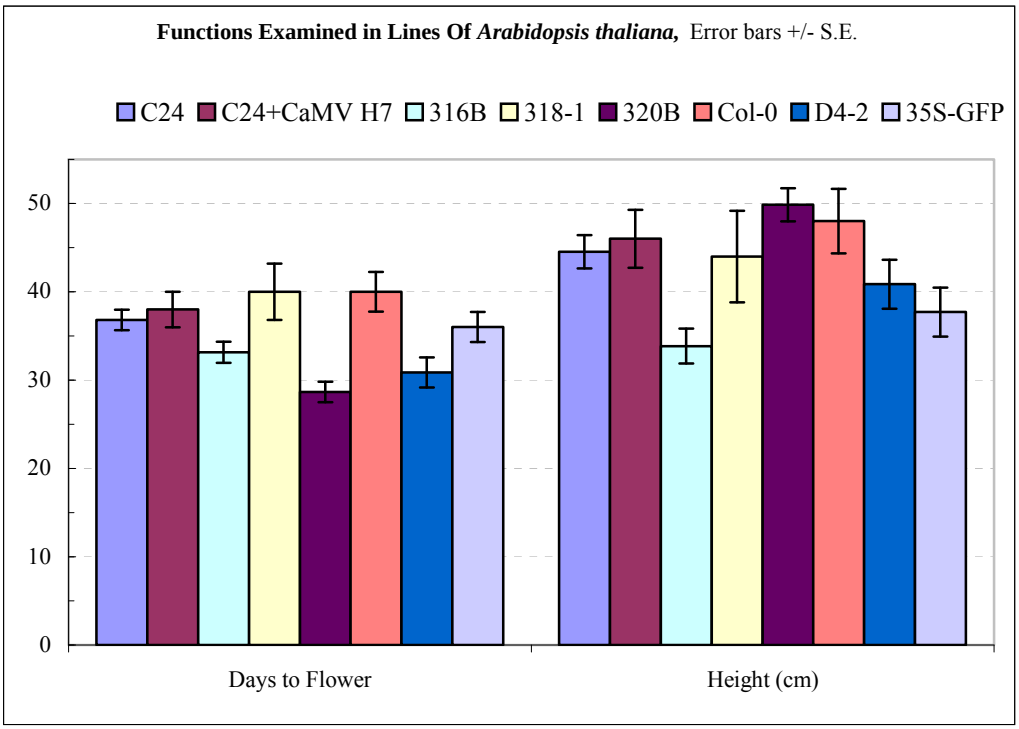

Figure 11. Histogram showing effects on *A. thaliana* vis-à-vis days to flowering and plant height. The data shown are for non-GM plants (ecotypes C24 and Col-0), non-GM plants infected with CaMV strain H7, GM plants transformed with a full-length copy of the CaMV genome (316B and 318-1), GM plants transformed with an empty vector (320B), GM plants transformed with CaMV gene VI (D4-2), and GM plants transformed with 35S-GFP.

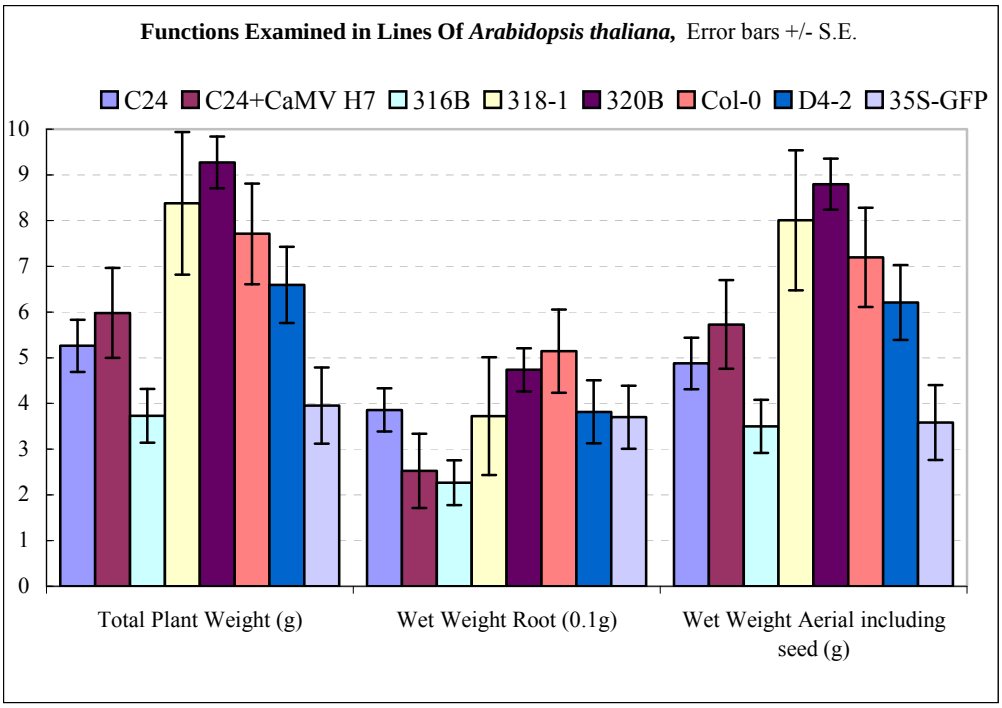

Fig. 12. Histogram showing effects on *A. thaliana* vis-à-vis total plant weight, wet weight of roots and wet weight of the aerial parts of the plants. The data shown are for the same lines described in the legend to Fig. 11.

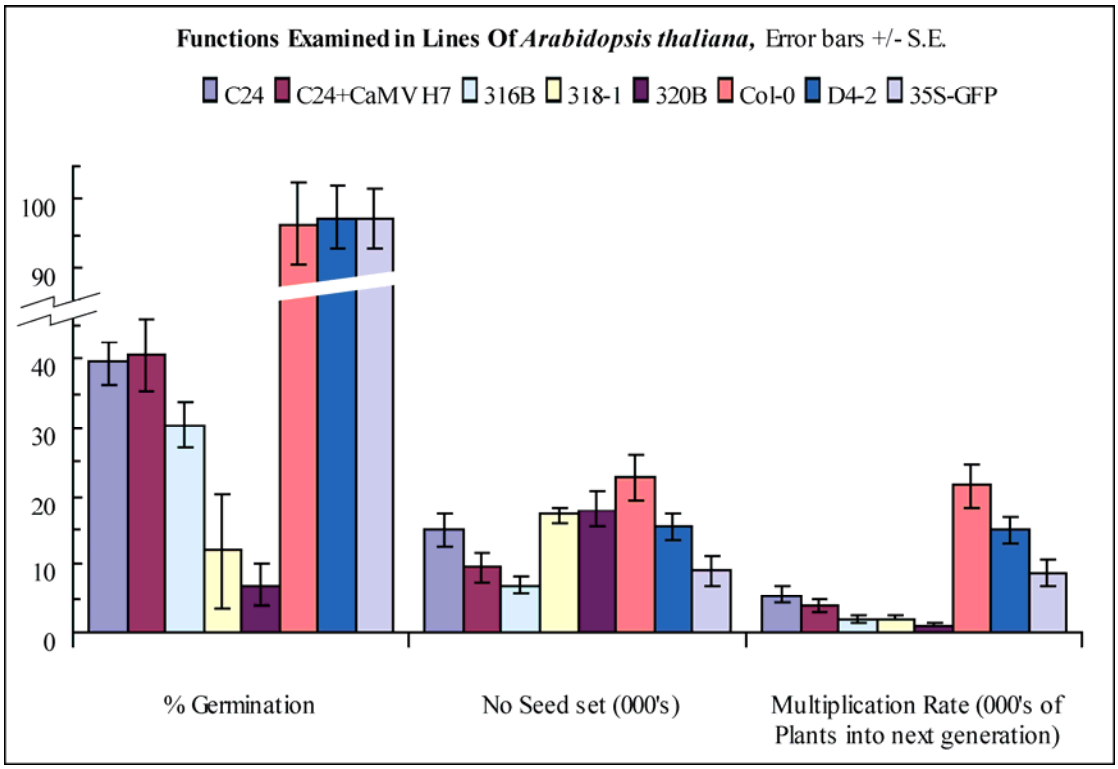

Fig. 13. Histogram showing effects on *A. thaliana* vis-à-vis percent germination, number of seed set and multiplication rate. The data shown are for the same lines described in the legend to Fig. 11.

By contrast, the line transformed with 35S-GFP showed a much higher germination rate than the non-transformed plants, indicating that this phenotype was quite variable between transgenic lines [Fig. 13]. Therefore, overall, the effects observed were mostly due to the integration itself or the location of the insertion and not due to the presence of the CaMV genome, since viral infection itself did not affect the germination rate [Fig. 13]. This was supported by the observation that plants expressing either CaMV gene VI of the GFP transgene showed small effects on time to flowering and plant height [Fig. 11], no significant effects on plant or root biomass [Fig. 12], and a decrease in the number of seed set, but no decrease in the percent germination, although still leading to a significant decrease in the multiplication rate [Fig. 13].

Since the above list of properties is not exhaustive and cannot measure phenotypes for uncharacterised genes, we used a microarray analysis to examine the effects of CaMV integration into plant genomes on gene expression. For this analysis, three plants of each of CaMV-genome GM lines 316B and 318-1 were propagated, along with three plants of empty vector-transgenic line 320B, three plants of wild type C24-M3 infected with CaMV H7, and three non-infected wild type C24-M3, as control plants. RNAs were extracted from these plants at 20 dpi (at the onset of flowering). The RNAs were processed by the SCRI Microarray Unit for further analysis. These RNAs were of sufficient quality and quality to proceed, and they were dually labelled. Microarray chips were hybridised and processed with the data from the scanning of the chips subjected to statistical analysis by BioSS. The detailed analysis of the gene expression profiles is presented in Annex 2.

Of the 22,000 sequences representing genes on the Affymetrix ATH1-22K Genechip, 358 were altered in their expression pattern at the time of sampling as a result of either infection by CaMV or transformation with or without the CaMV genome. These 358 genes were subjected to cluster analysis. [See Fig. 14].

Based on the cluster analysis of statistically significant differences in gene expression profiles, the various gene expression profiles could be grouped by into nine clusters of responses:

Cluster 1 consisted of the 94 genes for which the expression pattern was increased due virus infection, but was reduced by transformation;

Cluster 2 consisted of 176 genes for which the expression pattern was increased due virus infection, but was reduced less or not all by transformation;

Cluster 3 consisted of 10 genes for which the expression pattern was increased due virus infection and slightly by CaMV integration, but was reduced slightly by transformation with the empty vector;

Cluster 4 consisted of 4 genes for which the expression pattern was increased due to both virus infection and transformation;

Cluster 5 consisted of 8 genes for which the expression pattern either was increased slightly due virus infection, but was reduced by transformation, or was decreased slightly due to virus infection and was decreased further by transformation;

Cluster 6 consisted of 2 genes for which the expression pattern was increased due virus infection, but showed different effects due to the various transformation events; i.e., expression of these genes also increased in the vector-only line and in line 316B, but not in line 318-1;

Cluster 7 consisted of 56 genes for which the expression pattern was decreased due virus infection, and either increased or decreased by transformation;

Cluster 8 consisted of 2 genes for which the expression pattern was decreased slightly or not affected by virus infection or transformation with the empty vector, but was increased by integration of CaMV;
Cluster 9 consisted of 6 genes for which the expression pattern was increased slightly due virus infection, but was reduced considerably by transformation.

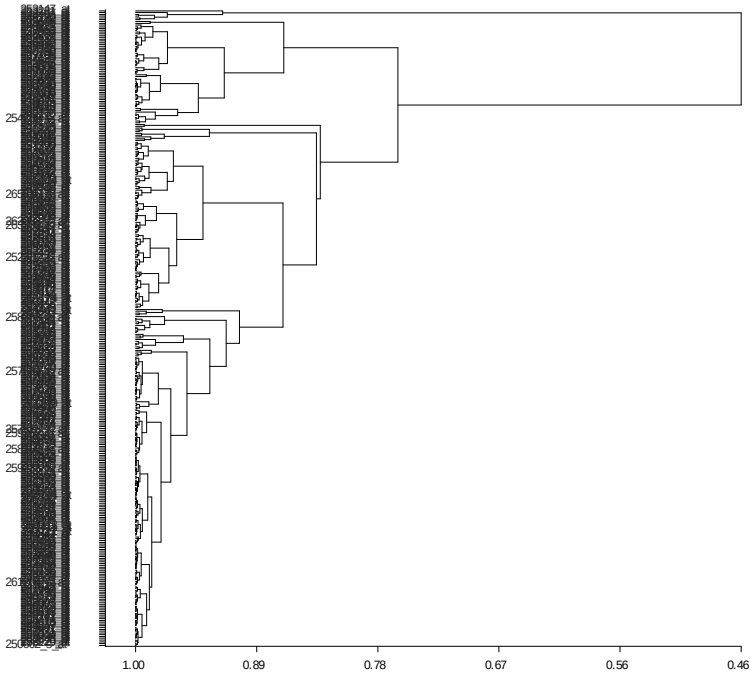

Fig. 14. Dendogram showing grouping of gene expression profiles by cluster analysis. Cluster analysis was done for the various types of samples (CaMV infection, transformation with empty vector, and two lines transformed with vector containing the CaMV genome) vs. the non-transformed, non-infected samples.

In examining the genes affected, many encode transcription factors, kinases, or environmental response proteins, but many are also encode proteins of unknown function or hypothetical proteins. Many of these genes also have been described previously as responding to environmental stresses, chemical inducers, or pathogen infection. These are found in all clusters except Cluster 4. None of the four genes upregulated by both CaMV infection and transformation (Cluster 4) appears to have been reported previously as altered by other elicitors or stresses, including after infection by several other viruses, as ascertained from interrogating DRASTIC database.

Transformation itself seemed to lead to significant reduced gene expression much more often than it led to significant increased gene expression with more than 85 % of the 358 genes showing reduced expression, whereas infection by CaMV increased the expression of many more genes than it repressed (see below). The effects on plant gene expression of integration of the entire CaMV genome led to few major changes not seen by integration of the empty vector (four genes in Clusters 6 and 8). Therefore, as the only CaMV gene expressed in the GM plants containing the integrated CaMV genome is gene VI, this gene product of the mildly virulent H7 strain of CaMV did not appear to alter gene expression much more than the empty vector. By contrast, the replication of CaMV during infection and the concomitant expression of the entire genome showed much greater effects on gene expression,

especially on the upregulation of gene expression with more than 80 % of the 358 genes showing increased gene expression. It is noteworthy that while CaMV infection led to gene expression for about half of the genes being increased by more than two-fold, transformation led to gene expression for about two-thirds of the genes being reduced by less than two-fold. After CaMV infection, the largest increase in gene expression was 44-fold and the largest decrease was 6-fold, while after transformation with the empty vector, the largest increase was 16-fold and the largest decrease was 16-fold.

Of the 358 genes altered in their expression, 144 are unknown with respect to their putative molecular function, while 179 are unknown with respect to a specific putative biological process. Of the remaining 214 genes for which a putative molecular function can be ascribed, 119 encode binding proteins and 115 encode some sort of catalytic activity (there is overlap between these groupings), while 23 genes encode proteins with transporter activity, 22 genes encode proteins with transcription regulatory activity, and six genes encode proteins with signal transducer activity. Of the 179 genes involved in various biological processes, 162 are involved in physiological processes, 152 are involved in cellular processes (again, there is overlap between these groupings) and 14 are specifically involved in plant development. [See Annex 2.]

Overall, integration by the CaMV genome into the genome of *A. thaliana* does not in itself affect vital functions, although it may affect the expression of specific genes. Finally, since CaMV does not have a mechanism to invade the meristem and become incorporated into seeds, it cannot show the effect we observed here on seed germination due to transformation of the germline tissues.

Objective 10. Determine whether CaMV infection affects expression and stability of a transgene.

The effects on suppression of transgene expression following CaMV infection of plants expressing transgenes driven by the CaMV 35S RNA promoter were evaluated based on data available from the literature (Al-Kaff *et al*., 1998; 2000). Those two studies showed that the CaMV 35S RNA promoter-mediated expression could be suppressed in one host species, which itself normally showed recovery (a form of suppression) from CaMV infection. However, this effect only occurred with a few of the CaMV strains in the host tested (oilseed rape). Moreover, a subsequent report has shown that this effect on suppression of transgene expression by CaMV infection of GM plants is also host-specific (Dale and Al-Kaff, 2006). Therefore, further work was done here to evaluate several hosts for such effects on gene expression driven by a 35S RNA promoter, as well as for the effect of CaMV integration on 35S-mediated foreign gene expression.

To determine whether CaMV infection itself affected the expression and stability of a transgene, four GM species (*A. thaliana*, *B. napus*, *N. benthamiana* and *N. tabacum*) were used in laboratory experiments, all expressing the GFP gene under the control of the CaMV 35S RNA promoter. Fifty plants of each of the above species were inoculated and the plants were incubated for either 30 days (*A. thaliana* and *N. tabacum*) or up to 90 days (*B. napus* and *N. benthamiana*). The above GM plants were assessed at 10-day intervals, relative to (10) non-inoculated control GM plants, for effects on gene expression or stability, by examination of the level of fluorescence of the GFP under a hand-held UV lamp in the first instance, and then

under a binocular microscope equipped with a camera capable of quantifying the fluorescence intensity. This experiment was then repeated with a second set of 50 plants for each species. Prior to inoculation all seedlings were screened first for GFP expression under UV, to eliminate any segregating variants that did not contain the GFP transgene.

Although there was some variation in the level of GFP expressed between individual plants during the time course of the experiment, as measured by fluorescence microscopy, in the case of *N. benthamiana*, *N. tabacum* and *B. napus*, this level of variation was not significantly different in mock-inoculated vs. CaMV-infected plants in both sets of experiments [Fig. 15 and data not shown]. Some silencing of GFP expression was seen in the oldest leaves, but this occurred in both mock-inoculated and CaMV-infected plants [Fig. 16A and data not shown], and was probably related to senescence. Localised silencing was observed in *N. benthamiana,* around viral-induced chlorotic lesions [Fig. 16B], but a generalised silencing of transgene expression was not seen [Figs. 15 and 16A].

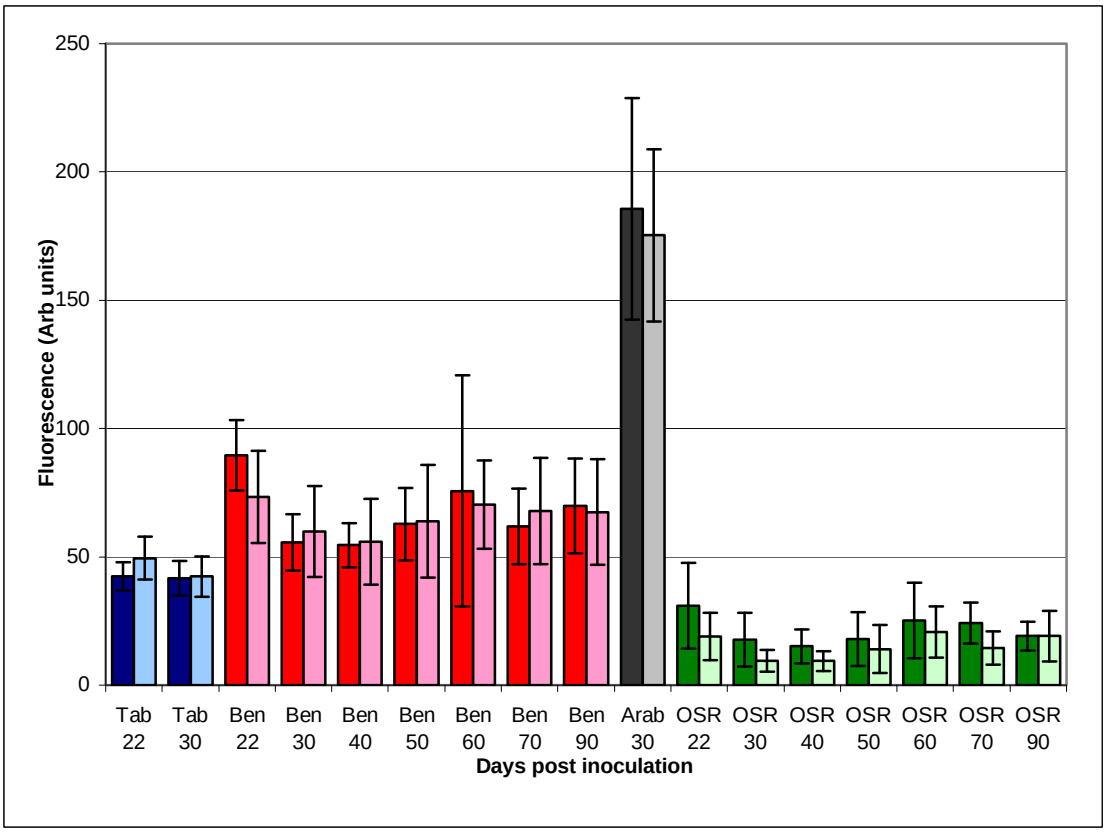

Fig. 15. **Mean fluorescence of 35S RNA promoted GFP plants measured at various days post inoculation.** Fluorescence was measured using a Leica Microscope, fitted with GFP3 filter, under U.V. light, 5x magnification and a Coolview camera. Images were analysed with ImagePro, using chip integration of 15 frames, and a gain setting of 18. *A. thaliana* [Arab] (grey), *N. benthamiana* [Ben] (red) and *B. napus* [OSR] plants (green) were inoculated with CaMV strain H7 and *N. tabacum* [Tab] (blue) plants were inoculated with CaMV strain W260. Mock inoculated plants shown as deep colour bars. Virus inoculated plants shown as lighter tint bars. Error bars are +/- standard deviation.

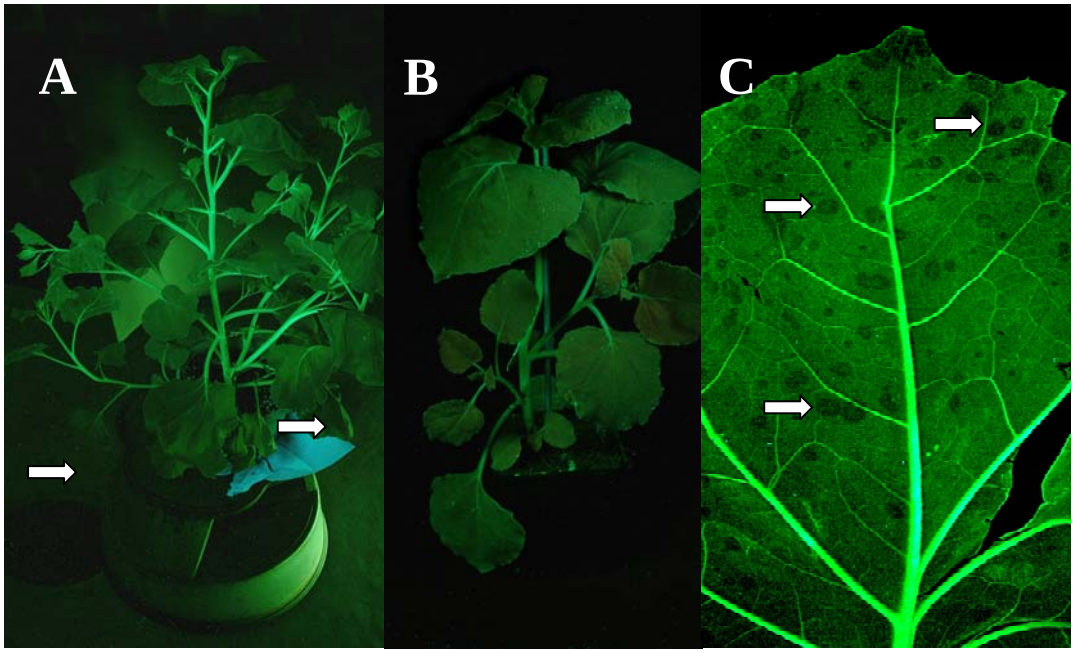

Fig. 16. *N. benthamiana* expressing 35S-driven GFP. (A) Plant showing silencing of GFP expression in the lower regions of the plant (arrows), 50 days post inoculation with CaMV H7 as compared to non-GM *N. benthamiana* plant (B). Plants were photographed under UV light with a Kodak Wratten 58 Tricolour Green Filter (under which green fluorescence to appear green). The arrows point to older leaves no longer expressing GFP. (C) Close up of a leaf viewed under UV light showing GFP silenced regions across the leaf at sites of chlorotic lesions (arrows), 90 days post inoculation with CaMV H7.

From the first experimental series, seeds of four *B. napus* and three *N. benthamiana* plants showing reduction in GFP fluorescence or unusual growth forms were germinated and their fluorescence again was screened. All seedlings exhibited a typical profile of GFP fluorescence, and did not show silencing (data not shown).

The experiments with *N. tabacum* showed that the limited movement of CaMV (W260) in the inoculated leaves did not lead to local or systemic silencing of the 35S RNA promoter-driven transgene [Fig. 15 and data not shown]. Seeds of six selected plants exhibiting a reduction in their fluorescence were germinated, and the seedlings were examined for silencing. Complete loss of fluorescence indicating gene silencing was not observed, but in some seedlings the fluorescence seemed to be reduced slightly. Final plant height also was recorded in *N. tabacum* with no significant difference observed between the W260-inoculated plants and mock-inoculated plants, (46.1 ± 12.1 cm vs. 57.9 ± 7.8 cm, mean height ± standard deviation, respectively). This has significance for GM plants which are not hosts of CaMV expressing transgenes driven by the CaMV 35S RNA promoter. That is, if infection was established in a limited number of cells in the inoculated tissue, then this would not be sufficient to cause silencing of the transgene in upper, non-infected leaves, even if local silencing had occurred. This is supported by the results from CaMV infection of *N. benthamiana* plants, where the only transgene silencing associated with infection that was observed occurred in local lesions in upper leaves induced by virus infection [Fig. 16B], which did not lead to a generalised silencing of transgene expression in those leaves [Fig. 16A].

In the case of *A. thaliana*, infection by CaMV strain H7 killed 10 of the 50 plants from the first experiment and led to severe symptoms with production of flower

stalks, flowers and siliques in most of the remaining 40 CaMV-infected plants. In two of these latter plants, silencing of GFP expression was observed in the rosette and cauline leaves and flower stalks [data not shown]. For the surviving plants, there was no other difference observed in fluorescence recorded with the binocular microscope and camera, but there was individual plant variation [Fig. 15]. Silencing was observed in the lower parts of the plant, presumably due to senescence, as this was seen in both mock-inoculated and virus-infected plants [Fig 17]. The surviving *A. thaliana* plants were assessed for suppression of transgene mRNA expression vs. modification of the transgene, by examining expression of transgene expression in the next generation. Seeds (10) obtained from the first experiment of all the *A. thaliana* plants remaining alive were sown and germinated and the seedlings were screened for fluorescence. All showed fluorescence, indicating that any silencing observed was not carried through to the next generation. This indicates that any such effect was transient in the infected population, and was not affecting the stability of the transgene in subsequent populations.

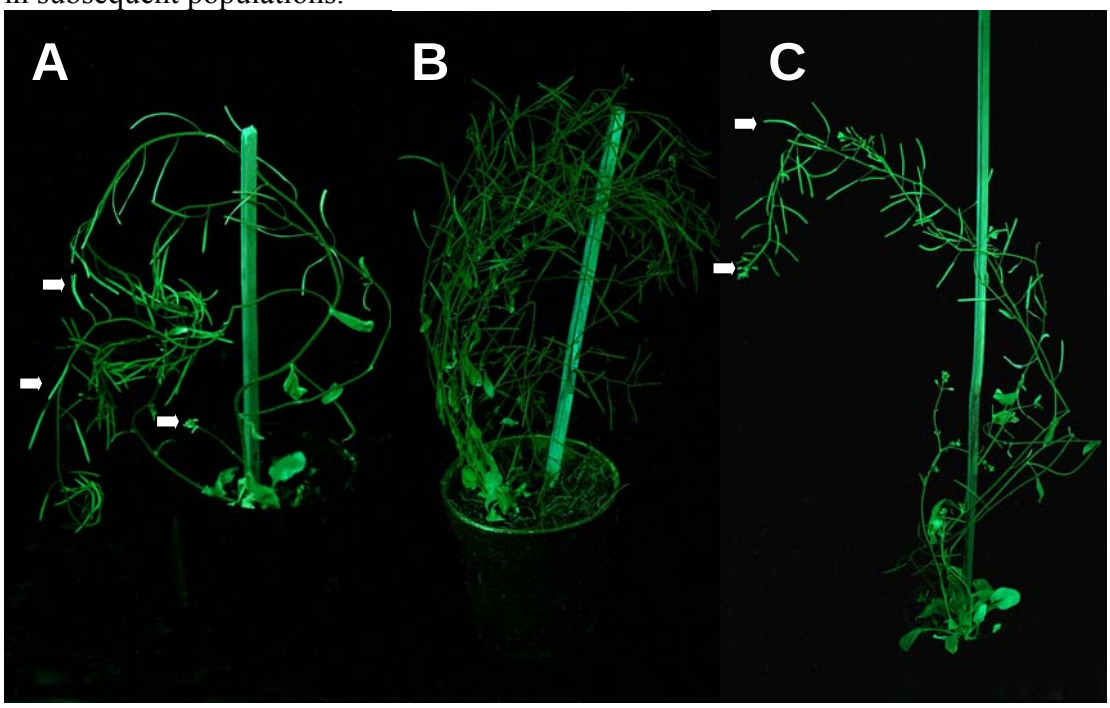

Fig. 17. *Arabidopsis thaliana* expressing 35S-driven GFP. Plants were photographed under UV light with a Kodak Wratten 58 Tricolour Green Filter, 30 days post inoculation with CaMV H7 (C) or mock inoculated (A). Lower regions of both plants showed GFP silencing, with siliques and new flower initials glowing (arrows), as compared with non-GM *A. thaliana* plants (B).

In the second experiment involving *A. thaliana*, infection by CaMV strain H7 killed eight of the 50 plants, but led to severe symptoms with production of flower stalks, flowers and siliques in most of the remaining 42 CaMV-infected plants. In one of these latter plants, silencing of GFP expression was observed in the rosette and cauline leaves and flower stalks, with only the basal root expressing GFP, (data not shown). Two of these plants have shown atypical growth forms, one being extremely stunted, with very hairy leaves [Fig 18A]. The second plant exhibited thickened stems and delayed flowering, typical of the effect of the *clavata* gene [Fig 18B]. For the remaining 40 plants, there was no other difference observed in fluorescence recorded with the binocular microscope and camera, but there was individual plant

variation, as seen for the first series [Fig. 15]. Silencing was observed in the lower parts of the plant, again presumably due to senescence, as this was seen in both mock-inoculated and virus-infected plants. These plants were assessed for suppression of transgene mRNA expression vs. modification of the transgene, by examining expression of transgene expression in the next generation. Fifteen *A. thaliana* seedlings were germinated to follow plant development right through. The plants developed normally and showed normal GFP fluorescence, seed set and growth.

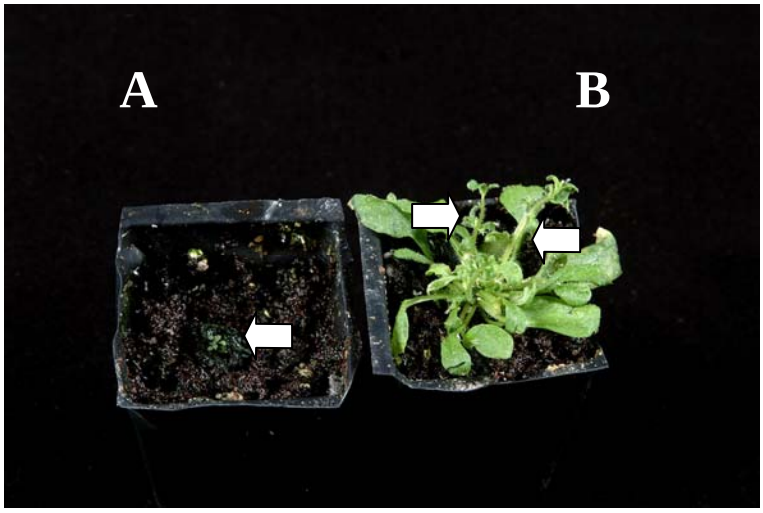

Fig. 18. *Arabidopsis thaliana* expressing 35S-driven GFP, 30 days post inoculation with CaMV H7. (A) Plant exhibiting reduced growth form and very hairy leaves (arrow). (B) Plant exhibiting flattened broadened stems, with distorted cauline leaves and new flower initials (arrows), typical of *clavata* gene expression. Plants were photographed under white light.

Seeds of selected plants of *N. benthamiana* and *B. napus*, which showed suppression of GFP from the second experimental series, were planted and the seedlings were examined for their GFP fluorescence profile. Once again, there was no indication of any silencing that was vertically transmissible.

In summary, it is concluded that infection by CaMV may affect expression of a transgene in very limited circumstances or tissues, while the stability of the transgene is not affected in subsequent generations.

The lack of silencing observed in *B. napus* may have been due to the fact that only some strains of CaMV cause gene silencing effect in this host (Al-Kaff *et al.*, 2000). That is, 67 strains of CaMV were assessed in that study for their virulence on *B. napus*, and the extent to which recovery from infection occurred, giving rise to six symptom types. One strain of each symptom type was tested for the ability to suppress the expression of a transgene controlled by a CaMV 35S RNA promoter, but only four of the six strains did suppress transgene expression, regardless of whether the plants recovered partially or completely from infection. However, since most of the strains were not tested, it is not known to what extent the groupings based on pathogenicity in *B. napus* had anything to do with the ability to suppress transgene expression. Al-Kaff *et al*. (2000) did not test the strains we have used in this study. However, our data showed that strain H7 (derived from CM1841 – a mild strain) did not induce systemic silencing of the GFP transgene in *A. thaliana, B. napus* or *N. benthamiana*, while CaMV strain W260, which was severe on turnip but did not infect *N. tabacum* systemically, did not induce systemic silencing in *B. napus*. Moreover, recent data from Dale and Al-Kaff (2006) showed that the gene silencing

observed by their strain of CaMV occurred in oilseed rape (*B. napus*), but not in pakchoi (*B. rapa*). Therefore those studies demonstrated that there is both a CaMV strain and host genotype component determining whether a given GM plant species can be subject to silencing of the transgene, while our results show that one strain that did not silence transgene expression in one GM species also did not do so in three others.

Objective 11. Assess whether transcription, reverse transcription (RT) or episomal replication occurs in CaMV-integrated plants.

The potential consequences of integration of CaMV sequences have been evaluated previously from two published studies (Shewmaker *et al*., 1985; Young *et al*., 1987). Those studies showed that gene expression from the CaMV 19S RNA promoter could occur in such plants, but that episomal expression did not occur when only a single copy of the CaMV genome was integrated. However, as that integration was limited to galls on plant, the consequences of other stresses on the potential for episomal replication were not evaluated.

The GM plants containing an integrated CaMV genome (line 316B) were screened initially for transcription and expression of gene VI, since gene VI and its 19S RNA promoter were expected to be active in the GM plants. Western blotting using an antiserum to the CaMV gene VI product showed the presence of the expected p62 protein [Fig. 19A, lanes 10 and 11]. This demonstrated that transcription could occur from the CaMV 19S RNA promoter present in the integrated copy of the CaMV genome and that the encoded p62 protein was expressed.

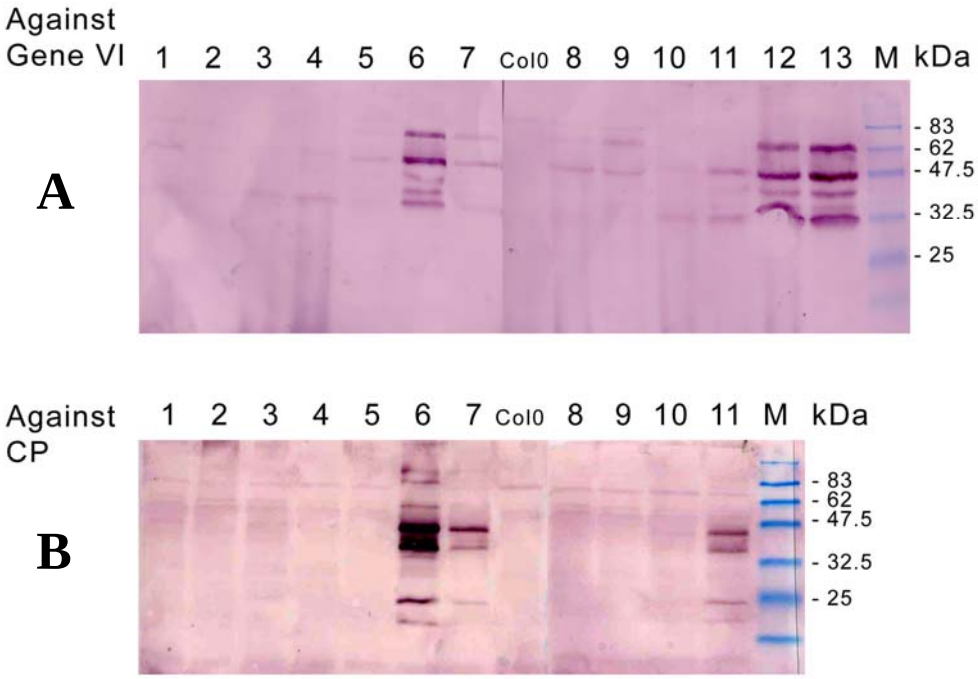

Fig. 19. Western blots for detection of the CaMV gene VI (A) or gene IV (coat protein - CP) (B) in plants stressed or not stressed. Samples: 1-3, CaMV-GM plants exposed to high light intensity; 4-6, CaMV-GM plants infected by CMV; 7-9, CaMV-GM plants exposed to drought; 10-11, CaMV-GM plants not stressed; 12-13, CaMV gene VI expressing GM plants (D4-2); Col-0, non-GM plant; M = mol. wt. markers. [The expressed p57 CP is normally processed *in vivo* to p44, p39 and p37.]

The ability of CaMV to replicate in the cytoplasm of plants cells and not as part of the plant genome is defined as episomal replication. To detect episomal replication of CaMV, western blot analysis also was done initially using an antiserum to the coat protein of CaMV. Subsequently, using primers that spanned the gene II - gene III junction, the PCR was used to confirm the presence of CaMV DNA replicating episomally, which occurs in the inclusion body in the cytoplasm and not in the nucleus. [The transgene construct splits gene II and therefore only episomal, replicating CaMV would be able to yield a PCR product.] This was done using the CaMV-GM line 316B alongside GM line D4-2 plants (which expressed CaMV gene VI), and control plants (non-GM wildtype C24-M3, wildtype Col-0 plants and empty-vector GM plants). Initially, the plants were examined at the development stage approaching flowering. In a second experiment, they were examined at flowering, at seed set, and again at senescence.

In the first experiment, episomal DNA replication occurred in several of the plants tested, although due to the mild nature of the strain, "disease" (i.e., symptoms) was not seen in these plants. Episomal replication was identified initially by western blotting for the presence of CaMV coat protein, which could only be produced by episomal infection [Fig. 19B, lane 11]. At the same time, these plants should also express CaMV gene VI, either from the 19S RNA promoter present in the integrated CaMV genome, or from the 19S RNA promoter present in the episomal CaMV DNA. The gene VI product also was detected by western blotting from the same plants [see Table 7]. Subsequently, the presence of the episomal, replicating CaMV DNA was confirmed by PCR to show the presence of an intact gene II - gene III junction [see Fig. 20], present in the virus and not the transgene [see Tables 7 and 8]. Then, immune-specific electron microscopy (ISEM) was used to show that typical CaMV particles were present in these plants. Finally, the PCR products were sequenced, and the exact positions of the recombination events as well as the sequences at the recombination junctions were determined (see below). The results will be discussed below together with those derived from GM plants containing an integrated CaMV genome that were subjected to various stresses (Objective 12).

Table 7. Expression of episomal replication following stress treatment of the plants for 30 days to the start of flowering.

	Light			**CMV-infected**			**Drought**			**No Stress**		**Line D4-2**		**Col-0**
Sample no.	1	2	3	1	2	3	1	2	3	1	2	1	2	1
Gene IV[a]	-	-	-	-	-	+++	++	-	-	+	++	ND	ND	-
Gene VI[a]	+	-	+	+	+	+++	++	+	++	+	++	+++	+++	-
Genes II-III[b]	+	+	++	+	+	++	++	-	+	++	++	-	-	-

All plants are line 316B (transgenic for full length insert split within gene II) with the exception of D4-2 (transgenic for expression of gene VI) and non-transgenic type Col-0. [a] Detection by western blot; [b] Detection by PCR. +++ Strongly positive signal; ++ positive signal; + weak but discernibly positive signal; — negative; ND not determined.

In the second experiment [see Table 8], additional sampling times (seed set and senescence) and repeated sampling of plants were done (plants 1-4, 11-14, 21-24, and 31-34 were sampled each time, while samples 5-8, 15-18, 25-28 and 35-38 were sampled only at both seed set and senescence and 9-10, 19-20 and 29-30 were sampled only at senescence) to determine if episomal replication was random, or was due to wounding of the plants during sampling for confirmation of transgene insertion. In this experiment, western blots of the gene IV and gene VI products as well as the PCR for the gene II – gene III junction were used to evaluate episomal replication.

Table 8. Expression of episomal replication following stress treatment of plants.

Sample	Stress	Flower Gene IV	Flower Gene VI	Flower PCR	Seed Gene IV	Seed Gene VI	Seed PCR	Senesce Gene IV	Senesce Gene VI
1	Control	-	+	+	+	+	+	+	+
2	Control	-	-	+	+	+	+	+	+
3	Control	-	+	+	+	+	ND	?	+
4	Control	-	+	+	+	+	ND	+	+
5	Control				+	+	+	-	+
6	Control				-	-	+	-	+
7	Control				+	+	+	+	+
8	Control				+	+	+	+	+
9	Control							+	+
10	Control							+	+
11	Light	+	+	+	-	+	-	- D	-
12	Light	-	+	-	-	+	-	- D	-
13	Light	-	+	-	-	-	-	- D	-
14	Light	-	+	-	+	+	-	- D	-
15	Light				-	+	-	-	-
16	Light				-	+	-	-	-
17	Light				-	+	-	-	+
18	Light				+	+	-	+	+
19	Light							+	+
20	Light							+	+
21	Heat	-	+	-	+	+	-	+	+
22	Heat	-	+	+	+	+	+	+	+
23	Heat	-	+	+	+	+	+	NS	NS
24	Heat	+	+	+	-	+	-	NS	NS
25	Heat				+	+	-	-	+
26	Heat				+	+	-	-	+
27	Heat				NS	NS	NS	- D	-
28	Heat				NS	NS	NS	- D	-
29	Heat							+ D	-
30	Heat							+ D	-
31	Drought	-	-	+	-	+	+	-	+
32	Drought	-	+	+	+	+	+	+	+
33	Drought	+	+	-	+	+	-	+	+
34	Drought	+	+	-	+	+	-	+	+
35	Drought				+	+	+	+	+
36	Drought				+	+	+	+	+
37	Drought				-	+	+	-	+
38	Drought				-	+	+	-	+
39	CMV	+	-	-	+	+	-	+	+
40	CMV	+	-	-	NS	NS	NS	NS	NS
41	CMV	-	-	-	NS	NS	NS	NS	NS
42	CMV	-	+	+	+	+	-	+	+
43	D4-2	-	+	ND	-	+	ND	-	-
44	Empty Vector	-	-	ND	-	-	ND	-	-
+CaMV	Virus	+	++		+	+ +		+	++

Plant selections were assessed at flowering, seed set and/or senescence (Senesce), by either western blotting for viral-encoded proteins (Gene IV and Gene VI) or the PCR for Gene II – Gene III. ND = not determined; NS = not sampled (plant died); + = discernable band present; - = no band present; ? = band not discernable by western analysis; D = plant dead but still sampled.

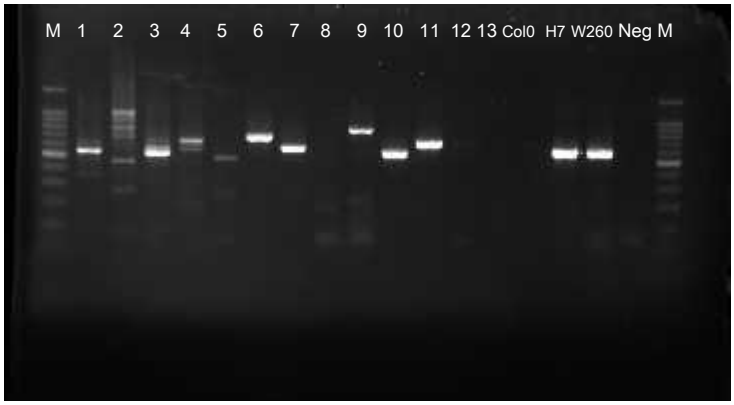

Fig. 20. **Agarose gel showing PCR products of the CaMV gene II through gene III junction region, of episomally replicating CaMV DNAs from plants subjected to various environmental stresses.** Lanes 1-3 = light stress, 4-6 = CMV-infection stress; 7-9 = drought stress; 10, 11 = No stress; 12, 13 = D4-2 gene VI transgenic plants. H7 = CaMV-H7 infected material; W260 = CaMV-W260 infected material; Col-0 = Non infected wild type plant. Neg = negative PCR control (no template added). The expected size of PCR product is 600bp. M = 100bp Promega Markers.

These results confirmed that episomal replication did occur in some of the plants and that this was not a consequence of wounding due to prior sampling. The results will be discussed further in Objective 12.

In a separate experiment, plants of line 318-1 also were tested for the presence of episomal replication of the integrated CaMV. Here again, episomal replication of CaMV was detected by western blotting in some of the plants [see Table 9 and Fig. 21], demonstrating that the generation of episomally replicating virus was not unique to line 316B.

Table 9. Expression of episomal replication in line 318-1

Line 318-1	Detection by	Drought (+)	No stress (+)
At 25 Days	PCR (II-III)	0 of 2	3 of 5
	CP	0 of 2	3 of 5
	VI	1 of 2	2 of 5
At Flowering	PCR (II-III)	1 of 2	4 of 4
	CP	2 of 2	4 of 4
	VI	0 of 2	4 of 4

All plants are line 318-1 (transgenic for full length insert split within gene II). (+) = the number of plants for which a positive reaction was obtained is indicated, out of the number of plants tested. One of the infected non-stressed plants had died before flowering.

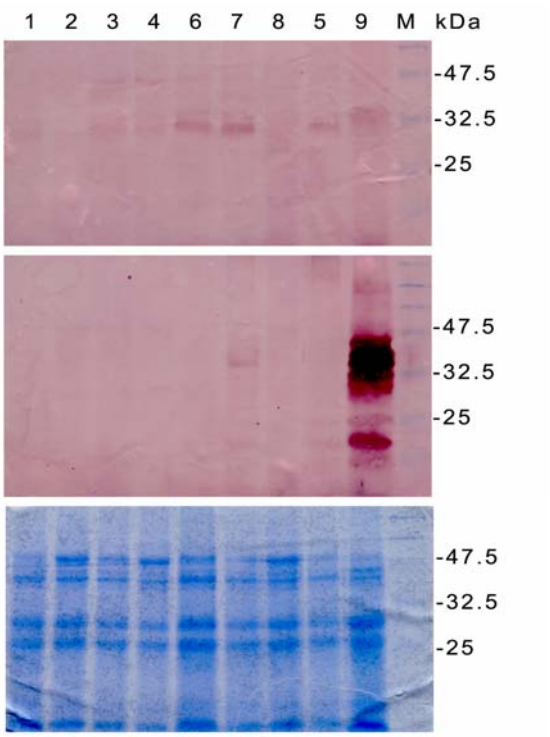

Fig. 21. Western blots for detection of (A) the CaMV gene VI, or (B) gene IV (coat protein - CP) in plants stressed or not stressed. (C) Coomassie blue-stained gel to show total protein loading. Samples: 1-2, CaMV genome GM-plants exposed to drought; 3-7, CaMV genome GM-plants not stressed; 8, C24-M3 non GM-plant; 9, turnip Infected with CaMV H7; M = mol. wt. markers.

Objective 12. Determine effects of environmental conditions on transcription, reverse transcription, or episomal disease in CaMV integrated plants.

It was thought that perhaps stressing plants might lead to or increase transcription, reverse transcription, or episomal disease from the integrated CaMV genomes where normal growth conditions did not do so, based on experiences with other integrated and episomally replicating viruses (reviewed by Harper *et al.*, 2002; Richert-Pöggeler *et al.*, 2003). Therefore, specific changes in environmental conditions that were examined were those likely to place considerable stress on the plant that could facilitate such expression. These included drought, strong light, heat, and infection by the unrelated virus cucumber mosaic virus (CMV). The plants subjected to higher light intensities were grown at 22 ºC day /18 ºC night, but under 450 µmol m^{-2} s^{-1} vs. 250 µmol m^{-2} s^{-1} for the other plants. The plants subjected to temperature stress were grown at 32 °C day and night. The plants subject to drought stress were watered every other day, while all other plants were watered twice a day. The plants infected with CMV were inoculated with a buffered extract taken from a CMV-infected *N. benthamiana*.

The expression of CaMV genes IV and VI as well as episomal replication in plants were assessed, following stress treatment for 30 days. These experiments were done at the same time as those described in Objective 11, when the plants were not stressed environmentally. Most of the CaMV-GM plants were from line 316B (transgenic for a full length insert split within gene II) [Tables 7 and 8 and Fig. 19], although in one experiment plants of line 318-1 were also assessed for effects of

drought stress [Table 9 and Fig. 21]. In addition, plants of line D4-2 (transgenic for expression of CaMV gene VI) and non transgenic ecotype Col-0 plants were tested. The results [see Tables 7-9] indicated that transcription and episomal replication were seen on plants stressed by high temperature, high light intensity, drought, or infection by CMV. However, it did not appear that there was an effect specific to any of these stresses, since the control plants also showed CaMV transcription and episomal replication.

The numbers of samples assessed were not sufficient in the first experiment [Table 7] to ascertain whether any particular stress treatment might increase the proportion of plants that showed episomal replication. Similarly, that test was not designed to ascertain whether a particular stress stimulated episomal CaMV replication to occur earlier. However, it can be noted that while some samples did not show the presence of CaMV coat protein, they did show the presence of the PCR product, indicating that those plants must not have accumulated sufficient coat protein at the time of sampling. [See Tables 7 and 8. Note: The same sample extracts were used for western blotting and PCR analysis.]

Selected plants showing episomal replication were characterised further by EM visualisation of virus particles [Fig. 22] and by cloning and sequencing of the PCR products [See Annex 3].

ISEM was performed on E316B tissues testing positive for episomal replication using immuno-capture onto antibody-labelled grids, and these confirmed the presence of spherical particles of size appropriate for CaMV (50nm). These particles were seen for samples E316-6, E316-7, E316-10 and E317-11. [Fig. 22 and data not shown.]

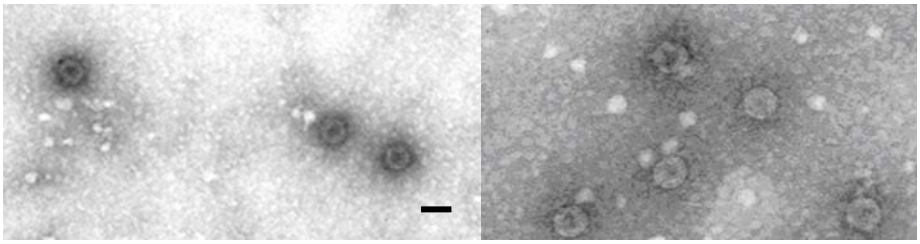

Fig. 22. Electron micrographs of CaMV particles. Anti-CaMV particle antibody was used to coat EM grids. Virus extract were made using phosphate buffer pH 6.5, and a standard antibody-coated grid protocol was used. The grids were stained with ammonium molybdate (AM) pH 6.5, or 2% uranyl acetate (UA). Images are samples from control CaMV-H7 infection on turnip (left) and the plant containing CaMV E316-10 (smaller PCR product from Fig. 20) on *A. thaliana* (right). Bar = 50 nm.

The various PCR products obtained were of slightly different sizes as obtained from the various plants analysed [Fig. 20 and data not shown], with different molecules containing various deletions in gene II and some also losing part or nearly all of gene III. Moreover, most contained some additional sequences [See Annex 3]. Alignments of the sequences indicated that these additional sequences were from the T-DNA in the integrated portion of the plasmid vector DNA. Apparently, this sequence was recruited to facilitate the circularisation of the genome enabling production of episomally replicating virus genomes.

Specifically, the data from line 316B plants showed the following: The sequence of the DNA from clone E316-7 (from a drought stressed plant) is the same as for CaMV strain H7, except for a 6bp deletion (TGATTA) at nucleotide 79 and a 12 nt insertion vector DNA (T-DNA) at position 196. Otherwise it matches the H7

strain DNA sequence. Therefore, the virus in E316-7 potentially circularised through this point using the *Xho*I site and the small 12 nt stretch at position 196.

The sequence of the DNA from clone E316-10 (from a non-stressed plant) has a substitution from the *Xho*I site (nucleotide 90) through to the end of gene II. This has led to a replacement of 278 nt of CaMV gene II with 135 nt of vector DNA sequence, giving a reduction in size of 143. This was due to removal of the 3' end of the gene II after the *Xho*I site. The genome then continues with gene III.

The sequence of the DNA from clone E316-11 (also from a non-stressed plant) again flows normally up to the *Xho*I site, and it also has an insertion of vector DNA through to a partial duplication of the *Xho*I site, although this is a shorter insertion at 64 nt. The genome then continues with gene III.

The sequence of the DNA from clone E316-6 (from a CMV-infected plant) again has the insertion of vector DNA after the *Xho*I site in the gene II region. It has a combination of 97 nt of right, flanking vector DNA (of the 135 nt seen in E316-10) and 22 nt of left, flanking vector DNA (9 nt longer than found in E316-7). The addition is across the region nucleotides 90–208. This has increased the genome size by approximately 120 nt and also duplicated the *Xho*I site.

By contrast, the data from line 318-1 plants showed the following: (a) Most samples showed only minor variation in sister clones of the same PCR products, with two exceptions [E318-4Fma,b and E318-6Fa,b]. (b) The sequence of PCR products taken from samples before flowering had larger deletions than those take from the same plants after flowering [E318-4 vs. E318-4Ft/m/b, and E318-5 vs. E318-5F]. (c) Some of the virus generated in the plants after flowering may have been generated by recombination of the defective virus (lacking large parts of genes II and II) with the integrated CaMV genome, although these also did not generate wild type virus [E318-4 and the subsequent sample yielding clones 318-4Ft/m-a,b/b, as well as E318-5 and the subsequent sample yielding clones E318-5Fa,b]. (d) Some of the viruses found after flowering represented new recombination-excision events, as they contained additional T-DNA or the unknown sequences not seen in the virus sequence amplified earlier [E318-4Ft and the two clones of E318-4Fm vs. E318-4]. And (e), several of the clones [E318-4, E318-5 and E318-7] were lacking part or almost all of gene III and therefore these viruses would not be able to move cell-to-cell, unless complemented by the gene III product from a low level of virus in the same cells that did contain gene III and would represent an independent recombination-excision event. However, we have no evidence supporting this scenario.

It was conceivable that the stress induced by wounding (i.e., the harvesting of leaves prior to stress application to confirm the presence of the CaMV transgene) could have induced episomal replication rather than the particular environmental or disease applied stress. [Note: All of the GM plants should have produced the gene VI product, since in the CaMV genome-transgenic plants, gene VI is expressed under the control of its own (19S) subgenomic RNA promoter. The assay for the gene VI p62 protein was done using the same protein extracts made for assaying the coat protein produced by gene IV.] Therefore, the second experiment with line 316B plants was undertaken involving more seeds, to determine whether environmental factors did show some effect, if the plants were not previously selected on antibiotic plates and stressed by wounding to assess their genetic composition. In this case [Table 8], plants were sampled at each of the major developmental stress times (flowering, seed set and senescence). This was done by repeat sampling of the plants through time and sampling of fresh plants at each time point, to assess if the episomal replication was random or initiated by plant wounding; i.e., removing leaves to sample for analysis.

The results indicated that removing leaves for sampling did not induce episomal replication. On the other hand, the applied stresses caused only low levels of episomal replication at the earliest time point, while either developmental changes associated with seed production or simply aging appeared to stimulate episomal replication in most of the plants. It is also possible that it is simply a probability event and the older the plants get, the greater the chance of recombination-excision occurring.

A number of samples in Table 8 showed the presence of capsid protein, but did not test positive by the PCR. These were not examined further; however, given that sequences of gene II and III were subject to deletion in the episomally replicating virus [See Annex 3], it is possible that the binding site for the gene II primer or gene III primer used to screen for episomal replication had been deleted in these samples.

In summary, episomal replication of CaMV genetically engineered into the *A. thaliana* genome occurred, and it occurred at a similar frequency whether the plants were subjected to various environmental stresses including infection by another virus. However, since integration of a fragment of the CaMV genome occurred at a very low frequency in non-GM infected plants in this study and does not occur in the germline, the ability of integrated CaMV to excise its genome from the plant genome and again replicate episomally has little consequence. This is because the plants are already infected by CaMV. In fact, it is conceivable that in an infected plant the rate of excision may be reduced due to crossprotection. However, it is clear that integration of the CaMV genome itself does not make the plants resistant to infection by CaMV, since those plants were not resistant to the generation of episomally replicating virus. Moreover, if integration did occur into the 35S RNA promoter of GM plants containing a CaMV 35S RNA promoter, which we did not detect in 3856 samples examined here (Objective 4), then the integrated virus would excise from the genome by homologous recombination from the genome with only flanking viral (35S RNA promoter) sequences and no flanking host sequences. Finally, we cannot exclude the possibility that the high incidence of non-homologous recombination (not frequency, since it is unknown how many cells per plant actually showed this recombination even leading to episomal replication) we observed leading to CaMV excision was due to the nature of the flanking sequences here being derived from the T-DNA bordering sequences, which themselves are involved in the recombination events that lead to plant transformation.

References

Al-Kaff, N.S., Covey, S.N., Kreike, M.M., Page, A.M., Pinder, R. and Dale, P.J. (1998). Transcriptional and posttranscriptional gene silencing in response to a pathogen. *Science* **279,** 2113-2115.

Al-Kaff, N.S., Kreike, M.M., Covey, S.N., Pitcher, R., Page, A.M. and Dale, P.J. (2000). Plants rendered herbicide-susceptible by cauliflower mosaic virus-elicited suppression of a 35S promoter-regulated transgene. *Nature Biotechnol*. **18,** 995-999.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403-410.

Ashby, A., Warry, A., Bejarano, E., Khashoggi, A., Burrell, M. and Lichtenstein, C. (1997). Analysis of multiple copies of geminiviral DNA in the genome of four closely related *Nicotiana* species suggest a unique integration event. *Plant Mol. Biol*. **35,** 313-321.

Assaad, F.F. and Signer, E.R. (1990). Cauliflower mosaic virus P35S promoter activity in *Escherichia coli*. *Mol. Gen. Genet.* **223,** 517-520.

Ballas, N., Broido, S., Soreq, H. and Loyter, A. (1989). Efficient functioning of plant promoters and poly(A) sites in *Xenopus* oocytes. *Nucl. Acids Res.* **17,** 7891-7903.

Bejarano, E.R., Khashoggi, A., Witty, M. and Lichtenstein, C. (1996). Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. *Proc. Natl. Acad. Sci. USA* **93,** 759-64.

Brunt, A.A., Crabtree, K., Dallwitz, M.J., Gibbs, A.J., Watson, L. and Zurcher, E.J. (eds.) (1996 onwards). `Plant Viruses Online: Descriptions and Lists from the VIDE Database. Version: 20th August 1996.' URL http://biology.anu.edu.au/Groups/MES/vide/ Description for Caulimoviruses.

Clough, S.J and Bent, A.F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16,** 735-743.

Cullen, D.W., Lees, A.K., Toth, I.K. and Duncan, J.M. (2001). Conventional PCR and real-time quantitative PCR detection of *Helminthosporium solani* in soil and on potato tubers. *Eur. J. Plant Pathol.* **107,** 387-398.

Dale, P.J. and Al-Kaff, N.S. (2006). http://www.defra.gov.uk/science/project_data/DocumentLibrary/CB02017/CB02017_3425_FRP.doc

Franck, A., Guilley, H., Jonard, G., Richards, K. and Hirth, L. (1980). Nucleotide sequence of cauliflower mosaic virus DNA. *Cell* **21,** 285-294.

Gal, S., Pisan, B., Hohn, T., Grimsley, N. and Hohn, B. (1991). Genomic homologous recombination *in planta*. *EMBO J.* **10,** 1571-1578.

Gal, S., Pisan, B., Hohn, T., Grimsley, N. and Hohn, B. (1992). Agroinfection of transgenic plants leads to viable cauliflower mosaic virus by intermolecular recombination. *Virology* **187,** 525-533.

Geering, A.D.W., Olszewski, N.E., Harper, G., Lockhart, B.E.L., Hull, R. and Thomas, J.E. (2005a). Banana contains a diverse array of endogenous badnaviruses. *J. Gen. Virol.* **86,** 511-520.

Geering, A.D.W., Pooggin, M.M., Olszewski, N.E., Lockhart, B.E.L. and Thomas, J.E. (2005b). Characterisation of Banana streak Mysore virus and evidence that its DNA integrated into the B genome of cultivated *Musa*. *Arch. Virol.* **150,** 787-796.

Gracia, O. and Shepherd, R.J. (1985). Cauliflower mosaic virus in the nucleus of *Nicotiana*. *Virology* **146,** 141-145.

Gregor, W., Mette, M.F., Staginnus, C., Matzke, M.A. and Matzke, A.J. (2004). A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. *Plant Physiol.* **134,** 1191-1199.

Halfhill, M.D., Richards, H.A., Mabon, S.A. and Stewart Jr., C.N. (2001). Expression of GFP and Bt genes in *Brassica napus* and hybridization with *Brassica rapa*. *Theor. Appl. Genet.* **103,** 659-667.

Harper, B.K. and Stewart Jr., C.N. (2000). Patterns of green fluorescent protein expression in transgenic plants. *Plant Mol. Biol. Rep*. **18,** 121a-141i.

Harper, G., Osuji, J.O., Heslop-Harrison, J.S. and Hull, R. (1999). Integration of banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. *Virology* **255,** 207-213.

Harper, G., Hull, R., Lockhart, B. and Olszewski, N. (2002). Viral sequences integrated into plant genomes. *Annu. Rev. Phytopathol.* **40,** 119-136.

Haseloff, J., Siemering, K.R., Prasher, D.C. and Hodge, S. (1997). Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proc. Natl. Acad. Sci USA* **94,** 2122-2127.

Ho, M.-W., Ryan, A. and Cummins, J.(1999). Cauliflower mosaic viral promoter - a recipe for disaster? *Microb. Ecol. Health Dis.* **11,** 194-197.

Ho, M.-W., Ryan, A. and Cummins, J. (2000). Hazards of transgenic plants containing the cauliflower mosaic virus viral promoter. *Microb. Ecol. Health Dis.* **12,** 6-11.

Hull, R. (1984). Caulimovirus group. *In* CMI/AAB Description of Plant Viruses No. 295. Unwin Bros. Ltd., Old Woking, Surrey, UK.

Hull, R., Covey, S.N. and Dale, P. (2000a). Genetically modified plants and the 35S promoter: assessing the risks and enhancing the debate. *Microb. Ecol. Health Dis*. **12,** 1-5.

Hull, R., Lockart, B.E., Reddy, D.V.R. and Schoelz, J.E. (2000b). Family *Caulimoviridae*. In: *Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses* (van Regenmortel, M.V.H., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B., eds), pp. 335-347, Academic Press, San Diego.

Jakowitsch, J., Mette, M.F., van der Winden, J., Matzke, M.A. and Matzke, A.J.M. (1999). Integrated pararetroviral sequences define a unique class of dispersed repetitive DNA in plants. *Proc. Natl. Acad. Sci. USA* **96,** 13241-13246.

Kohli, A., Griffiths, S., Palacios, N., Twyman, R.M., Vain, P., Laurie, D.A. and Christou, P. (1999). Molecular characterization of transforming plasmid rearrangements in transgenic rice reveals a recombination hotspot in the CaMV 35S promoter and confirms the predominance of microhomology mediated recombination. *Plant J.* **17,** 591-601.

Lheureux, F., Carreel, F., Jenny, C., Lockhart, B.E. and Iskra-Caruana, M.L. (2003). Identification of genetic markers linked to banana streak disease expression in inter-specific *Musa* hybrids. *Theor. Appl. Genet.* **106,** 594-598.

Lockhart, B.E., Menke, J., Dahal, G. and Olszewski, N.E. (2000). Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *J. Gen. Virol.* **81,** 1579-1585.

Lovisolo, O., Hull, R. and Rösler, O. (2003). Coevolution of viruses with hosts and vectors and possible palaeontology. *Adv. Virus Res.* **62,** 325-379.

Mette, M.F., Kanno, T., Aufsatz, W., Jakowitsch, J., van der Winden, J., Matzke, M.A. and Matzke, A.J.M. (2002). Endogenous viral sequences and their potential contribution to heritable virus resistance in plants. *EMBO J.* **21,** 461-469.

Morel, J.B. and Tepfer, M. (2000). Pour une evaluation scientifique des risques: le cas du promoteur 35S. *Biofutur* **201,** 32-35.

Ndowora, T., Dahal, G., LaFleur, D., Harper, G., Hull, R., Olszewski, N.E. and Lockhart, B. (1999). Evidence that badnavirus infection in *Musa* can originate from integrated pararetroviral sequences. *Virology* **255,** 214-220.

Noreen, F., Akbergenov, R., Hohn, T. and Richert-Pöggeler, K.R. (2007). Distinct expression of *Petunia vein clearing virus* and the DNA transposon *dTph1* in twp *Petunia hybrida* lines is correlated with differences in histone modification and siRNA production. *Plant J.* **50,** 219-229.

Pobjecky, N., Rosenberg, G.H., Dinter-Gottlieb, G. and Kaufer, D. (1990). Expression of the beta-glucuronidase gene under the control of the CaMV 35S promoter in *Schizosacharomyces pombe*. *Mol. Gen. Genet.* **220,** 314-316.

Richert-Pöggler, K.R., and Shepherd, R.J. (1997). Petunia vein-clearing virus: a plant pararetrovirus with the core sequences for an integrase function. *Virology* **236,** 137-146.

Richert-Pöggler, K.R., Noreen, F, Schwarzacher, T., Harper, G. and Hohn, T. (2003). Induction of infectious petunia vein clearing virus from endogenous provirus in petunia. *EMBO J.* **22,** 4836-4845.

Ruiz, M.T., Voinnet, O. and Baulcombe, D.C. (1998). Initiation and maintenance of virus-induced gene silencing. *Plant Cell* **10,** 937-946.

Rüth, J., Hirt, H. and Schweyen, R.J. (1992). The cauliflower mosaic virus 35S promoter is regulated by cAMP in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **235,** 365-372.

Rüth, J., Schweyen, R.J. and Hirt, H. (1994). The plant transcription factor TGA1 stimulates expression of the CaMV 35S promoter in *Saccharomyces cerevisiae*. *Plant Mol. Biol.* **25,** 323-328.

Schoelz, J.E. and Shepherd, R.J. (1988). Host range control of cauliflower mosaic virus. *Virology* **182,** 30-37.

Schoelz, J.E. and Wintermantel, W.M. (1993). Expansion of viral host range through complementaion and recombination in transgenic plants. *Plant Cell* **5,** 1669-1679.

Schoelz, J.E., Wiggens, B.E., Balaji, B., Palukaitis, P.F. and Squires, J. (2005). Integration of Cauliflower mosaic virus into the genome of Arabidopsis thaliana through a mechanism involving homologous recombination. *In* Abstracts of the XIII International Congress of Virology. IUMS 2005. Microbes in a Changing World. San Francisco, CA, USA, July 23-28, 2005. p. 107.

Seemanpillai, M., Dry, I., Randles, J. and Rezaian, A. (2003). Transcriptional silencing of geminiviral promoter-driven transgenes following homologous virus infection. *Mol. Plant-Microbe Interact*. **16,** 429-238.

Shewmaker, C.K., Caton, J.R., Houck, C.M. and Gardner, R.C. (1985). Transcription of cauliflower mosaic virus integrated into plant genomes *Virology* **140,** 281-288.

Staginnus, C. and Richert-Pöggeler, K.R. (2006). Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends Plant Sci.* **11**, 485-491.

Staginnus, C, Gregor, W., Mette, M.F., Teo, C.H., Borroto-Fernández, E.G., Laimer da Câmara Machado, M., Matzke, M. and Schwarzacher, T. (2007). Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. *BMC Plant Biol.* **7:24,** 16 pages.

Sun, L., Cai, H., Xu, W., Hu, Y. and Lin, Z. (2002). CaMV 35S promoter directs beta-glucuronidase expression in *Gonoderma lucidum* and *Pleurotus citrinopileatus*. *Mol. Biotechnol.* **20,** 239-244.

Tanne, E. and Sela, I. (2005). Occurrence of a DNA sequence of a non retro RNA virus in a host genome and its expression: evidence of recombination between viral and host RNAs. *Virology* **332,** 614-622.

Tepfer, M., Gaubert, S., Leroux-Coyau, M., Prince, S., Houdebine, L-M. (2004). Transient expression in mammalian cells of transgenes transcribed from the *Cauliflower mosaic virus* 35S promoter. *Environ. Biosafety Res.* **3,** 91-97.

Vaden, V.R. and Melcher, U. (1990). Recombination sites in cauliflower mosaic virus DNAs: Implications for mechanisms of recombination. *Virology* **177,** 717-726.

Vlasák, J., Šmahel, M., Pavlik, A., Pavingerová, D. and Bříza, J. (2003). Comparison of hCMV intermediate early and CaMV 35S promoters in both plant and human cells. *J. Biotechnol.* **103,** 197-202.

Wintermantel, W.M. and Schoelz, J.E. (1996). Isolation of recombinant viruses between cauliflower mosaic virus and a viral gene in transgenic plants under conditions of moderate selection pressure. *Virology* **223,** 156-164.

Young, M., Shewmaker, C.K., Daubert, S.D. and Shepherd, R.J. (1987). Expression of cauliflower mosaic virus proteins from an integrated form of the viral genome. *J. Gen. Virol.* **68,** 3217-3222.