THE CONSTRUCTION OF AN INFECTIOUS CLONE OF GRAPEVINE VIRUS A (GVA)

by

Jacques du Preez

Presented in partial fulfillment of the requirements for the degree of Master of Science at the Department of Genetics, University of Stellenbosch.

April 2005
Supervisor: Prof JT Burger

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

_____ Date: _____

J. du Preez

Abstract

An infectious clone of a viral RNA genome is one that can be used, either as an *in vitro* transcript or as cDNA, to produce an infection in a susceptible plant. Infectious clones serve as a tool to study viral RNA genomes at a molecular level to gain deeper insight into genome organization, viral gene function, presence of regulatory sequences and gene expression. In the Western Cape (and elsewhere) a new crippling grapevine disease, known as Shiraz disease, is emerging of which the aetiology and pathogenic agents involved are not yet fully understood. *Grapevine virus A* (GVA), genus *Vitivirus*, family *Flexiviridae*, is thought to be the associated with this disease. The aim of this study was to construct a full-length infectious cDNA clone of GVA, which will aid in the molecular study of the viral genome. This clone could ultimately be used to investigate GVA's involvement in Shiraz disease, which could lead to the unravelling of the aetiology and control of the disease. A full-length clone of GVA, named GVA-IC2/T7-2972-3, was constructed in several steps using restriction digestion/ligation and primer overlap extension PCR. Grapevine virus A cDNA fragments were obtained from GVA-infected *Nicotiana benthamiana* and *Vitis vinifera* plants using three different techniques, of which the Rapid direct-one-tube RT-PCR was most successful. A 5' T7 promoter and a 3' poly-A tail were incorporated and the full-length clone was cloned into pBluescript®II SK (+). Full-length sequencing of the clone, revealed two significant frameshift mutations. The first mutation was a single base pair insertion (one G) in a slippery site of 6 G's at position 1380 – 1385 in open reading frame one (ORF 1) of the viral genome. This mutation was corrected by PCR-based site-directed mutagenesis, which resulted in pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4. The second mutation was a single base pair deletion (one G) at position 6959 in ORF4, which coded for the coat protein (CP). Several techniques were attempted to correct this mutation, but none were successful. Even though the second mutation could not be corrected, *in vitro* transcriptions were performed on three clones followed by subsequent infections of *N. benthamiana* plants. The three clones included pSK-GVA-mutagen-3, pSK-GVA-mutagen-4 (both hosting the mutation at position 6959) and GVA-IC2/T7-2972-3 (hosting both mutations). At 21 days post-inoculation no significant visual symptoms were observed in plants infected with *in vitro* RNA or in plants infected with wild type GVA. Rapid direct-one-tube RT-PCR results revealed the presence of viral RNA in infected leaves and apical leaves of infected plants, and provided preliminary evidence that the mutated clones were still capable of systemic infection and viral movement. These results are still inconclusive, and several post-infection studies will have to be performed to confirm these findings. Koch's postulates will also have to be proved in order to confirm

the infectious nature of the clones. The effect of the two mutations in the constructed clones will be investigated further and post-infection analysis performed to deduce whether the viral progeny are devoid of the mutations. Three full-length GVA cDNA clones (hosting mutations) seemingly capable of systemic infection in *N. benthamiana* plants were constructed in this study and have laid the foundation for molecular and mutational analysis of the GVA genome. This could lead to the study of pathogen-host interactions in order to unravel the aetiology of Shiraz disease in the future.

Opsomming

‘n Infektiewe kloon is ‘n cDNA kloon van ‘n virale RNA genoom wat *in vitro* of *in vivo* getranskribeer kan word. Infektiewe klone dien as hulpmiddel om virale RNA genome op molekulêre vlak te bestudeer om sodoende dieper insig te verkry oor genoomorganisasie, virale geenfunksie, teenwoordigheid van regulatoriese volgordes en geenuitdrukking. In die Wes-Kaap (onder andere) is ‘n nuwe kruppelende wingerdsiekte, bekend as Shiraz siekte, besig om pos te vat. Die etiologie van die siekte en patogeniese agense daarby betrokke word nog nie ten volle verstaan nie. Studies toon dat *Grapevine virus A* (GVA), genus *Vitivirus*, familie *Flexiviridae*, met die siekte-toestand geassosieerd is. Die doel van hierdie studie was om ‘n vollengte infektiewe cDNA kloon van GVA te konstrueer om sodoende te dien as instrument in die molekulêre studie van die virusgenoom. Hierdie kloon kan in die toekoms ingespan word om GVA se betrokkenheid by Shiraz siekte te bestudeer, wat weer kan lei tot die ontrafeling van die etiologie en beheer van die siekte. ‘n Vollengte kloon van GVA, genaamd GVA-IC2/T7-2972-3, was in verskeie stappe gekonstrueer deur gebruik te maak van beperkingsvertering/ligasie en inleier oorvleuelings-verlenging PCR. Grapevine virus A cDNA fragmente is vanuit GVA-geïnfekteerde *N. benthamiana* en *Vitis vinifera* plante verkry deur van drie verskillende tegnieke gebruik te maak, waarvan die vinnige direkte enkel-buis RT-PCR die suksesvolste was. ‘n 5’ T7 promotor en ‘n 3’ poli-A stert is geïnkorporeer en die vollengte kloon is in pBluescript®II SK (+) gekloneer. Vollengte volgordebepaling van die kloon het twee beduidende leesraam-verskuiwingsmutasies getoon. Die eerste mutasie was ‘n enkel basispaar invoeging (een G) in ‘n “glyende setel” van ses G’s by posisie 1380-1385 in oop leesraam een (ORF1) van die virusgenoom. Hierdie mutasie is met PCR-gebaseerde setelgerigte mutagenese herstel, wat gelei het tot pSK-GVA-mutagen-3 en pSK-GVA-mutagen-4. Die tweede mutasie was ‘n enkel basispaar delesie (een G) by posisie 6959 in ORF4, wat vir die kapsiedproteïen (CP) kodeer. Verskeie tegnieke is aangepak om dié mutasie te herstel, maar geen was suksesvol nie. Alhoewel die tweede mutasie nie herstel kon word nie, is *in vitro* transkripsies gedoen op die drie klone, gevolg deur die onderskeie infeksies van *N. benthamiana* plante. Die drie klone het pSK-GVA-mutagen-3, pSK-GVA-mutagen-4 (wat die 6959 mutasie bevat) en GVA-IC2/T7-2972-3 (wat beide mutasies bevat), ingesluit. Geen beduidende visuele simptome is teen dag 21, post-innokulasie, waargeneem in plante wat geïnfekteer is met *in vitro* RNA of in plante geïnfekteerd met wilde tipe virus nie. Vinnige direkte enkel-buis RT-PCR resultate het die teenwoordigheid van virale RNA in geïnfekteerde en apikale blare van geïnfekteerde plante getoon, en voorlopige bewyse gelewer dat die gemuteerde klone op een of ander manier in staat is tot sistemiese

infeksie en virale beweging. Hierdie resultate is steeds voorlopig en verskeie post-infeksie studies moet nog gedoen word om hierdie bevindings te bevestig. Koch se postulate sal ook moet bewys word om die infektiewe status van die klone te bevestig. Die effek van die twee mutasies in die gekonstrueerde klone sal verder getoets moet word en post-infeksie analises gedoen word om te ondersoek of die virus-nageslag mutasie-vry is. Drie vollengte GVA cDNA klone (bevattende mutasies) wat skynbaar die vermoë het om *N. benthamiana* plante sistemies te infekteer, is in hierdie studie gekonstrueer. Dit het die grondslag gelê vir molekulêre en mutasie-analise van die GVA genoom. Dit kan uiteindelik lei tot die studie van patogeen-gasheer interaksies om die etiologie van Shiraz siekte te ontrafel.

Abbreviations

A_{600}	Absorpsion value at 600nm
A	Adenine
Amp	Ampicillin
ATP	Adenosine Triphosphate
bp	Base pairs
β-ME	β-Mercaptoethanol
BMV	Brome mosaic virus
BNYVV	Beet necrotic yellow vein virus
BSA	Bovine Serum Albumin
C	Cytosine
CaMV	Cauliflower mosaic virus
cDNA	Complementary deoxyribonucleic acid
CP	Coat protein
CTAB	N-Cetyl-N,N,N-trimethyl Ammonium Bromide
CTV	Citrus tristeza virus
DI RNA	Defective interfering ribonucleic acid
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside triphosphate(s)
dRNA	Defective ribonucleic acid
DTT	Dithiothreitol
ddH_2O	Double distilled water
dsRNA	Double stranded ribonucleic acid
EDTA	Ethylene Diamine Tetra-acetic Acid di-sodium salt
ELISA	Enzyme-Linked Immunosorbent Assay
EtOH	Ethanol
g	Gram(s)
G	Guanine
GDP	Gross domestic product
GES	Glysine-NaOH/EDTA/Sodium
GFP	Green Fluorescent Protein
GGP	Gross geographic product
GLRaV-3	Grapevine leafroll-associated virus 3
GVA	Grapevine virus A
GVB	Grapevine virus B
GVC	Grapevine virus C
GVD	Grapevine virus D
h	Hour(s)
HLV	Heracleum latent virus
IEM	Immuno-electronmicroscopy
IPTG	Isopropylthio-β-D-galactoside
kb	Kilobase
kDA	Kilo Dalton
LB	Luria Berthani broth
MOPS	(3-[N-morpholino]propanesulfonic acid)
MP	Movement protein

µg	Microgram(s)
µl	Microliter
µM	Micromolar
M	Molar
min	Minute
ml	Millilitre(s)
mM	Millimolar
mRNA	Messenger ribonucleic acid
NaOAc	Sodium acetate
nt	Nucleotide
ORF	Open reading frame
PCR	Polymerase Chain Reaction
PEG	polyethylene glycol
PIPES	1,4-Piperazinediethanesulfonic acid
PTGS	Post-transcriptional gene silencing
RE	Restriction endonuclease
RSP	Rupestris stem pitting virus
PVA	Potato virus A
PVP	Polyvinyl pyrrolidone
PVY	Potato virus Y
RdRp	RNA-dependant RNA polymerase
RNA	Ribonucleic Acid
rpm	Revolutions per minute
RT	Reverse transcription
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
Sec	Second(s)
sgRNA	Sub-genomic ribonucleic acid
ssRNA	Single stranded ribonucleic acid
SD	Shiraz disease
SDS	Sodium Dodecyl Sulphate
SSCP	Single-strand conformational polymorphisms
STE	Sodium/Tris/EDTA
T	Thymine
TAE	Tris/acetic acid/EDTA
Tris	Tris(hydroxymethyl)aminomethane
U	Unit(s)
U	Uracil
UV	Ultra Violet
UTR	Un-translated region
V	Volt
v\v	Volume per volume
w\v	Weight per volume
X-Gal	5-bromo-4-chloro-3-indocyl-β-D-galactoside

Acknowledgements

I would like to express my sincerest gratitude to the following :

- My supervisor, Prof JT Burger for his support, guidance and encouragement throughout this study.
- Michael-John Freeborough for his advice, encouragement, support and interest in my work.
- Julia Robson, Hano Maree and Chris van Eeden for their technical assistance and advice.
- Harry Crossley Foundation, National Research Foundation, THRIP, Winetech and the University of Stellenbosch for their financial support.
- The Vitis lab: Anthony, Beverly, Chris, Hano, Helena, Julia, Mandi, Michael-John, Mike and Ndiko for creating an enjoyable working environment.
- My friends and family.

In questions of science the authority of a thousand is not worth the humble reasoning of a single individual.
-Galileo Galilei

Of course, if one ignores contradictory observations, one can claim to have an 'elegant' or 'robust' theory. But it isn't science.
- Halton Arp, 1991, from *Science News*, Jul 27.

Contents

Chapter 1: Introduction

1.1. GENERAL INTRODUCTION

The wine industry plays an important role in the South African economy and contributes 3% to the South African gross domestic product (GDP) (World development report, 2003). According to the Wines of South Africa statistics report in 2003 (http://www.wosa.co.za), there are currently 110 200 hectares of vines producing wine grapes, under cultivation in South Africa. Red varietals account for approximately 45% of the national vineyard, with Cabernet Sauvignon representing 15% and Shiraz 9% of the total. Due to the growth in red wine consumption and shifting market demands, the industry is rapidly increasing its plantings of red wine varietals with Shiraz showing the most dramatic growth. Shiraz was the most planted variety in 1999 and 2000. In 2003 South Africa produced an annual harvest of 965m litres, of which 75% was devoted to the making of premium wine. South Africa is the ninth ranked wine producing country worldwide, producing 3.1% of the world's wine. In the 2003 calendar year, 237.3m litres of natural bottled wines were exported by South Africa, showing an increase of 10% on the previous year. Furthermore, red wine exports grew by 13% to account for 45% of all natural wines exported. In the case of bottled wines, Shiraz, Merlot, Chardonnay and Sauvignon Blanc varietals showed the most export growth in 2003 compared to the previous year. The wine industry currently directly or indirectly employs 348 500 people, including farm labourers, those involved in packaging, retailing and wine tourism. Wine tourism employs some 48 350 of these people. A study conducted by the SA Wine Industry Information & Systems (SAWIS) showed that in 1999, the wine industry contributed 9.7% to the Western Cape's gross geographic product (GGP). This study also concluded that some R3.5 billion of the R14.6 billion contributed by the wine industry to the regional economy was generated indirectly through wine-tourism activities centered in the winelands (http://www.sawis.co.za). These statistics indicate that it is important to protect this resource so that the economy of the Western Cape, and ultimately South Africa, is not negatively affected.

As with all crop plants, grapevines are susceptible to disease causing organisms and pests, which include, bacteria, fungi, insects, nematodes, phytoplasmas and viruses. Of these pests and pathogens, viruses are some of the most devastating pathogens over the long term, causing losses of millions of Rands, although it is difficult to make a formal estimate. Virus control is made difficult by the fact that

there are no cures or treatments, and furthermore no grapevine resistance genes or other natural resistance genes have been discovered against grapevine viruses.

It has been suggested that GVA is currently the second most important virus in South African vineyards, next to *Grapevine leafroll-associated virus* 3 (GLRaV-3), genus *Closterovirus*, family *Closteroviridae*. It has a negative impact on South African vineyards, for it is associated with two destructive diseases namely, grapevine leafroll disease and Shiraz disease (Pietersen, G., Pers. Comm., Department of Plant Pathology, University of Pretoria, South Africa).

The main interest in GVA is its role in Shiraz disease, as Shiraz is one of the most economically important red cultivars in South Africa. Shiraz disease is an emerging disease of grapevine in the Western Cape (and elsewhere) with very distinctive symptoms, and has a crippling effect in vineyards. This deadly disease infects the noble grapevine cultivars Shiraz and Merlot in South Africa. Infected vines deteriorate relatively fast and normally die within five years. The aetiology of the disease is largely unknown, but is believed to be of multiple viral origin with GVA consistently associated with the disease (Goszczynski, D., Pers. Comm., Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa).

In order to elucidate disease and ultimately control viruses in South African vineyards, research needs to be performed on important grapevine diseases and their agents. Grapevine virus A is the most likely agent of Shiraz disease. Grapevine virus A has a single stranded RNA genome, which makes it difficult to study at a molecular level.

Viral genomes are amenable to investigations into their organization and expression, by recombinant DNA technology, because of their small size. Deeper insight can be gained by analyzing and modifying viral genomes at a molecular level. The potential of investigations has greatly been enhanced by the possibility of obtaining infectious clones (as *in vitro*-transcribed RNA copies or as cDNAs) corresponding to the genomes of RNA viruses (Boyer & Haenni, 1994). This will serve as a powerful tool to study RNA viral genomes at the molecular level to gain insight into gene organization and replication of RNA viruses. This will lead to the unravelling of gene expression and pathogen-host interactions to ultimately develop resistance, which will lead to the elucidation of disease.

1.2. PROJECT PROPOSAL

The aim of this study was to create a tool to investigate GVA's role in Shiraz disease and to study pathogen-host interactions on a molecular level. We aimed to construct a full-length infectious cDNA clone of the virus in order to ultimately develop a transient expression vector for grapevine in order to assist in elucidating the aetiology of this disease, and possibly obtaining resistance to Shiraz disease. The full-length infectious clone can shed light on the genome organization and gene expression of GVA.

Specific tasks/steps of this study were:

- Design of GVA-specific primers
- Total RNA extraction or dsRNA extraction from GVA-infected grapevine or tobacco or direct one-step RT-PCR on GVA-infected grapevine or tobacco
- Cloning of GVA PCR fragments into pGEM®-T Easy Vector
- Sequencing of cloned fragments
- Construction of a full-length cDNA clone by joining of overlapping cloned fragments by restriction digestion and ligation or primer extension PCR
- Incorporation of a T7 phage promoter upstream of the 5'-end of GVA
- Incorporation of a poly-A tail downstream of the 3'-end of GVA
- Cloning of the full-length clone into pBluescript®II SK (+)
- *In vitro* RNA transcription followed by infection of tobacco
- Post-infection analysis

Chapter 2: Literature Review

2.1. INTRODUCTION

Shiraz (known as Syrah elsewhere) is one of the most important red grapevine cultivars worldwide. The wine industry in South Africa is drastically increasing its plantings of Shiraz in order to keep up with the growth in red wine consumption (http://www.sawis.co.za). An increase in produce is hampered by the fact that a new disease is emerging in the Western Cape (and elsewhere), known as Shiraz disease. This disease, of which the aetiology is largely unknown, affects the noble grapevine cultivars of Shiraz and Merlot in South Africa. It causes infected grapevines to deteriorate very fast, show distinctive crippling symptoms and eventually die within five years (Goszczynski, D., Pers. Comm., Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa). It is believed that multiple viruses are associated with the disease. A recent study undertaken by Goszczynski *et al.* (2003) revealed that GVA is consistently associated with the disease (Goszczynski, D., Pers. Comm., Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa). Another distinct new emerging disease, known as Syrah decline, is affecting Syrah vineyards in France and California. As with Shiraz disease, very little is known about this mysterious condition and care should be taken not to confuse these two distinct conditions with each other. To further complicate matters, the symptoms of Syrah decline in France and California differ drastically from each other. In France the vines eventually die, whereas in California this condition does not lead to death. For this reason, the condition in California was provisionally called Syrah disorder (Stamp, 2004).

Grapevine virus A, consistently associated with Shiraz disease, is a member of the recently established genus *Vitivirus* (Martelli *et al.,* 1997). It is a phloem-limited virus (Minafra *et al*., 1997), with a linear, positive sense single-stranded (ss) RNA genome (www.dpvweb.net/dpv/showdpv.php?dpvno=383). The complete nucleotide sequence of GVA has been determined (Minafra *et al.,* 1994; 1997) and it was found that the genome is translated by means of five open reading frames (ORF) (www.dpvweb.net/dpv/showdpv.php?dpvno=383). In 1999, the construction of an infectious clone of GVA isolate PA3 was reported by Galiakparov *et al.* Mutation analysis was performed on the clone to experimentally define the role of every GVA gene. Putative translation products were assigned for every ORF except for ORF 2 (Galiakparov *et al.,* 2003). Utilization of the clone to unravel the

regulation of gene expression and the synthesis of viral sgRNAs in GVA, led to the characterization of two nested sets of sub-genomic RNAs (sgRNAs) (Galiakparov *et al.,* 2003). sgRNA-producing viruses are used increasingly for expression of vaccines and other pharmaceutically useful proteins in plants (McCormick *et al.,* 1999). Insight can be gained into the origin and diversification of viruses by recombination and RNA virus replication if the mechanism of sgRNA synthesis is fully understood. The possibility that defective RNAs may arise during vitivirus replication was investigated by Obreque *et al*., (2003). This led to the detection of deleted chimaeric RNAs, with conserved 5' and 3' ends, resembling defective RNAs in GVA-infected *N. benthamiana* plants. Recently, clear-cut post-transcriptional gene silencing (PTGS) suppression was observed in *N. benthamiana* transgenic plants, expressing a GFP (green fluorescent protein) transgene, following GVA infection. Identification of the GVA protein(s) determining PTGS suppression is currently under investigation (Turturo *et al*., 2003). Due to the fact that no function could be assigned to the ORF 2 protein, it could be that this protein plays a role in PTGS suppression, even though it has no sequence homology to any other known proteins.

As mentioned previously, in 1999, an infectious clone of GVA isolate PA3 was constructed by Galiakparov *et al.* An infectious clone is a full-length cDNA clone of a viral RNA genome that can be transcribed, either *in vitro* or *in vivo.* Infectious clones can aid as a tool for the investigation and modification of RNA viral genomes at a molecular level. They can facilitate studies of viruses whose isolations are problematic and viruses that are present only in very low titers. They can also provide useful information in the study of the genetic expression and replication of RNA viruses by the use of deletions, insertions, mutagenesis and complementation studies. Furthermore it can aid in the study of natural or induced RNA recombination, plant-virus movement such as the mechanisms of cell-to-cell movement and pathogen-host interactions. These clones can also be considered as viral gene pools for the design of antiviral strategies, trans-complementation studies and the development of new viral vectors. There are a few limitations and pitfalls when it comes to the construction of infectious clones. Furthermore there are factors that strongly influence infectivity of the infectious clone (Boyer & Haenni, 1994).

The construction of an infectious clone of GVA in South Africa, will supply us with a tool to investigate the genome of GVA. The GVA genome will be represented in complementary DNA (cDNA) form, which will allow us to study the genome organization and replication at a molecular level. Mutation analysis (introducing mutations into all ORFs) will shed light on gene expression of

GVA. RNA, produced *in vitro* from the cDNA clone, could be used to infect *N. benthamiana* plants directly, which could lead to the study of plant-pathogen interactions at a molecular level. This could ultimately lead to the unravelling of the aetiology of Shiraz disease, which could lead to the elucidation of this destructive disease in South Africa.

2.2. GRAPEVINE VIRUS A

2.2.1. Genus *Vitivirus*

The genus *Vitivirus* is one of 8 genera in the family *Flexiviridae.* It contains viruses that were originally tentatively included in the genus *Trichovirus.* The genus name is derived from the host of the type member, *Vitis vinifera* (www.dpvweb.net/notes/showgenus.php?genus=Vitivirus). Grapevine virus A is a member of this recently established genus *Vitivirus* which includes four defined species: GVA, *Grapevine virus B* (GVB), genus *Vitivirus*, family *Flexiviridae*, *Grapevine virus D* (GVD), genus *Vitivirus*, family *Flexiviridae*, and *Heracleum latent virus* (HLV), genus *Vitivirus*, family *Flexiviridae*. A tentative species of this genus is *Grapevine virus C* (GVC), genus *Vitivirus*, family *Flexiviridae* (Martelli *et al*., 1997). The complete genome sequences are known for both GVA and GVB (Saldarelli *et al*., 1996; Minafra *et al*., 1994; 1997) and only partial sequences are available for GVD (Abou-Ghanem *et al*., 1997) and HLV. The genomes of vitiviruses have five ORFs, one of which is downstream of the CP cistron (www.dpvweb.net/notes/showgenus.php?genus=Vitivirus).

2.2.2. Morphology

Grapevine virus A is a phloem-limited virus, which has flexuous, non-enveloped, filamentous particles of 800 nm in length and 12 nm in diameter (Minafra *et al*., 1997) (figure 2.1). The nucleocapsid is transversely cross-banded, obliquely striated, and has rope-like features. The virions contain approximately 5 % nucleic acid (www.dpvweb.net/dpv/showdpv.php?dpvno=383).

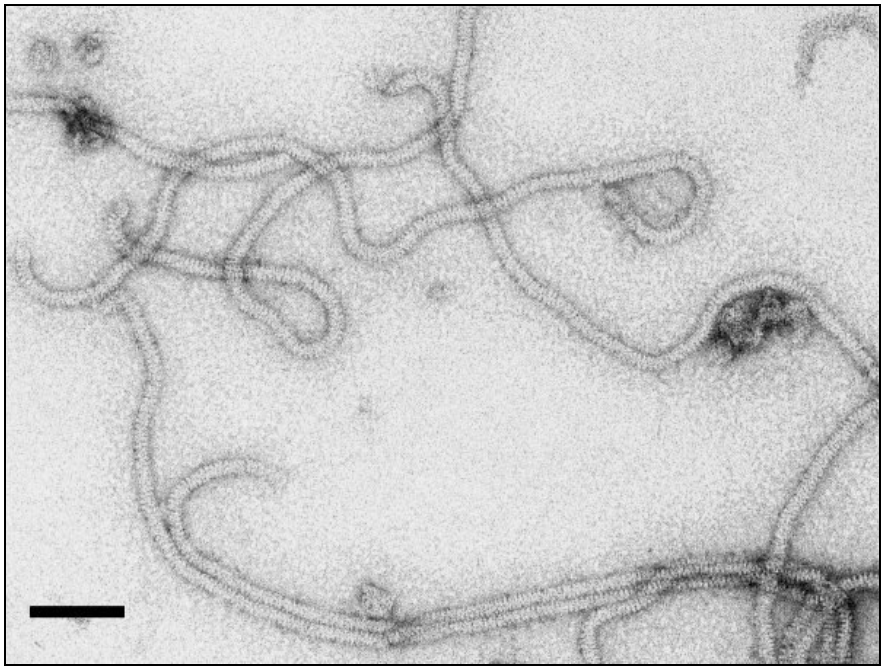


Figure 2.1. Electron micrograph of GVA showing its flexuous, filamentous particles. Bar represents 100nm. (www.dpvweb.net /dpv/showfig.php?dpvno=383&figno=06).

2.2.3. Genome and Genomic organization

Grapevine virus A has a monopartite, linear, positive sense single-stranded (ss) RNA genome of 7349 nucleotides (nt), excluding the 3' poly-A tail. It also possesses a 5' methylated nucleotide cap. The complete nucleotide sequence of a GVA isolate has been determined (Minafra *et al.,* 1994, 1997). The GVA genome is translated by means of five ORFs into a single protein species of Mr c. 22,000 (www.dpvweb.net/dpv/showdpv.php?dpvno=383) and two (recently identified) nested sets of subgenomic RNAs (Galiakparov *et al.,* 2003) (section 2.2.7.2). The roles of some of the ORFs have been deduced based on sequence homology to known genes (Minafra *et al*., 1997). In 1999, the construction of an infectious clone of GVA isolate PA3 was reported by Galiakparov *et al.* (section 2.3.2.1). Mutation analysis was performed on the clone to experimentally define the role of every GVA gene. Putative translation products were assigned for every ORF except for ORF 2 (Galiakparov *et al.,* 2003) (figure 2.2). In 2002, epitope mapping of the GVA capsid protein was performed (Dell'Orco *et al*., 2002). Results suggested that GVA particles carry a highly structured epitope centered on a common peptide region of the CP sequence.

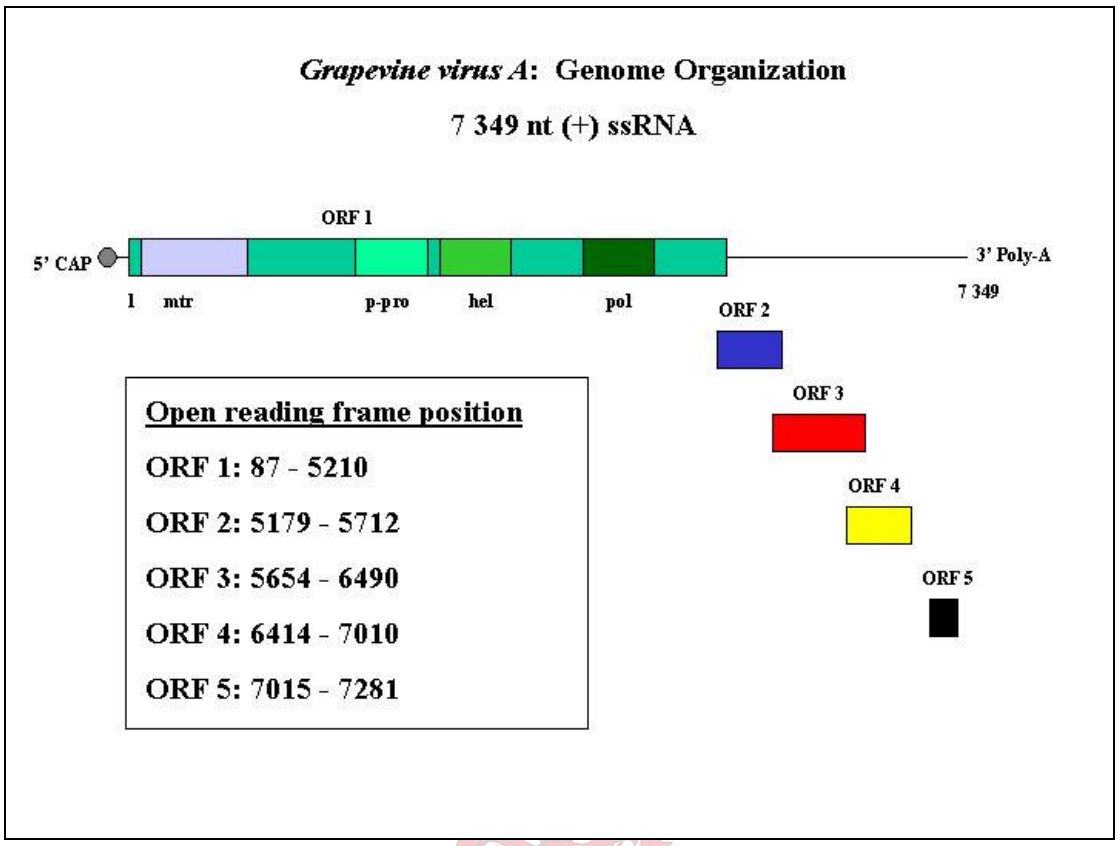


Figure 2.2. Genome organization of GVA depicting the five ORFs and their position in the genome. Grapevine virus A has a monopartite, linear, positive sense single-stranded (ss) RNA genome of 7349 nucleotides (nt), excluding the 3' poly-A tail. It also possesses a 5' methylated nucleotide cap. The genome is translated by means of five ORFs (mtr – methyl transferase, p-pro – potential product, hel – helicase, pol – RNA-dependent RNA polymerase) (Minafra *et al.,* 1994, 1997).

2.2.4. Transmission

As GVA's name implies, it is a pathogen of grapevine from which it can be transmitted with difficulty by sap inoculation to a very narrow range of herbaceous hosts. Grapevine virus A is in fact the first phloem-limited virus to be transmitted by sap inoculation to herbaceous hosts. (Conti *et al*., 1980) (fig. 2.3) It can be transmitted to *N. clevelandii* and *N. benthamiana* plants (Galiakparov *et al.,* 1999) where most of the known virus isolates induce systemic vein clearing in 10-12 days, deformation of the leaves and dwarfing (fig. 2.3) (www.dpvweb.net/dpv/showdpv.php?dpvno=383). It is spread by the propagation of infected material over medium and long distances (www.dpvweb.net/dpv/showdpv.php?dpvno=383) and is naturally transmitted between plants by species of the pseudococcid mealybug genera *Pseudococcus* and *Planococcus.* (Engelbrecht & Kasdorf, 1997; Garau *et al*., 1995; Rosciglione *et al*., 1983)

2.2.5. Diseases and Geographical distribution

Grapevine virus A's involvement in grapevine diseases is still not clear. It was reported more than 20 years ago (Conti *et al*., 1980), and it is one of the most frequently detected viruses in vineyards worldwide

Figure 2.3. Stunting and systemic mottling in GVA- infected *Nicotiana benthamiana*. Healthy plant on the right. (www.dpvweb.net/dpv/showfig.php?dpvno=383&figno=03, DPV383 Figure 03)

(Goszczynski & Jooste, 2003). Grapevine virus A probably occurs wherever *V. vinifera* is grown. It has been detected worldwide from Europe, the Mediterranean basin, Middle East, South Africa, China, Australia, North and Latin America (Boscia *et al*., 1997a). Results obtained in different laboratories worldwide suggest that GVA is involved in the aetiology of Kober stem grooving (Digiaro *et al.,* 1994; Chevalier *et al*., 1995; Garau *et al*., 1995), which is one of the four economically important diseases of the grapevine rugose wood complex (Martelli, 1993). The virus is responsible for crop losses from 5 to 22 % in wine grape cultivars in Italy (Gurau *et al*., 1997), and for decline and death of table grapevines affected by leafroll disease (Digiaro *et al*., 1997). As mentioned previously, GVA is important in South African vineyards for its role in grapevine leafroll disease and Shiraz disease (Pietersen, G., Pers. Comm., Department of Plant Pathology, University of Pretoria, South Africa).

2.2.5.1. Shiraz disease and Syrah decline

Syrah is the French name for the grapevine cultivar, known commonly as Shiraz, in South Africa. Shiraz disease and Syrah decline are the names given to two distinct new emerging diseases. They can be easily confused with one another.

2.2.5.1.1. Shiraz disease (SD)

Shiraz disease is an emerging disease of grapevine in the Western Cape (and elsewhere) with very distinctive symptoms, and a crippling effect in vineyards. This deadly disease infects the noble grapevine cultivars Shiraz and Merlot in South Africa. Infected vines deteriorate relatively fast and normally die within five years. Affected vines are rubbery (do not lignify) and keep their leaves longer than healthy vines. The disease causes delayed budding in Merlot. The aetiology of the disease is largely unknown, but is believed to be of multiple viral origin with GVA consistently associated with the disease (Goszczynski, D., Pers. Comm., Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa). Burger & Spreeth (1993) reported that 50 SD-affected grapevines were infected with GLRaV-3 and/or GLRaV-2 or GVA using ELISA and immuno-electronmicroscopy (IEM). Goszczynski *et al.,* (2003) re-examined their findings by RT-PCR. Results showed that 91 SD-affected grapevines cv. Shiraz and 35 SD-affected grapevines cv. Merlot were all infected with GLRaV-3 and GVA. Further results suggested that only GVA is required for inducing the disease, although detection of both GLRaV-3 and GVA suggested involvement of both viruses in SD. They also suggested that a certain concentration of GVA may be necessary to induce Shiraz disease and furthermore that a specific group of GVA variants may be associated with Shiraz disease. The successful transmission of SD by the mealybug *Planococcus ficus,* strengthens the argument that GVA may be the causative agent. GLRaV-3's involvement in SD could be accidental, as this virus is very common in vineyards in the Western Cape (Goszczynski, D., Pers. Comm., Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa). GLRaV-3 may also be required for the efficient transmission of GVA by mealybugs (Engelbrecht & Kasdorf, 1990a; La Notte *et al.,* 1997).

2.2.5.1.2. Syrah decline

Syrah decline is a mysterious condition that affects Syrah vineyards in California and France about which very little is known. In France, symptoms include the development of swollen and cracked graft unions ("graft union collapse") followed by leaf reddening in mid to late summer, and eventual death over a four-to-ten year period. Syrah decline in France has been of concern since the early 1990s, because Syrah is the fourth most important red variety in France constituting 56 000 ha of vineyard in 2002. No causative agents have been identified as yet, although a range of fungi, playing a secondary role in accelerating the decline, including *Phomopsis, Verticillium* and *Alternaria* have been associated with the graft unions of symptomatic vines (Stamp, 2004).

In California, on the other hand, affected Syrah vines develop red leaves, vines may remain green, graft unions may or may not become cracked and swollen and the disease does not kill the plant. It is still not clear whether Syrah decline in France is related to this similar condition in California, provisionally called **Syrah disorder** (Stamp, 2004).

A possible breakthrough in Syrah decline was made when Dr. Adib Rhowani (Stamp, 2004) determined that specific Syrah decline/disorder symptomatic plants from France and California tested negative for all grapevine viruses except for *Rupestris stem pitting virus* (RSP), genus *Foveavirus*, family unallocated. RSP has not yet been thought to be a threat to vineyard productivity and it is widely dispersed worldwide in grapevine plant material. An interesting fact is that studies have shown that the RNA of the RSP strains isolated from Syrah decline-affected vines in both France and California differ substantially from the previously characterized RSP virus. The two RSP strains in France and California, associated with Syrah decline, also substantially differ from each other, but it is not clear whether this genetic sequence difference is sufficient to cause a condition such as Syrah decline or whether it is in any way related to it. It could, in fact be quite possible that an altogether previously unknown and undetected virus is responsible for Syrah decline, because of the presence of uncharacterized genetic material in Syrah decline-affected vines (Stamp, 2004).

2.2.6. Molecular diversity

Theoretically, all RNA viruses have the potential to establish very large population diversity, because of their error-prone replication, mutation rates in the range of 10^{-3}–10^{-5} misincorporations per

nucleotide copied (Smith & Simmonds, 1997; Drake & Holland, 1999), and their short generation times. However, there is little reliable information on polymerase fidelity and generation time of RNA plant viruses, which make diversity estimates difficult. Virus diversity can also be overestimated according to the strategy used for its assessment, due to errors introduced during analysis (Smith *et al*., 1997). Diversity has been observed between virus genomes within the same individual, with the characterization of a population of closely related variants termed viral quasispecies (Martel *et al*., 1992; Forns *et al*., 1999). Viral quasispecies are closely related (but non-identical) mutant and recombinant viral genomes subjected to continuous genetic variation, competition, and selection. Quasispecies structure and dynamics of replicating RNA enable virus populations to persist in their hosts and cause disease (Domingo *et al*., 1998).

Several techniques are used to study diversity. A quasispecies distribution within a population can be studied using a sequence-based approach, a technique that is time-consuming. To obtain results rapidly, some techniques using differential gel electrophoresis mobility can be applied to amplified product, such as temperature gradient gel electrophoresis (Lu *et al*., 1995), single-strand conformation polymorphism analysis (SSCP) (McKechnie *et al*., 2001), and heteroduplex tracking analysis (Gretch *et al*., 1996). Whatever the strategy adopted to study quasispecies distribution within a population of a RNA virus, nucleotide sequence information is obtained through PCR amplification of virus-specific cDNA produced by reverse transcription of viral RNA. The enzymes used through these two steps, a reverse transcriptase and a thermostable DNA polymerase, exhibit relatively high error rates which should be taken into account when considering the actual heterogeneity within a viral population (Malet *et al*., 2003).

Recent studies indicated a correlation between mutation frequency and virus host range, suggesting that diverse populations constitute an advantage for RNA plant viruses. An insect feeding is likely to transmit a virus to a variety of plant hosts and host adaptability could mean survival for a plant virus in a natural setting (Schneider *et al*., 2000; Schneider *et al*., 2001; Roossinck, 2003). Recently, single-strand conformation polymorphism (SSCP) analysis of GVA isolates in South Africa recovered in *N. benthamiana,* revealed extensive molecular heterogeneity of the virus (Goszczynski & Jooste, 2002). Following this study, Goszczynski and Jooste recovered eight isolates of GVA, which induced different symptoms in leaves of *N. benthamiana,* from various grapevines. Four kinds of symptoms were induced on leaves: (1) mild vein clearing; (2) vein clearing plus interveinal chlorosis; (3) vein clearing, interveinal chlorosis plus strong curling of top leaves; and (4) extensive “patchy” necrosis. They found

that the dsRNA patterns of two isolates, that consistently induced mild vein clearing (referred as mild isolates of GVA) were similar, but different from those of other isolates of GVA. Their analysis based on overall nucleotide (nt) sequence identity in the 3' terminal part of the GVA genome, comprising part of ORF 3, entire ORF 4, entire ORF 5 and part of 3' UTR, revealed that GVA isolates separate into three groups (I, II, III), sharing 91.0-99.8% nt sequence identity within groups and 78.0% - 89.3% nt sequence identity between groups. Mild isolates of the virus were located in group III and shared only 78.0% - 79.6% nt sequence identity with the other isolates. Their comparison of predicted amino acid sequences for MP and CP revealed many amino acid alterations, revealing distinct local net charges of these proteins for mild isolates of the virus. Based on both conserved and divergent nt regions in the CP and ORF 5, they designed oligonucleotide primers for simultaneous RT-PCR detection of all GVA isolates and for the specific detection of the most divergent virus variants represented by mild isolates of the virus (Goszczynski & Jooste, 2003).

2.2.7. Replication mechanism of GVA

Grapevine virus A replicates in the cytoplasm, possibly in association with membranous vesicles (www.dpvweb.net/dpv/showdpv.php?dpvno=383). After the construction of a full-length infectious clone of GVA isolate PA3 had been reported (Galiakparov *et al*., 1999) (section 2.3.2.1), the clone was utilized in an attempt to experimentally define the roles of various GVA genes. Mutations were inserted into every ORF, and the effect on viral replication, gene expression, symptom expression and viral movement were studied in *N. benthamiana* protoplasts (Galiakparov *et al*., 2003). Results showed (figure 2.4) that ORF 1 (nt 87-5210) encodes a 194 kDa polypeptide with conserved motifs similar to replication-related proteins (methyl-transferase, RNA-dependant RNA polymerase and RNA helicase) of the "Sindbis-like" supergroup of positive-strand ssRNA viruses (Minafra *et al*., 1997). A frameshift mutation, introduced at nucleotide position 3632 of this ORF, abolished RNA replication (Galiakparov *et al*., 2003). This truncated polypeptide lacked the RNA-dependant RNA polymerase domain located between amino acids 1234 and 1673 (Galiakparov *et al*., 2003). ORF 2 (nt 5179-5712) encodes a protein of 19 kDa with no significant homology with any other proteins (Minafra *et al*., 1994). None of the aforementioned parameters were affected after inserting a 28 nucleotide deletion, between two *Acc* I sites, within ORF 2 (Galiakparov *et al*., 2003). ORF 3 (nt 5654-6490) encodes a polypeptide of 31 kDa, with amino acid similarity to putative movement proteins (MPs) from the 30 K superfamily (Minafra *et al*., 1994). A 355 nucleotide GUS gene substitution in ORF 3, and two introduced frameshift mutations in ORF 4 (nt 6414-7010), coding for the 22 kDa CP, restricted viral movement.

This implies that ORF 3 codes for the GVA movement protein (Galiakparov *et al*., 2003). A mutation introduced in ORF 5 (nt 7015-7281), by changing the start codon to ATC, partially restricted viral movement and rendered the virus asymptomatic. ORF 5 codes for a small 10 kDa protein (putative RNA binding protein) with no homology to other known proteins (Galiakparov *et al*., 2003). However, the corresponding ORF 5 of the closely related vitivirus GVB shares homology with small, 3'-terminal polypeptides of various plant viruses that contain the zinc finger domain of nucleic acid binding proteins (Minafra *et al*., 1994; 1997).

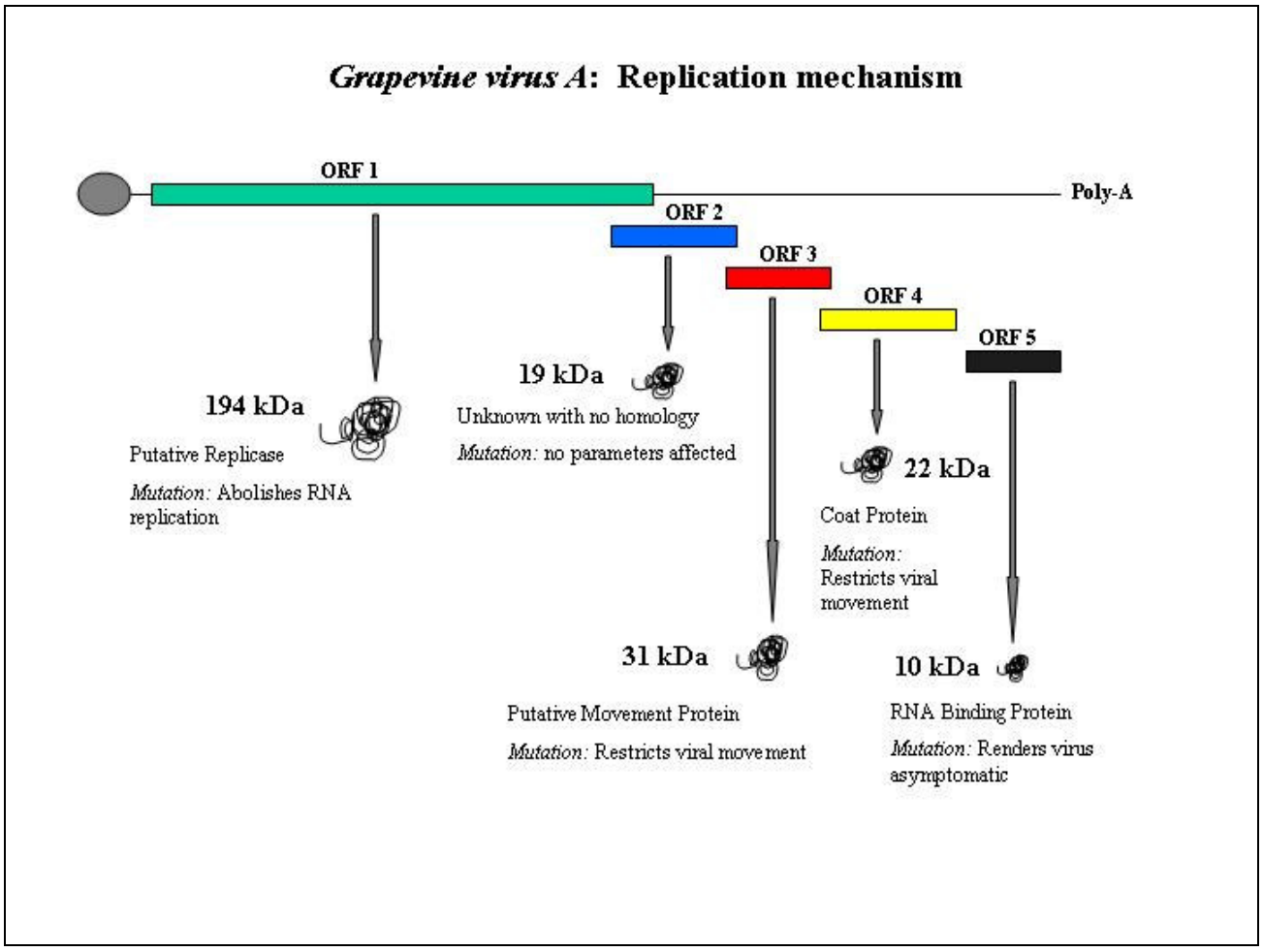


Figure 2.4. Replication mechanism of GVA. The translation products of all 5 ORFs are shown and the effect that an introduced mutation would have on viral replication given. ORF 1 codes for a 194 kDa putative replicase, ORF 2 codes for an unknown protein, ORF 3 codes for a putative movement protein, ORF 4 is the CP and ORF 5 codes for a RNA binding protein (Galiakparov *et al*., 2003).

2.2.7.1. Synthesis of Subgenomic RNAs by Positive-Strand RNA Viruses

Many RNA viruses encode several genes on a single genomic RNA, but only the first 5'-proximal ORF, on a normal eukaryotic mRNA is translated. Downstream genes on viral genomes are thus expressed either via novel translation events or, more commonly, by deployment of subgenomic mRNAs (sgRNAs). Subgenomic RNAs of positive-strand viruses have the same 3' ends as genomic RNA, but have deletions at the 5' ends to bring the 5' end of the RNA in proximity with the start codon of downstream (on genomic RNA) ORFs. The RNA-dependant RNA polymerase (RdRp) is always translated first, directly from genomic RNA of positive-strand RNA viruses, because translation is required for sgRNA synthesis. sgRNAs express products needed during intermediate and late stages of infection, such as structural or movement proteins (Miller *et al.,* 2000). Understanding mechanisms of sgRNA synthesis is important because it will shed light on how these RNA viruses, and GVA in particular, replicate. It may also provide insight into the origin and diversification of viruses by recombination.

2.2.7.2. Subgenomic RNAs in GVA?

A study undertaken by Galiakparov *et al.* (2003) to unravel the regulation of gene expression and the synthesis of viral sgRNAs in GVA led to the characterization of two nested sets of sgRNAs. They explored the production of viral RNAs in a GVA-infected *N. benthamiana* herbaceous host. This led to the characterization of one nested set of three 5'-terminal sgRNAs of 5.1, 5.5, and 6.0 kb, and another of three 3'-terminal sgRNAs of 2.2, 1.8, and 1.0 kb that could serve for expression of ORFs 2-3, respectively. Results obtained, suggested that expression of ORF 5 occurred via bi- or polycistronic mRNA, because neither 3'- nor 5'-terminal sgRNAs, which would correspond to this ORF, was detected. The 5'-terminal sgRNAs were abundant in dsRNA-enriched extracts. Cloning and sequence analysis of the 5' end of the 1.8 kb 3'-terminal sgRNA and the 3'-end of 5.5 kb 5'-terminal sgRNA suggested that a mechanism other than specific cleavage was involved in production of these sgRNAs. Sequences upstream of the 5'-terminus of each of ORFs 2-4 apparently controlled the production of the 5'- and 3'-terminal sgRNAs. Detection of both plus and minus strands of the 5'- and 3'-terminal sgRNAs, though in different levels of accumulation, suggested that each of these *cis*-acting elements is involved in production of four RNAs: a 3'-terminal plus-strand sgRNA which could act as an mRNA, the corresponding 3'-terminal minus-strand RNA, a 5'-terminal plus-strand sgRNA, and the corresponding 5'-terminal minus-strand (Galiakparov *et al*., 2003).

2.2.8. Defective RNAs

Some single-stranded positive sense RNA viruses are known to support the replication of smaller-than-genomic defective RNAs (dRNAs) in infected host cells which, when they interfere with viral RNA accumulation and symptom expression, are referred to as defective interfering RNAs (DI RNAs) (Lewandowski *et al*., 1998; Simon *et al*., 1994; White *et al*., 1999). These RNAs are a mosaic of genome fragments originated by deletion or recombination events. A study was undertaken by Obreque *et al*. (2003) to investigate the possibility that during vitivirus replication, defective RNAs may arise, which contain essentially unmodified 5'- and 3'-ends. Deleted chimaeric RNAs, with conserved 5' and 3' ends, resembling defective RNAs were detected in GVA-infected *N. benthamiana* plants by RT-PCR amplification. A first class of molecules about 790 bp in size contained the 5' UTR and the first 459 nucleotides of ORF 1 fused to a 3' end region (from position 7010 up to the 3' end primer sequence) consisting of the whole ORF 5 and the 3' UTR. Obreque *et al.* sequenced three clones from this class of molecules, which showed a slight variability (3% average) at the nucleotide level in the coding sequences. A second class of molecules, about 420 bp in size, contained the 5' UTR, the first 307 nucleotides of ORF 1 and the 3' end primer, without any residue of ORF 5 and 3' UTR. Two clones from this class of molecules were sequenced, which again showed minor sequence divergence. What is interesting from these results is the fact that ORF 5 is conserved in the larger GVA dRNA, since this gene does not seem to be expressed via a subgenomic RNA. Its presence as dRNA may account for an involvement of the expression product of this ORF, supposed to be involved in suppression of gene silencing (Galiakparov *et al*., 2003). Further investigations are required to investigate the influence of sgRNAs and dRNAs on virus replication and accumulation and *de novo* production in plants.

2.2.9. RNA gene silencing

A number of living organisms use a general mechanism of post-transcriptional regulation of gene expression, known as RNA silencing. In plants this has evolved into a defence mechanism against viruses based on target sequence-specific degradation (Voinnet, 2001). Plant viruses, in response, have developed the ability to counteract host-induced silencing by means of proteins encoded in their genomes (Li *et al*., 2002; Voinnet *et al.*, 1999). Diverse virus species use distinct strategies to target the host gene silencing machinery by expressing structurally and functionally different proteins, which effectively suppress PTGS (Voinnet *et al*., 2000). Recently, clear-cut PTGS suppression was observed

in *N. benthamiana* transgenic plants, expressing a GFP transgene, following GVA infection. Identification of the GVA protein(s) determining PTGS suppression is currently under investigation (Turturo *et al*., 2003).

2.3. INFECTIOUS CLONES

Worldwide viruses cause massive damage to crop quality and yield. There are currently no known resistance genes to grapevine viruses and no cures or treatments against plant viruses exist. By analyzing viral genomes at a molecular level, deeper insight can be gained into their genome organization viral gene function, presence of regulatory sequences (and their function) and gene expression. A positive aspect in this regard is the fact that viral genomes are relatively small, which make them amenable to such investigations by recombinant DNA technology. Molecular analysis of viruses whose replication cycle encompass a DNA intermediate step, are much easier to perform than nonretroviral RNA viruses. Reverse genetics, the ability to produce specific mutations followed by examination of phenotype, revolutionized the study of RNA viruses. The potential of investigations has also greatly been enhanced by the possibility of obtaining infectious clones (as *in vitro*-transcribed RNA copies or as cDNAs) corresponding to the genomes of RNA viruses (Boyer & Haenni, 1994). An infectious clone is a full-length cDNA copy of a viral genome generated by reverse transcription. This full-length clone can be transcribed *in vitro* by use of a bacteriophage RNA polymerase (T7, T3 or SP6) or linked to a *Cauliflower mosaic virus* (CaMV), genus *Caulimovirus*, family *Caulimoviridae*, 35S promoter to produce *in vivo* transcripts (Boyer & Haenni, 1994; Baulcombe *et al.,* 1995; Gal-On *et al.,* 1995; Fakhfakh *et al.,* 1996). Infectious clones have several advantages. They can facilitate studies of viruses whose isolation is problematic or viruses that are present in very low titres. By introducing mutations into the cDNA clone, information can be provided in the study of genetic expression and replication of RNA viruses. Furthermore it can aid in the study of natural or induced RNA recombination, plant-virus movement such as the mechanisms of cell-to-cell movement and pathogen-host interactions. By obtaining cDNA clones of most RNA viruses, viral gene pools will be established for the design of antiviral strategies, trans-complementation studies and the development of new viral vectors (Boyer & Haenni, 1994).

2.3.1. Construction of infectious clones

There are currently two ways of constructing infectious clones: namely (1) The construction of a full-length cDNA clone of a viral genome, from which an *in vitro* infectious transcript can be synthesized under the influence of a bacteriophage (T7, T3 or SP6) RNA polymerase promoter, and (2) The expression of infectious viral RNAs by *in vivo* transcription of cDNA-containing vectors through a CaMV 35S promoter (Boyer & Haenni, 1994; Baulcombe *et al.,* 1995; Gal-On *et al.,* 1995; Fakhfakh *et al.,* 1996).

2.3.1.1. Full-length cDNA clones

There are a few limitations and pitfalls when it comes to the construction of a full-length infectious clone of a viral genome from which an *in vitro* infectious transcript can be produced. It can be a long and tedious process to construct a full-length clone that is infectious. The possibility of producing infectious transcripts from incomplete viral cDNA clones has been reported, even though the presence of the entire viral sequence is generally, thought to be required to obtain infectious clones, but this does not ensure biological activity (Davis *et al*., 1989; Klump *et al*., 1990). In recent studies, it was found that cDNA synthesis, cloning strategies and the design of sequences bordering the viral insert, strongly influences the infectivity of the viral insert (Boyer & Haenni, 1994).

2.3.1.1.1. Basic approach in construction of full-length cDNA clones

In general, the construction of a full-length infectious cDNA clone consists of purification of virus from infected material and purification of the viral RNA. By using a primer hybridizing specifically to the 3' end of the viral genome, the viral RNA is transcribed into single-stranded DNA. The viral RNA is then removed and eliminated. DNA synthesis is initiated with a second primer encompassing the sequence corresponding to the nucleotides at the 5' end of the viral RNA. The single-stranded DNA is thus converted into double-stranded form. Typically, the recognition sequence for a RNA polymerase promoter fused to the viral sequence, are included in the second primer. The synthesis of the full-length first cDNA strand appears to be a serious limiting factor for some viral genomes, because the polymerization step is hampered by strong secondary structures on the viral template RNA (Boyer & Haenni, 1994). It is difficult to sort out an optimum protocol for any virus, because several variations on the general scheme for construction of full-length cDNA clones have been reported (Boyer &

Haenni, 1994). A trend has developed where researchers use thermostable and high fidelity reverse transcriptases and DNA polymerases, both with proofreading capability, to improve the sequence accuracy of the infectious clone. This will ultimately increase the infectivity of the clone and decrease the number of mutations incorporated during reverse transcription and PCR. Interestingly, it has been reported that PCR with *Taq* polymerase, can be successfully applied to obtain infectious clones (Hayes & Buck, 1990), despite the high error rate of the enzyme (Keohavong & Thilly, 1989).

2.3.1.2. Infectious cDNAs

As stated previously, the expression of infectious viral RNAs by *in vivo* transcription of cDNA-containing vectors through a CaMV 35S promoter is the second approach that can be followed when constructing infectious clones or transcripts.

2.3.1.2.1. Advantages

This approach has several advantages. Firstly, the replication process can overcome detrimental effects resulting from RNA degradation, because infectivity of the clone is less dependent on RNA degradation since it presumably occurs only within cells where the RNAs are synthesized. Secondly, an *in vitro* transcription step is not required. This is particularly important for RNA viruses for which the production of a good yield of highly infectious full-length transcripts can be problematic. This is also less expensive, because costly reagents such as the cap analogues and RNA polymerases are not required (Boyer & Haenni, 1994). Lastly, the viral replication process and the expression of viral genes are rendered largely independent of each other, which might be very convenient when studying the role and/or localization of proteins expressed by mutant viral RNAs unable to replicate in cells. *In vivo*-viral transcripts produced in this way would then behave like messenger RNAs produced by a host RNA polymerase, still able to express native or mutant proteins without being replicated (Van Bakoven *et al.,* 1993).

2.3.2. cDNA clones for plant RNA viruses

Infectious transcripts for members of most plant virus groups have been generated from cDNA clones of plant RNA viruses using both described methods (section 2.3.1) (Boyer & Haenni, 1994). Among potyviruses for example, transcripts have been synthesized *in vitro* (T7, T3 or SP6) and proved to be

infectious for among others, *Potato virus A* (PVA), genus *Potyvirus*, family *Potyviridae* (Puurand *et al*., 1996) and *Potato virus Y* (PVY), genus *Potyvirus*, family *Potyviridae* (Jakab *et al*., 1997). *In vivo* (CaMV 35S promoter) infectious transcripts have also been reported for PVY (Fakhfakh *et al*., 1996; Jakab *et al*., 1997) among others.

For vitiviruses, full-length copies of the genomes of GVA (Galiakparov *et al*., 1999) and GVB under the control of bacteriophage T7 RNA polymerase promoter have been synthesized and both transcribed cDNAs were infectious when mechanically inoculated to *N. benthamiana* plants (Saldarelli *et al*., 2000). Capped *in vitro*-transcribed RNA was infectious in *N. benthamiana* and *N. clevelandii* plants (Galiakparov *et al*., 1999) (section 2.3.2.1).

2.3.2.1 Infectious RNA transcripts from GVA cDNA clone

In 1999, Galiakparov *et al.* constructed an infectious clone of GVA isolate PA3 (provided by A. Minafra and G.P. Martelli). They partially purified GVA by polyethylene glycol precipitation and extracted RNA from virions. They extracted dsRNA by CFII chromatography followed by CC41 chromotography. Reverse transcription was performed on the dsRNA followed by PCR. Fragments were cloned into vectors and the full-length clone was constructed with restriction digestion. A T7 promoter was incorporated immediately adjacent to the 5'-end of the GVA genome and a poly-A tail adjacent to the 3'-end. After *in vitro* transcription, resulting RNA was used to mechanically inoculate *N. benthamiana* and *N. clevelandii* plants. After 10-14 days, the symptoms induced by the RNA transcripts were indistinguishable from the parental virus. Observation of the virions by electron microscopy and serological detection of the virus coat and movement proteins were used to confirm the infectivity of the *in vitro*-produced transcripts. This was the first report of infectious RNA transcripts derived from a full-length clone of a member of the *Vitivirus* genus (Galiakparov *et al*., 1999).

2.3.3. Factors influencing infectivity of the cDNA clone

There are a few factors that influence the infectivity of an infectious clone: namely (1) the heterogeneity of transcript population, (2) the presence of point mutations, (3) presence of nonviral nucleotides, (4) Instability in bacteria, and (5) RNA polymerase used.

2.3.3.1. The heterogeneity of transcript population

Heterogeneity of transcript size was reported to be a problem when using the *E. coli* RNA polymerase-λPm promoter system (Ahlquist *et al*., 1984; Dawson *et al*., 1986; Janda *et al*., 1987; Hamilton & Baulcombe *et al*., 1989). Competition between incomplete non-replicatable viral copies and full-length transcripts for interaction with viral and/or host factors involved in the replication process, could account for the relatively low infectivity of most preparations. This problem can be circumvented by using a bacteriophage promoter (Boyer & Haenni, 1994). The relative low fidelity of most reverse transcriptases and polymerases, increase the amount of misincorporations into the genome and contribute to heterogeneity.

2.3.3.2. The presence of point mutations

When working with long viral genomes, like GVA, point mutations are to be expected. This is mainly due to the relatively poor fidelity of the RNA- and DNA-synthesizing enzymes. Furthermore, the low fidelity of viral dependent RNA polymerases can lead to the faithful reverse transcription and amplification of an initial virion RNA, which is mutated (lethal mutant) and would most probably be eliminated in the next round of viral replication. The resulting full-length cDNA clone, or the *in vitro* transcript, would not be infectious. It has been reported that in some cases infectivity can be restored by exchanging a specific region of the cDNA with a fragment corresponding to the same region from an independent cDNA clone (Ahlquist *et al.,* 1984). Conversely, possible mutations favouring infectivity of the full-length transcripts may occur, as it has been reported that certain synthetic transcripts induced more severe symptoms than the parental virion RNAs (Hamilton & Baulcombe, 1989, Hayes & Buck, 1990). As mentioned previously, it has been reported that PCR with *Taq* polymerase, can be successfully applied to obtain infectious clones (Hayes & Buck, 1990), despite the high error rate of the enzyme (Keohavong & Thilly, 1989).

2.3.3.3. Presence of nonviral nucleotides

Many authors have studied the effect of nonviral nucleotides at the extremities of viral transcripts. It is generally admitted that extensions at the 5'-end of viral transcripts, strongly reduces infectivity, whereas 3'-extensions are more easily tolerated (Boyer & Haenni, 1994).

2.3.3.3.1. Effect of 3'-extensions

Different studies performed on the effect of 3'-extensions, revealed the following facts (Boyer & Haenni, 1994). Short 3'-extensions of 1-7 nucleotides don't seem to influence the biological activity of viral transcripts (Dawson *et al.,* 1986), whereas long extensions between 82 (Dzianott & Bujarski, 1989) and 2700 nt (Ahlquist *et al.,* 1984) for different viruses abolishes infectivity. For many viruses, transcripts bearing a long (>30 nt) additional sequence at the 3'-end are infectious (Dzianott & Bujarski, 1989), as are *in vivo* transcripts presumably polyadenylated by host cell enzymes. It is not clear whether the structure of additional 3'-sequences affects the biological activity of the transcripts as Sarnow (1989) showed that 4 extra bases (after a short poly-A tail) at the 3'-end of synthetic poliovirus transcripts did not lower infectivity, but 17 cytosine residues decreased infectivity. The same study revealed that the presence of long homopolymeric adenine sequences (poly-A tail) at the 3'-end of *in vitro*-produced poliovirus transcripts increases infectivity, whereas an adverse effect is seen for long heteropolymeric nucleotide sequences (Sarnow *et al.,* 1989). Dzianott and Bujarski (1989) reported that infectivity of *Brome mosaic virus* (BMV), genus *Bromovirus*, family *Bromoviridae*, transcripts presenting 19 extraviral nucleotides was higher than those harboring 6 or 7 nonviral nucleotides at the 3'-end. This could be due to the fact that longer extensions could confer protection *in vivo* against ribonucleases, or due to the fact that it could alternatively modify the 3'-end. This modification could cause the 3'-end to acquire an enhanced affinity for the BMV replicase (Dzianott & Bujarski, 1989).

2.3.3.3.2. Effect of 5'-extensions

Different studies performed on the effect of 5'-extensions, revealed the following facts (Boyer & Haenni, 1994). In most cases, infectivity is greatly decreased for 5'-extensions, even for only 1 or 2 nucleotide extensions. When transcripts derived from plant viruses harbor moderately long 5' additional sequences of 14 to 17 nucleotides (Heaton *et al.,* 1989), infectivity is completely abolished. Commandeur *et al.* (1991) made an interesting observation when cloning cDNA sequences of *Beet necrotic yellow vein virus* (BNYVV), genus *Benyvirus*, family unallocated, downstream of the CaMV 35S promoter. The resulting *in vivo* transcripts, containing up to 40 extra-viral nucleotides at the 5'-end, were infectious *in planta. In vitro*-derived transcripts containing the same extensions, were biologically inactive in the host plant. Another interesting fact is that the infectivity of animal viruses, compared to plant viruses, is less affected when transcripts contain relatively large 5'-extensions. It is assumed that the problem with extra 5'-non-viral sequences is that its presence could seriously hamper

proper initiation of (+) RNA synthesis from the 3'-end of the (-) strand. These nucleotides do not possess initiation codons, so it seems unlikely that they interfere with *in vivo* viral gene translation, and the *in vitro* translation products are similar to those of wild-type RNA (Janda *et al*., 1987; Verver *et al*., 1987; Angenent *et al*., 1989; Eggen *et al*., 1989; Dore *et al*., 1990; Gal-On *et al*., 1991; Viry *et al*., 1993).

2.3.3.3.3. Effect of the cap structure

Studies performed on the effect of the cap structure (for viruses without a viral genome linked protein, VPg), revealed the following facts (Boyer & Haenni, 1994). For optimum infectivity, as a general rule, a cap structure (m7GpppG) is required at the 5'-end, possibly because it enhances translation initiation (Shih *et al.,* 1976) and/or improves their stability by conferring a better resistance to host cell nucleases (Furuichi *et al.,* 1977). In a few cases both capped and uncapped transcripts proved to be infectious, although uncapped transcripts were almost always either not infectious or had a highly reduced level of infectivity (Angenent *et al.,* 1989).

2.3.3.4. Instability in bacteria

Another limiting factor is the high instability of full-length cDNA clones in bacteria. This problem is especially manifested in (+) RNA viruses (of which GVA is one). As a consequence, infectious clones for some viruses are either impossible to assemble, or once assembled, are very difficult to maintain in *E. coli* due to their predisposition to spontaneous rearrangements and/or to acquisition of stabilizing mutations (Yamshchikov *et al.,* 2001). One plausible explanation is that the instability results from unanticipated expression of viral cDNA resulting in products that are toxic to the bacterial host. While stable infectious clones for some viruses have been reported (Boyer & Haenni, 1994; Khromykh & Westaway, 1994; Meyers *et al*., 1996; Geigenmuller *et al*., 1997), similar constructs for other viruses exhibit acceptable stability only after assembly in low copy number vectors (Gritsun & Gould, 1998; Gualano *et al*., 1998; Mendez *et al.,* 1998; Hurrelbrink *et al*., 1999; Almazan *et al*., 2000). In order to circumvent the instability problem for certain viruses, more sophisticated procedures have been used for assembly of infectious clone cDNA templates. Methods used include, *in vitro* ligation (Sumiyoshi *et al*., 1992; Kapoor *et al*., 1995), long high-fidelity PCR (Campbell & Pletnev, 2000), or a combination of both (Herold *et al*., 1998), the insertion of short introns into problematic expression regions (Yamshchikov *et al.,* 2001), and the introduction of frameshift mutations in infectious cDNA

clones (Satyanarayana *et al.,* 2003). The use of nonbacterial cloning systems (Polo *et al*., 1997) has been reported as well.

2.3.3.5. RNA polymerases

The choice of the RNA polymerase promoter is important when dealing with the design of a full-length clone from which infectious RNAs are expected to be produced, *in vitro* or *in vivo*, because it directly affects the yield of transcripts and the nucleotide sequence at the extremities. Allthough several types of promoters have been used such as the *E. coli* Pm promoter, from bacteriophage λ (Ahlquist & Janda 1984), and the promoters of bacteriophages SP6 (Melton *et al*., 1984), T3 and T7 (Dunn & Studier, 1983). The bacteriophage T7 promoter is more commonly used, because of the more thoroughly studied genetics of bacteriophage T7. Compared to phage RNA polymerase-based systems, the the *E. coli* RNA polymerase-based system produces much lower transcript yields, because it leads to a large proportion of premature termination products (Melton *et al*., 1984; Janda *et al*., 1987; Angenent *et al*., 1989; Heaton *et al*., 1989).

2.3.4. Conclusion

Even though GVA was reported more than 24 years ago (Conti *et al.,* 1980), the surface of GVA's involvement in disease hasn't even been scratched. Grapevine virus A is one of the most frequently detected viruses worldwide (Goszczynski & Jooste, 2003) and cause dramatic losses to vineyards worldwide. In South Africa (and elsewhere), a new disease known as Shiraz disease is threatening to become a problem. Since GVA is thought to be associated with the disease (Goszczynski & Jooste, 2003), the construction of an infectious clone of GVA could greatly benefit in unravelling the aetiology of the disease.

Infectious cDNA clones of RNA viral genomes have been constructed for many viruses. They serve as tools to study RNA viral genomes at a molecular level. There are many pitfalls when it comes to the construction of such clones. When these pitfalls are overcome, infectious clones can lead to the unraveling of genome organization, gene expression and pathogen-host interactions of RNA viruses. This could ultimately lead to development of resistance to, and proper control over, disease.

Chapter 3: Materials and Methods

3.1. PLANT MATERIAL

3.1.1. Plant material

Genomic fragments of GVA were obtained from two different plant sources, namely GVA infected grapevine, I 3973 (obtained from Mr. Nolan Africander, South African Agricultural, Food, Quarantine & Inspection Services, Stellenbosch, South Africa), and GVA (GTR1-2) infected *N. benthamiana* (obtained from Dr Dariusz Goszczynski, Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa).

3.1.2. Plant cultivation

N. benthamiana seeds were germinated on damp filter paper in Petri dishes in the dark. Plants were grown in heat sterilized soil in a growth room with controlled conditions of, temperature between 18ºC and 24ºC, relative humidity of approximately 70% and a 16 hour to 8 hour light/dark cycle (Freeborough, 2003). Plants were fed with SEAGRO® (Premier Fishing SA (Pty) Ltd) every two weeks according to the manufacturer's instructions.

3.2. GENERATION OF GVA cDNA FRAGMENTS

GVA cDNA fragments were generated with three different methods: namely (1) double stranded RNA extraction followed by 2-step RT-PCR, (2) total RNA extraction followed by 2-step RT-PCR, and (3) rapid direct-one-tube RT-PCR.

3.2.1. Double stranded RNA (dsRNA) extraction: CFII cellulose method

Double stranded RNA extraction was performed according to the method of Rezaian and Krake (1987). Grapevine petioles and bark scrapings were prepared for dsRNA extraction and stored at -80°C.

Ten grams of phloem tissue was ground to a fine powder in liquid nitrogen, using a sterile mortar and pestle. Ground material was resuspended in 112 ml extraction buffer (45 ml 2x STE [1x STE (100 mM NaCl, 50 mM Tris-HCl pH 7, 1 mM EDTA)], 1.3% SDS, 32 mg Bentonite, 15 mM β-ME, 25 ml

Phenol, 25 ml Chloroform) and stirred (45min, 4°C). The mixture was centrifuged (8 000 rpm, 15 min, 4°C). The supernatant was retained and 1x STE was added to a volume of 83 ml. Seventeen ml of EtOH (100% (v/v)) and 2 g of sterile CF11 cellulose powder (Whatman) were added. The mixture was stirred (1 h, room temperature), loaded onto a chromatography column (Promega) and allowed to run through. The column was washed with 100 ml STE/EtOH (100 mM NaCl, 50 mM Tris-HCl pH 7, 1 mM EDTA, 17% ethanol v\v). The dsRNA was eluted with 9 ml 1x STE and precipitated by adding 2.5 volumes of 100% ethanol and 0.9 ml 3M NaOAc, pH 5.2 (-20°C, overnight). The RNA solution was centrifuged (7 300 rpm, 30 min) and resuspended in 500 µl 1x STE. The RNA was precipitated with 50 µl 3 M NaOAc (pH 5.2) and 2 volumes 100% EtOH and incubated (-80°C, overnight). The sample was centrifuged (13 200 rpm, 30 min, 4°C) and the resulting pellet was washed with 1 ml 70% EtOH, dried in a Speedy Vac and resuspended in 50 µl double distilled H_2O (ddH_2O). Ten microlitre aliquots were stored at -80°C.

3.2.2. Total RNA extraction

Five hundred micrograms of leaf veins/bark strips were excised, using a scalpel blade, and placed into a cold mortar. A pinch of carborandum (Saarchem, 180 grid) was added and ground to a paste using a pestle. One millilitre RNA extraction buffer (0.5 M Tris-Cl, pH 8.0, 1.4 M NaCl, 3% CTAB, 20 mM EDTA, 0.5% β-Mercaptoethanol) was added and ground further. Another 4 ml of extraction buffer was added, ground further and 1 ml of the slurry was pipetted into each of two microfuge tubes. The tubes were incubated in a water bath (15 min, 60°C) followed by centrifugation (12 000 rpm, 5 min, 4°C). Seven hundred and fifty µl of the supernatant was removed, and transferred to a clean tube on ice and extracted with an equal volume of chloroform:isoamylalcohol (24:1). The tubes were inverted vigorously to mix and centrifuged (12 000 rpm, 15 min, 4°C). Six hundred µl of supernatant was removed, 1/3 volume of 8 M LiCl was added and incubated on ice overnight. This was followed by a centrifugation (12 000 rpm, 30 min, 4°C). The supernatant was discarded, 500 µl 70% ethanol added to the pellet and centrifuged (12 000 rpm, 10 min, 4°C). The ethanol was discarded and the pellet was briefly dried in a Speedy Vac. The pellet was resuspended in 25 µl ddH2O and stored at –80°C until required.

3.2.3. Primer design

RT-PCR and PCR primers were designed from a known GVA sequence found in the GENBANK database (NCBI) (accession nr NC_003604) using PRIMER DESIGNER (Version 1.01. Serial number 50132, Copyright 1990, scientific and educational software). Primers were synthesized by Inqaba

Biotech (Pretoria, South Africa) and Integrated DNA Technologies, Inc (Coralville). Primers were designed to satisfy the following criteria: GC content of 45 – 55% and annealing temperature of 50 - 60ºC. All primer melting temperatures were calculated, based on nearest neighbour parameters, using the Oligonucleotide Analyzer tool on the website http://www.rnature.com.

Table 3.1. Characteristics of primers used in this study.

Primer name	Sequence	Annealing Temperature (°C)
GVA-ext-For	GAA TAT TTA ACT TGA TTC CCA TCG	51.5
GVA-ext-Rev	GTC TTC GTG TGA CAA CCT AG	53.6
GVA-int-For	GAG CAA TGA AGG AAG AGA AG	51.6
GVA-int-Rev	CGG CTC CTC TAT CTC TAT GT	53.5
GVA-2972-Rev	CCA ATC ATC TCT GAG CAC TT	53.3
GVA-1823-For	CGA GAG TGA TGT GGA GAA G	52.7
GVA-5252-Rev	CCA CTA CCT GAA TCA CAG C	53.4
GVA-4110-For	TTC AAG GAC ACA TAC CAG C	53.1
GVA-2793-For	CTC GGC AAG TAC AAT AAA GG	51.9
GVA-3367-Rev	CTT CTC TTC CTT CAT TGC TC	51.6
GVA-3794-For	ACA TAG AGA TAG AGG AGC CG	53.5
GVA-4370-Rev	GAA TCT ACA CAG CAC CGA AT	53.5
GVA-2193-Rev	GTA GCT CCT TCT TCT CGA TC	60.4
GVA-4821-For	CTC TCG ACC AGT TGA CAC T	60.2
GVA-5655-For	ATA GCT GTA ATC ACC TTC CG	58.4
GVA-2007-For	GGC CTA GAT GAT AGC TAC GA	60.4
GVA-1434-For	GTG AAC TTC CTC GAC ATC AG	60.4
GVA-5061-Rev	GCT GTG CAT CAA CAT CAA TA	56.3
GVA-6848-For	ACA GGT TAT GTT CGA CTT CG	58.4
GVA-941-For	GTA CTT GTA CGA TTC CTC CG	60.4
GVA-6979-Rev	GAC TGA GCG TTG AAT ACT CC	60.4
GVA-6141-Rev	GTG TAT CCG TTC AGC AGA TC	60.4
GVA-620-Rev	TTG CAT CTC GTC TAG ACT CC	60.4
GVA-5233-For	AGC TGT GAT TCA GGT AGT GG	60.4
GVA-5880-For	ATC TGG ACA TCA TGG ATG AG	58.4
GVA-1615-Rev	CTA GAG GAG CAC ACA TAC CG	62.5
GVA-508-For	CAT TCA GAG AAC TCG ACG TT	58.4
GVA-1027-For	AGC CAA CAC TTC TGT ACC TC	60.4
GVA-1188-Rev	CCA TCT TGT CAT ACA GAG CC	60.4
GVA-264-For	GGC TAT GGT ATG GAG TTC AG	60.4
GVA-5467-Rev	GAC TTC TTG GAC TGT TGC AC	60.4
GVA-922-For	CGG AGG AAT CGT ACA AGT AC	60.4
GVA-1598-Rev	CCC TGG AGG AGC ATA CAT AC	62.5
GVA-1434-Rev	CTG ATG TCG AGG AAG TTC AC	60.4
GVA-1188-For	GGC TCT GTA TGA CAA GAT GG	60.4
GVA-6979-FN	GGA GTA TTC AAC GCT CAG TC	60.4
GVA-5467-FN	GTG TAA CAG TCC GAG AAG CC	62.5

GVA-2972-FN	GGA AAG TGC GAG ATG ATT GG	62.8
GVA-1188-FN	GGC TCT GTA TGA CAA GAT GG	60.4
GVA-5061-FN	GAA TAT TGA TGT TGA TGC ACA GC	59.2
GVA-3794-FN	GGA CAT AGA GAT AGA GGA GCC	62.6
GVA-2193-FN	GGG AAG ATC GAG AAG AAG GAG	62.6
GVA-ER-Poly A	GTC TTC GTG TGA CAA CCT AG $(T)_{40}$	63.4
Geiser-For	GCG GCC GCG GGA ATT CGA TTT AAT ACG ACT CAC TAT AGA ATA TTT AAC TTG ATT CCC	73.4
T7	TAA TAC GAC TCA CTA TAG GG	55.0
SP6	TAC GAT TTA GGT GAC ACT ATA G	55.0
GVA-5'-mutagen-F	GAA GGG GGG CAA CGT TCA AAG	64.5
GVA-5'-mutagen-R	CTT TGA ACG TTG CCC CCC TTC	64.5
GVA-3'-mutagen-F	GGT GCA AAG GGA GTA TTC AAC	60.6
GVA-3'-mutagen-R	GTT GAA TAC TCC CTT TGC ACC	60.6

3.2.4. Reverse transcription polymerase chain reaction (RT-PCR)

3.2.4.1. Primer annealing

dsRNA

The reverse transcription (RT) reaction was performed using 2.5 µl of dsRNA, 0.25 µl Reverse primer (10 µM) and 3.625 µl H_2O as the primer annealing cocktail. This cocktail was incubated at 95°C for 15 min followed by flash cooling on ice for 5 min.

Single stranded RNA (ssRNA)

The primer annealing cocktail (4 µl of extracted RNA, 1 µM reverse primer per 11.75 µl reaction) was mixed in a 0.5 ml tube. The tube was incubated (70°C water bath, 15 min) to denature RNA secondary structure, followed by a quick chill in ice-water slurry to anneal primer.

3.2.4.2. cDNA synthesis

Reverse transcription was performed with Superscript™II, Superscript™III Reverse transcriptase (Invitrogen) or AMV Reverse transcriptase (Promega) according to the respective manufacturer's instructions.

dsRNA

The RT Cocktail (1x First Strand Buffer (50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$), 0.5 mM dNTPs, 10 mM dithiothreitol (DTT), 2U RNasin, 1 U Reverse transcriptase) was prepared and added to the primer annealing cocktail and incubated (42°C, 1 h 30 min).

ssRNA

The RT Cocktail (1x First Strand Buffer [50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$), 0.5 mM dNTPs, 10 mM dithiothreitol DTT, 2 U RNasin, 2.5 U reverse transcriptase per 8.25 µl reaction) was mixed in a separate 0.5 ml tube, added to template, mixed gently and incubated (42°C, 2 hours).

3.2.4.3. PCR amplification

All PCR reactions were performed in a GeneAmp® PCR System 9700 (Applied Biosystems).

PCR conditions for cDNA derived from dsRNA: 1 µl cDNA template, 1x Bioline PCR NH_4 reaction buffer (16mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20), 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM forward primer, 1 µM reverse primer, 0.5 U Bioline BIOTAQ™ DNA polymerase per 10 µl reaction.

PCR conditions for cDNA derived from ssRNA: 2 µl cDNA template, 1x Bioline PCR NH_4 reaction buffer, 1.5 mM $MgCl_2$, 0.15 mM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 0.5 U Bioline BIOTAQ™ DNA polymerase per 10 µl reaction.

PCR conditions for a universal reaction: approximately 20 ng of template DNA, 1x Bioline PCR NH_4 reaction buffer [16mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20], 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM forward primer, 1 µM reverse primer, 1x cresol/sucrose loading dye (20% w/v sucrose, 1mM cresol red), 0.5 U Bioline BIOTAQ™ DNA polymerase per 10 µl reaction.

PCR cycling conditions were as follows: 1 cycle of 94°C for 5 min; 35 cycles of 94°C for 30 sec, 54°C for 30 sec, 72°C for 45 sec and 1 cycle of 72°C for 7 min. (The primer annealing temperature varied depending on the primer pair used.) (table 3.1)

3.2.5. Rapid direct-one-tube RT-PCR

Rapid direct-one-tube RT-RCR was performed according to a modification of the method of La Notte *et al.,* (1997).

One hundred and fifty micrograms of petiole tissue was ground in 3 ml Grape-ELISA grinding buffer (1/20) (grinding buffer pH 9.6, Na_2CO_3, $NaHCO_3$, 2% PVP 40, 0.2% BSA, 0.05% Tween 20, 1% Sodium metabisulfide) and stored at –80 °C until required. Four µl of the 1/20 grape extract was pipetted into a tube containing 25 µl sterile 1x GES (GES, 0.1M glycine-NaOH pH 8.0, 50mM NaCl, 1mM EDTA, 0.5% Triton X-100) and denatured in a waterbath at 95°C for 10 min, followed by a 5 minute rapid cooling on ice. Two µl was pipetted into 23 µl One-Tube RT-PCR mix. (One-Tube RT-PCR mix final reagent concentrations: 1x NH_4 Buffer, 1.5 mM $MgCl_2$, 1x cresol/sucrose, 0.2 mM dNTPs, 0.5 µM forward primer, 0.5 µM reverse primer, 5 mM DTT, 1 U Superscript™II or Superscript™III Reverse transcriptase (Invitrogen), 0.5 U BIOLINE BIOTAQ™ DNA Polymerase/25 µl reaction) PCR cycling conditions were: 1 cycle of 48 °C for 45 min, 30 cycles of 94°C for 30 sec, 52°C for 45 sec, 72°C for 1 minute and 1 cycle of 72°C for 7 min.

3.3. AGAROSE GEL ELECTROPHORESIS

Agarose gel electrophoresis was performed as per Sambrook *et al.* (1989).

Separation of DNA fragments were performed on 1% (w/v) agarose D1 LE (Hispanagar) gels in 1x TAE (0.04 M Tris-acetate, 0.001 M EDTA pH 8.0) at 100 to 120 mV for 45 to 60 min. A 50x TAE stock solution (2 M Tris, 5.71% (w\v) glacial acetic acid, 0.05 M EDTA pH 8.0) was prepared and used in a 1x working solution. Ethidium bromide (0.5 µg/ml) was added to the agarose gel for visualization by ultra-violet (UV) transilluminator. One time (1x) loading buffer (20% (w/v) sucrose, 1 mM cresol red) was used to facilitate loading of DNA into wells. Molecular size markers used included the Hyperladder I (BIOLINE), 1Kb PLUS DNA ladder (GIBCO BRL, Life technologies), λ *Sty* I ladder and λ *Pst* I ladder. λ Sty I and λ *Pst* I ladders were prepared by digesting 30 µg of λ DNA with 100 U of either *Pst* I or *Sty* I in a total volume of 500 µl.

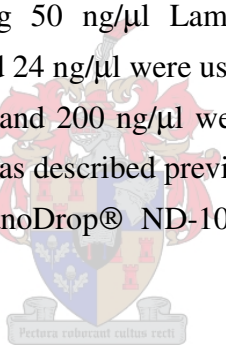
3.4. PCR PRODUCT GEL PURIFICATION

The GFX™ PCR DNA and Gel Band Purification Kit (Amersham Biosciences) was used according to the manufacturer's instructions.

All PCR fragments were excised from a TAE/agarose gel, using a scalpel, and weighed in a microfuge tube. For every 10 mg of agarose gel slice, 10 µl of capture buffer (buffered solution containing acetate and chaotrope) was added and vortexed. The tubes were incubated in a waterbath (60°C, 10 – 15 min) until the gel slice was completely dissolved. The sample was transferred to a GFX column placed within a collection tube, incubated (room temperature, 1 min) and centrifuged (13 200 rpm, 30 sec). The flow through was discarded and 500 µl of wash buffer (10mM Tris-HCl pH 8.0, 1mM EDTA, 80% EtOH absolute) was added to the column. The column was centrifuged (13 200 rpm, 30 sec), the flow through was discarded and the column recentrifuged (13 200 rpm, 30 sec). The column was transferred to a clean 1.5 ml microfuge tube; 20 µl of ddH_2O was added to the column and the sample incubated (room temperature, 1 min). The column was centrifuged (13 200 rpm, 1 min) and the eluted DNA was stored at 4°C till required.

3.5. DNA QUANTIFICATION

DNA quantification was performed using 50 ng/µl Lambda (λ) DNA standards (Promega). A concentration series of 8 ng/µl, 16 ng/µl and 24 ng/µl were used for insert and vector quantification and a series of 50 ng/µl, 100 ng/µl, 150 ng/µl and 200 ng/µl were used for plasmid DNA quantification. Agarose gel electrophoresis was performed as described previously in section 3.3. At later stages in this study, DNA was quantified with the NanoDrop® ND-1000 Spectrophotometer according to the manufacturer's instructions.



3.6. CLONING OF PURIFIED PCR PRODUCTS INTO PGEM®-T EASY VECTOR

Most fragments in this study were cloned into the pGEM®-T Easy Vector (Promega). Fragments not cloned into pGEM®-T Easy Vector (Promega) were cloned into pBluescript®II SK (+) (Stratagene) according to the manufacturer's instructions on website http://www.stratagene.com/manuals/212206.pdf.

3.6.1. Ligation systems

pGEM®-T Easy Vector Systems (Promega) (including X-Gal (Promega) and IPTG) was used according to manufacturer's instructions in Promega technical manual no. 042.

T4 DNA ligase and 10x Ligation buffer (New England Biolabs) were used according to manufacturer's instructions.

Ligation of PCR products into pGEM®-T Easy Vector was performed in 10 or 15 µl volumes depending on the concentration of the insert used. The concentration of insert used was calculated following the manufacturer's instructions. In most cases, a 3:1 insert:vector molar ratio was used. The ligation reaction (1x Rapid Ligation Buffer [30 mM Tris-HCl, 10 mM MgCl2, 10 mM DTT, 1 mM adenosine triphosphate (ATP), 1% Polyethylene glycol (PEG)] 25 – 50 ng of pGEM®-T Easy Vector, ± 75 ng insert, 3 Weiss units T4 DNA Ligase) was incubated overnight at 4ºC. When using the T4 DNA ligase and 10x Ligation buffer (50 mM Tris-HCl, 10 mM MgCl2, 10 mM DTT, 1 mM adenosine triphosphate, 25 µl/ml bovine serum albumin (BSA), pH 7.5 at 25ºC) (New England Biolabs), ligations was performed as described above with a few changes: namely a 10x Ligation buffer (New England Biolabs) was used and the ligation was incubated at 16ºC overnight. For difficult ligations 5% PEG was added to the ligation (NEB) and incubated at 20ºC overnight.

3.6.2. Preparation of ultracompetent cells: Rubidium chloride

DH5α competent cells were prepared using an adaptation from Hanahan (1985).

A single colony of *E.coli* DH5α was inoculated into 5 ml of LB medium and incubated (37ºC, 225 rpm, overnight). The overnight culture was used to inoculate 500 ml LB medium containing 20 mM $MgSO_4$, in a one-liter flask. The cells were incubated (37ºC, 225 rpm), and the absorption value (A_{600}) was taken after 2 h and every 30 min thereafter until an absorpsion value of 0.4-0.6 was reached. The cells were pelleted (7 000 rpm, 5 min, 4ºC) and resuspended in 100 ml ice-cold TFB1 (30 mM KaOc, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol (v/v) – pH adjusted to 5.8 with 1 M acetic acid, filter sterilised (0.22 Micron)). The resuspended cells were combined and incubated (5 min, ice) and for the remainder of the procedure the cells were kept on ice and microfuge tubes and flasks were also chilled. Due to their high sensitivity to temperature and handling, the competent cells were treated gently. Cells were centrifuged (7 000 rpm, 5 min, 4ºC) and gently resuspended in 1/25 volume of the original culture volume of ice-cold TFBII (10 mM MOPS or PIPES, pH 6.5, 75 mM CaCl2, 10 mM RbCl, 15% glycerol (v/v) – pH adjusted to 6.5 with 1 M KOH, filter sterilized (0.22 Micron) and stored at room temperature). The competent cells were incubated (15-60 min, ice), aliquoted into 100 µl aliquots, flash frozen in an ice-cold isopropanol (pre-cooled at -80ºC) bath, and stored at -80ºC.

3.6.3. *Escherichia coli* transformation

Transformation was performed as per Sambrook *et al.,* (1989). All plasmids were transformed into competent *E. coli* DH5α competent cells, unless specified otherwise.

All of the ligation reaction were added to 100 µl of ultra competent cells in a 1.5 ml microfuge tube, and flicked gently to mix. The cells were incubated on ice for 30 min, followed by a 42ºC heatshock for 1 minute, and immediately returned to ice for 2 min. 900 µl LB broth (Merck) was added to the tube and incubated (37ºC, 155 rpm, 90 min). 100 µl of these cells were plated out onto LB plates (agar bacteriological, Biolab) containing 100 µg/ml ampicillin (Roche), the selective antibiotic for pGEM®-T Easy. Blue-white colony selection was used and all LB plates were plated with 0.016 mg/ml 5-bromo-4-chloro-3-indocyl-β-D-galactoside (X-Gal) (Promega) and 0.16 mM IPTG before plating out the bacterial culture. The plates were incubated for 16-24 hours at 37ºC. When low transformation efficiencies were expected, the remaining 900 µl transformation mix was centrifuged (1400 rpm, 2 min), and the cell pellet resuspended in 100 µl LB broth, for all potential transformants to be plated out.

3.6.4. Screening for positive white colonies by colony PCR

White colonies were screened by PCR to confirm the presence of the correct insert and orientation.

One third of a white colony was picked with a sterile wooden toothpick and the tip was inserted briefly in a 10 µl PCR cocktail (1x Bioline PCR NH_4 reaction buffer [16mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20], 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM vector specific primer, 1 µM insert specific primer, 1x cresol/sucrose loading dye, 0.5 U Bioline BIOTAQ™ DNA polymerase). PCR cycling conditions were: 1 cycle of 94°C for 5 min, 25 cycles of 94°C for 30 sec, 54°C for 45 sec, 72°C for 60 sec – 150 sec and 1 cycle of 72°C for 7 min.

A confirmed positive white colony was inoculated in 5 ml LB medium (containing 100 µg /ml ampicillin (Roche)) and incubated (37 ºC, 225 rpm, overnight).

3.6.5. Plasmid DNA purification

The Wizard®Plus SV Minipreps DNA Purification System (Promega) was used according to the manufacturer's instructions.

One and a half ml of overnight culture was pelleted (13 200 rpm, 1 min) and the pellet was resuspended in 250 µl resuspension solution (50 mM Tris-Cl pH 7.5, 10 mM EDTA, 100 µg/ml RNase A). Two hundred and fifty µl of cell lysis solution (0.2 M NaOH, 1% SDS) was added, the tube inverted 4 times and incubated (5min, 37°C). Ten µl alkaline protease solution was added to the sample, the tube inverted 4 times and incubated (5 min, room temperature). Three hundred and fifty µl of neutralization solution (4.09 guanidine hydrochloride, 0.759 M potassium acetate, 2.12 M glacial acetic acid, pH 4.2) was added, inverted 4 times and centrifuged (13 200 rpm, 10 min). The cleared lysate was decanted into a spin column within a collection tube and centrifuged (13 200 rpm, 1 min). The flow through was discarded; 750 µl wash solution (60 mM potassium acetate, 8.3 mM Tris-Cl pH 7.5, 40 µM EDTA, 60% EtOH) added to the spin column and centrifuged (13 200 rpm, 1 min). The flow through was discarded, 250 µl wash solution was added to the spin column and centrifuged (13 200 rpm, 2 min). The spin column was transferred to a sterile 1.5 ml microfuge tube, 50 µl ddH_2O was added, incubated for 1 min and centrifuged (13 200 rpm, 1 min). Plasmid DNA was stored at –20°C.

Plasmid Alkaline lysis mini-prep method was used according to Sambrook *et al.,* (1989).

A 5 ml overnight culture of bacteria was grown with antibiotic selection at 37°C. One and a half ml of the overnight culture was centrifuged (1 min, 13 200 rpm) in a 1.5 ml eppendorf tube. The supernatant was discarded and the pellet resuspended in 0.1 ml of solution1 (0.25 M Tris-Cl pH 8, 0.5 M Glucose, 0.1 M EDTA pH 8) and incubated (10 min, room temperature). Two hundred µl of solution 2 (0.2 M NaOH, 1% SDS w/v) was added and left on ice for 10 min. Hundred and fifty µl pre-cooled solution 3 (3 M potassium acetate, 2 M acetic acid) was added, inverted, left on ice for 10 min and centrifuged (5 min, 13 200 rpm). Four hundred and fifty µl of the supernatant was recovered and transferred to a 1.5 ml eppendorf tube. Three hundred µl isopropanol was added, mixed, left for 2 min at room temperature and centrifuged (10 min, 13 200 rpm). The pellet was retained and washed with 0.5 ml 70% ethanol. The ethanol was poured off, the tube was dabbed dry and the pellet was air dried for 5 min. The pellet was then resuspended in 50 µl ddH_2O and stored at -20°C.

3.7. FREEZER CULTURES

Freezer cultures of positive clones were made by adding 300 µl 50% glycerol to 700 µl of overnight culture and stored at -80°C.

3.8. SEQUENCING OF CLONES

Plasmid DNA templates were sequenced with the Applied Biosystems ABI PRISM BigDye™ Terminator v3.0 Ready Reaction Cycle Sequencing Kit following the ½ reaction protocol according to the manufacturer's instructions. The T7 and SP6 primers were used for sequencing, unless stated otherwise. Sequencing was performed by the Core DNA Sequencing Facility, Stellenbosch University.

3.9. SEQUENCE ANALYSIS

Sequence editing was performed using Chromas (version 1.45 Copyright 1998-2004, Technelysium Pty. Ltd., all rights reserved) (www.technelysium.com.au/chromas.html). Sequence comparisons were performed using the BLAST algorithm (Altschul *et al.,* 1990) against database NCBI (www.ncbi.nlm.nih.gov). Overlapping clones were analysed with BioEdit (Version 5.0.9) (Hall, 1999) and restriction sites within the overlaps were determined with Webcutter (Heiman, 1997) (www.firstmarket.com/firstmarket/cutter/cut2.html).

3.10. JOINING OF OVERLAPPING CLONES

To join overlapping clones, two methods were used namely **restriction digestion** and **primer extension PCR**.

3.10.1. Restriction digestion and ligation of inserts and vectors

A unique restriction site was identified within the overlap that did not cut elsewhere within the two fragments to be joined.

Restriction enzymes were obtained from MIB Fermentas and Roche Biochemicals and were used according to the manufacturer's instructions. DNA was incubated for 90 min with the restriction enzyme at optimal temperature and in the appropriate buffer.

Ligations were performed as described previously using T4 DNA ligase and 10x Ligation buffer (New England Biolabs) (section 3.6.1).

3.10.2. Primer Overlap Extension PCR (figure 3.2)

An adaptation of the method described by Horton *et al.*, (1989) was used.

Two cloned overlapping fragments were re-amplified out of pGEM®-T Easy using the original PCR primers for fragment 1 and the original reverse primer and a newly designed forward primer for fragment 2 (figure 3.1). The new primer was designed so that it was complementary to more or less 20 bases of the 3'-end of the first fragment. A 10x and 100x dilution of the template DNA was made and 1µl of template was added to 9 µl PCR cocktail (1x Bioline PCR NH_4 reaction buffer [16mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20], 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM forward primer, 1 µM reverse primer, 1x cresol/sucrose loading dye, 0.5 U Bioline BIOTAQ™ DNA polymerase). PCR cycling conditions were 1 cycle of 94°C for 5 min, 25 cycles of 94°C for 30 sec, 54°C for 45 sec, 72°C for 1 min – 2 min 30 sec and 1 cycle of 72°C for 7 min.

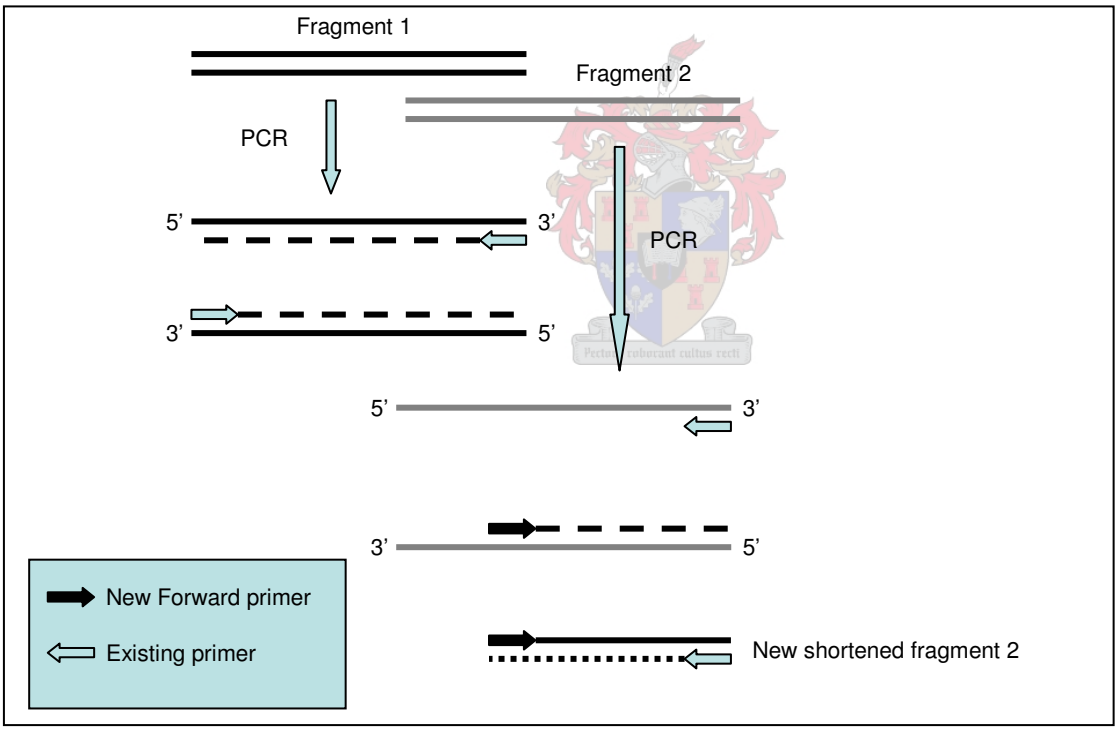


Figure 3.1. Generation of shortened fragments for primer overlap extension. Two cloned overlapping fragments were re-amplified out of pGEM®-T Easy using the original PCR primers for fragment 1 and the original reverse primer and a newly designed forward primer for fragment 2. The first 20 nt of the new shortened fragment overlap with the last 20 nt of fragment 1 to facilitate extension in the first PCR amplification step.

PCR products were electrophoresed as described previously in section 3.3. Fragments were excised from the gel, pooled into two identical excised fragments per tube (± 10-20 ng/µl), and purified as described previously in section 3.4.

Bioline BIOTAQ™ DNA polymerase adds a non-specific A at the 3'-end of PCR fragments, which had to be removed (blunted) before primer extension. This was performed with the Klenow enzyme (Roche). In a single microfuge tube 50ng of each fragment was blunted with 1 U Klenow enzyme, 1x Buffer H in a 15 µl reaction and incubated (room temperature, 40 min).

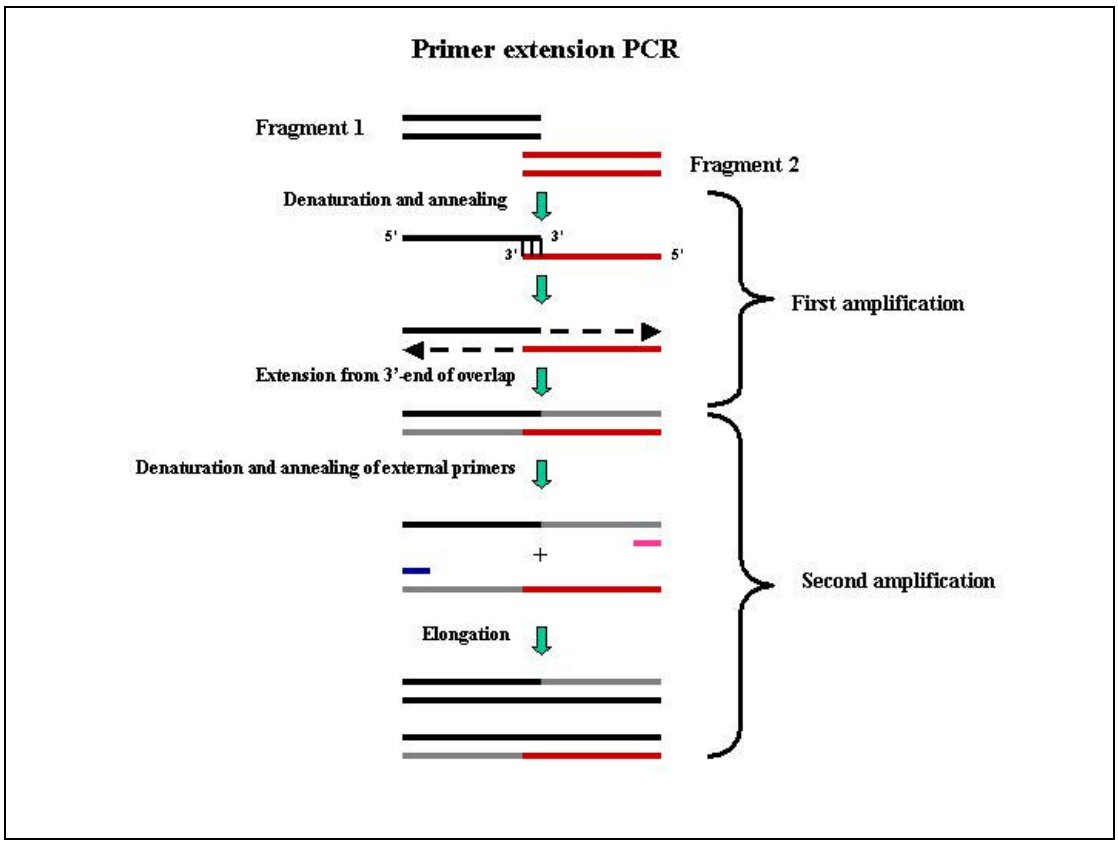


Figure 3.2. Primer overlap extension PCR. Overlapping fragment 1 and shortened fragment 2 are joined in two separate PCR steps. In the first amplification the overlap will act as primer to facilitate 5'→3' extension, followed by extension from two external primers in the second round of amplification.

3.10.2.1. First amplification

One, 2 and 3 µl of the Klenow digestion was used in 3 separate 10 µl reactions respectively to titre the amount of template required for optimal amplification. The PCR cocktail (1x Bioline PCR NH_4 reaction buffer [16mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20], 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.5 U Bioline BIOTAQ™ DNA polymerase) contained no primers, because the

generated overlap will act as primer. PCR cycling conditions were as follows: 1 cycle of 94°C for 5 min, 25 cycles of 94°C for 30 sec, 54°C for 45 sec, 72°C for 1 min – 2 min and 1 cycle of 72°C for 7 min.

3.10.2.2. Second amplification

In the second round of amplification, 1 µl of each first amplification PCR reaction was added to the 9 µl PCR cocktail (1x Bioline PCR NH_4 reaction buffer [16mM $(NH_4)_2SO_4$, 67 mM Tris-HCl pH 8.8 at 25°C, 0.01% Tween-20], 2 mM $MgCl_2$, 0.2 mM dNTPs, 1 µM **new** forward primer, 1 µM reverse primer, 1x cresol/sucrose, 0.5 U Bioline BIOTAQ™ DNA polymerase) and was subjected to the following PCR cycle: 1 cycle of 94°C for 5 min, 30 cycles of 94°C for 30 sec, 54°C for 45 sec, 72°C for 1 min – 2 min 30 sec and 1 cycle of 72°C for 7 min.

PCR reactions were electrophoresed as described previously in section 3.3. Fragments were excised from the gel, purified, quantified and cloned into pGEM®-T Easy as described previously in sections 3.4, 3.5 and 3.6.

3.11. INCORPORATION OF THE 3' POLY-A TAIL

The 3' poly-A tail was incorporated with normal PCR (figure 3.1) using primer pair GVA-5061F / GVA-ER-poly-A (table 3.1) and clone pGEM-GVA-2308 as template. The PCR cocktail was made as described previously (section 3.2.4.3) and subjected to the following PCR cycle: 1 cycle of 94°C for 5 min, 30 cycles of 94°C for 30 sec, 42°C for 30 sec, 72°C for 2 min 30 sec, 20 cycles of 94°C for 30 sec, 54°C for 30 sec, 72°C for 2 min 30 sec and 1 cycle of 72°C for 7 min. The resultant PCR product was cloned into the pGEM®-T Easy Vector (Promega) and sequenced as described previously in sections 3.6 and 3.8.

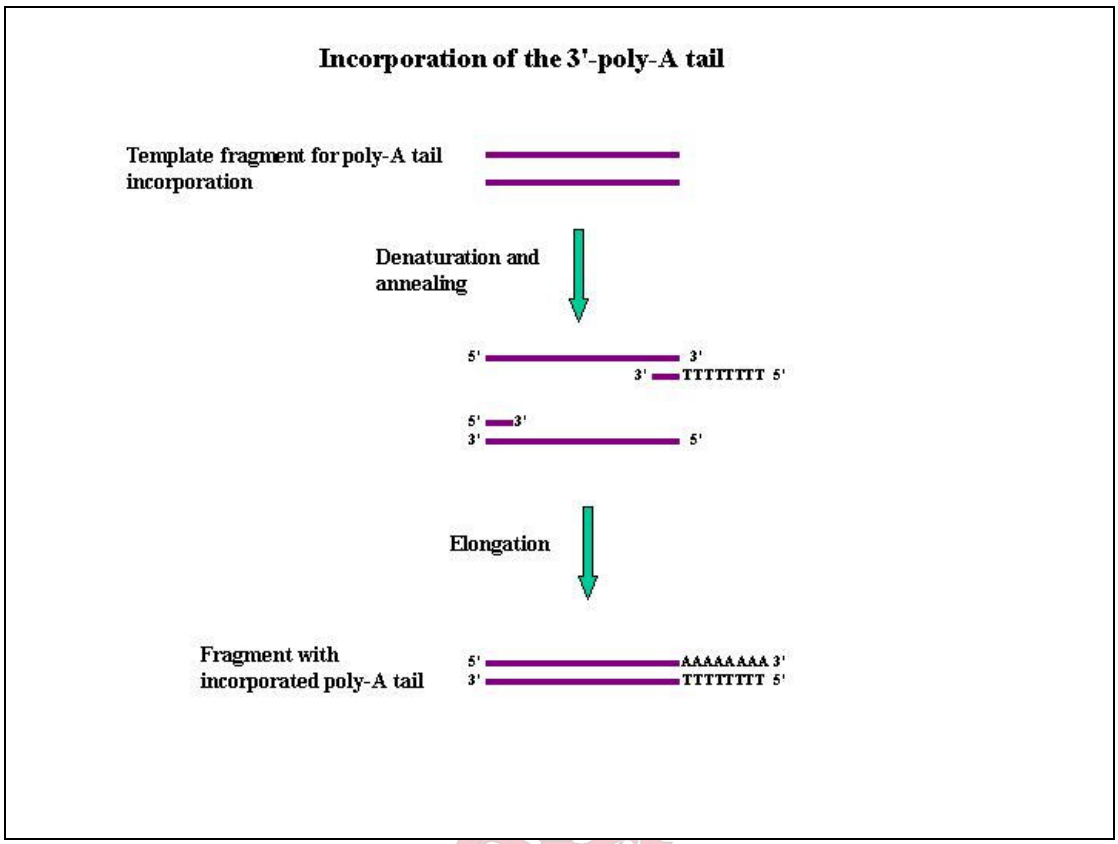


Figure 3.3. Incorporation of the 3' poly-A tail. The poly-A tail was incorporated by PCR amplification using an existing forward primer and a newly designed reverse primer containing 40 T's.

3.12. INCORPORATION OF THE 5' T7 PROMOTER

To ensure infectivity of the infectious clone, a T7 promoter had to be incorporated immediately adjacent to the 5'-end of the viral genome, as it is generally accepted that any foreign vector sequences strongly reduces infectivity (Boyer *et al*., 1994). This was performed according to Geiser *et al.,* (2001).

To incorporate a T7 promoter, a primer named Geiser-For (table 3.1), was designed. The primer consisted of 20 nucleotides at the 5'-end, corresponding exactly to the DNA sequence upstream from the point of insertion in the pGEM®-T Easy Vector, the T7 promoter sequence in the middle and the first 20 nucleotides of the 5'-end of the GVA genome, at the 3'-end. Two methods were tested for the incorporation: 3.12.1.) Incorporation of the T7 using the Geiser Technique and, 3.12.2.) Incorporation of the T7 promoter using Expand High Fidelity PCR.

3.12.1. Incorporation of the T7 using the Geiser Technique

3.12.1.1. Expand Two-step Geiser High Fidelity PCR

The DNA fragment used for mutagenesis was obtained by PCR with the Expand High Fidelity PCR system, using the GVA-Geiser-For primer together with the GVA-1188R primer. Amplifications were performed using the pGEM-GVA-3794 clone as template. High Fidelity PCR was carried out in a 50 µl reaction using 2.6 U enzyme, 20-100 ng of template plasmid DNA, 20 pmol of each primer, 0.4 mM dNTPs in the supplied 10x Expand High Fidelity PCR reaction buffer. The PCR reactions were subjected to the following cycling conditions: 1 cycle of 94°C for 2 min, 25 cycles of 94°C for 15 sec, 54°C for 30 sec, 68°C for 2 min and 1 cycle of 68°C for 10 min. The PCR reactions were run on an agarose gel as described previously in section 3.3. Three amplified fragments of expected size from separate reactions, were cut out of the gel, pooled and purified as described previously in section 3.4. The purified fragments were eluted in 20 µl ddH_2O. A second round of amplification was performed with the Expand High Fidelity PCR system, using 6 µl of purified product per reaction, and the pGEM-3794 clone as template. High Fidelity PCR was carried out in a 50 µl reaction using 2.6 U enzyme, 20-100 ng of template plasmid DNA, 0.4 mM dNTPs in the supplied 10x Expand High Fidelity PCR reaction buffer. The PCR reactions were subjected to the following cycling conditions: 1 cycle of 95°C for 2 min, 18 cycles of 95°C for 30 sec, 54°C for 30 sec, 68°C for 12 min and 1 cycle of 68°C for 12 min.

3.12.1.2. Dpn I digestion

At the end of the mutagenesis protocol, 10U *Dpn I* was added to completely remove the wild-type original methylated template vector. The incubation time was extended to 3 h at 37°C.

3.12.1.3. Purification of digestions

Digestions were purified with the DNA Clean & Concentrator™-5 system (Zymo Research) according to the manufacturer's instructions. The DNA was eluted 2 times with 6 µl ddH_2O.

3.12.1.4. Transformation

Two µl of the eluted DNA was used to transform *E. coli* DH5α cells as described previously (section 3.6.3). A few clones were tested to verify the incorporation of the T7 promoter by restriction digestion and PCR analysis.

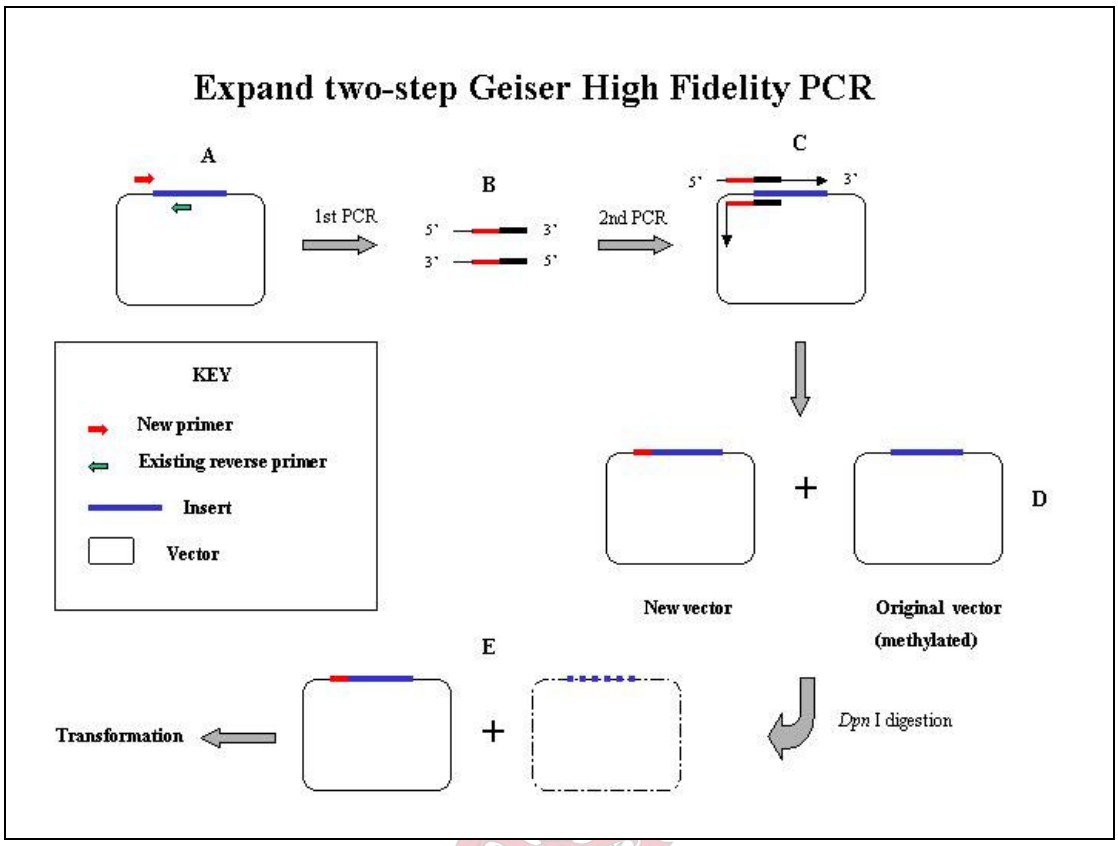


Figure 3.4. Expand Two-step Geiser High Fidelity PCR. A) Template vector showing primer binding sites. The new forward primer contains the mutation to be introduced. B) PCR product after the first amplification step, containing introduced mutation. C) New denatured product annealing to template, acting as primers in second amplification step. D) New non-methylated vector and wild-type methylated template vector resulting after second amplification. E) *Dpn* I digests wild-type methylated vector, resulting in only new non-methylated vector, with introduced mutation. This vector is transformed into *E. coli* DH5α (Geiser *et al*., 2001).

3.12.2. Incorporation of the T7 promoter using Expand High Fidelity PCR

The T7 promoter was incorporated with the Expand High Fidelity PCR system (Roche). Expand High Fidelity PCR was performed according to the manufacturer's instructions using primer pair Geiser-For/GVA-2972-R on template pGEM-GVA-2972-13. High Fidelity PCR was carried out in a 50 µl reaction using 2.6 U enzyme, 20-100 ng of template plasmid DNA, 20 pmol of each primer, 0.4 mM dNTPs in the supplied 10x Expand High Fidelity PCR reaction buffer. The PCR reactions were subjected to the following cycling conditions: 1 cycle of 94°C for 5 min, 25 cycles of 95°C for 30 sec, 52°C for 30 sec, 68°C for 2 min 30 sec and 1 cycle of 68°C for 10 min. PCR reactions were electrophoresed as described previously in section 3.3. Fragments were excised out of the gel, purified, quantified and cloned into pGEM®-T Easy as described previously in sections 3.4, 3.5 and 3.6. The

resulting clones were sequenced to confirm the incorporation of the T7 promoter as described previously in section 3.8.

3.13. PCR-BASED SITE-DIRECTED MUTAGENESIS

PCR-based site-directed mutagenesis was performed according to the protocol of Fisher & Pei, (1997). Two overlapping 21 mer primers (spanning the mutation) were designed on opposite DNA strands with the mutation incorporated near the 5'-end of the forward primer. The primers were PAGE gel purified to avoid n-1 primers. Contrary to popular opinion, the primers were not phosphorylated. Expand High Fidelity PCR was carried out in a 50 µl reaction using 2.6 U enzyme, 1-16 ng of template plasmid DNA, 50 pmol of each primer, 0.4 mM dNTPs in the supplied 10x Expand High Fidelity PCR reaction buffer. The PCR reactions were subjected to the following cycling conditions: 1 cycle of 94°C for 4 min, 16 cycles of 94°C for 1 min, the annealing temperature for 1 min, 68°C for 2 min/kb plus a cycle increment of 5 sec for each cycle and 1 cycle of 68°C for 2 min/kb. After PCR, 2 µl of the reaction was electrophoresed to confirm product formation as described in section 3.3. The remaining PCR product was digested with 60 U *Dpn* I (3h, 37°C) in the appropriate buffer. Five µl of the digestion was electrophoresed (as described in section 3.3) to confirm digestion of original methylated template. Five and 10 µl of the digestion were transformed into ultra-competent *E. coli* DH5α cells as described previously (section 3.6.3). Resulting colonies were sequenced (section 3.8) to confirm the incorporation of the mutation.

3.14. *IN VITRO* TRANSCRIPTION

In vitro RNA transcription was performed with the AmpliCap™ T7 High Yield Message Maker Kit (Epicentre®) according to the manufacturer's instructions. In a 20 µl reaction, 1 µg linearized template DNA (T7 incorporated), 1x transcription buffer, 8 µl Cap/NTP Premix, 10 mM DTT, 2 µl AmpliCap T7 Enzyme Solution, were added at room temperature in the order given, and incubated (2h, 37°C). Two µl of the reaction was electrophoresed to confirm transcription as described in section 3.3. *In vitro* RNA was used immediately to infect *N. benthamiana* plants.

3.15. INFECTION OF *N. BENTHAMIANA*

For each infection, 5 to 8 µl of *in vitro* RNA was diluted (1/20) in inoculation buffer (0.01M K_2HPO_4, 0.01M cysteine-HCl, 3% (v/v) nicotine). This was used to mechanically infect *N. benthamiana* leaves. Each leaf was mechanically injured by rubbing carborandum or celite onto it, followed by the

inoculation buffer containing the RNA or virus. The plants were observed after 14 to 28 days for the development of symptoms, followed by Rapid direct-one-tube RT-PCR (section 3.2.5) to confirm systemic infection.

Chapter 4: Results and Discussion

4.3. AMPLIFICATION OF GVA cDNA FRAGMENTS

All three described methods (section 3.2) were tested to identify the most effective method for generation of GVA cDNA fragments. These methods were dsRNA/2-step RT-PCR, total RNA/2-step RT-PCR, and direct-one-tube-RT-PCR. Grapevine virus A cDNA fragments were obtained using all three described methods.

4.1.1. Plant material

Initially, GVA cDNA fragments were generated from GVA-infected *Vitis vinifera* (I3973) plant material. Ideally, it would have been more viable to obtain all GVA cDNA fragments from GVA infected *N. benthamiana* plants. Unfortunately, no GVA infected *N. benthamiana* plants were available for fragment generation at the early stage of the project, whereas *Vitis vinifera* plant material was immediately available. We decided to start generating GVA cDNA fragments from *V. vinifera*, and revert to *N. benthamiana* if amplifications proved to be problematic. Good amplification was found when using *V. vinifera* as starting material for most cDNA fragments, but no or ineffective RT-PCR amplification was found when using some primer pairs. We therefore switched to *N. benthamiana* plants, infected with GVA isolate GTR1-2, obtained from Dr Dariusz Goszczynski (Agricultural Research Council Plant Protection Reseach Institute, Pretoria, South Africa). This specific isolate of GVA (GTR1-2) was used, because it had significant sequence homology to the published sequence from which the GVA primers were designed. GVA infected RT-PCR amplification was tested using GVA-infected *N. benthamiana* as starting material. Results showed that when using the Rapid Direct-One-Tube-RT-PCR method, a higher concentration of product could be generated from *N. benthamiana* than from *V. vinifera* (figure 4.3). This could have been due to a higher concentration of viruses in *N. benthamiana* material or due to the fact that it is easier to extract viruses/nucleic acids from tobacco than from grapevine. Five GVA cDNA fragments (generated with primers pairs that were problematic when using *V. vinifera*) were obtained from *N. benthamiana*, and seven fragments were obtained from GVA-infected *V. vinifera* (even though better amplification was achieved when using *N. benthamiana*). The fact that two different plant sources were used could result in the construction of a

hybrid virus. The aim of this thesis is to construct a universal GVA clone, and not a clone of a specific isolate. Grapevine virus A is thought to be associated with Shiraz disease, and there is no evidence that only a specific GVA isolate is involved. Therefore, we did not have to obtain GVA cDNA fragments from Shiraz diseased vines and the construction of a hybrid clone of GVA would not present a problem. To justify this statement, we managed to amplify GVA cDNA fragments from Shiraz diseased vines, using GVA specific primers (results not shown).

4.1.2. Total RNA extraction followed by RT-PCR

A yield of approximately 500 ng of total RNA was obtained from 500 µg of leaf veins/bark strips using this method (figure 4.1). A number of GVA cDNA fragments were obtained with RT-PCR amplification by using this method. The largest PCR fragment obtained by this method was 1,579 bp using primer pair GVA-2793-For/GVA-4370-Rev (figure 4.1). This fragment was the largest fragment obtained in this study, but was not selected for further analysis, because of spurious results. Inconsistent PCR amplification was found when re-amplifying the 1,579 bp fragment from the clone, and amplified bands had a lower electrophoretic mobility than was expected (results not shown). As mentioned previously, this method proved not to be successful when using some primer pairs. This could have been due to the negative influence of RNA secondary structure on amplification or due to the presence of a high GC content in some areas of the GVA genome to be amplified. This problem is not limited to this method and could have influenced amplification with the other methods as well. The main problem could have been the fact that the RNA:primer ratio in the reverse transcription step, and the cDNA:primer ratio in the PCR reaction was not optimal and had to be titered. Incorrect ratios could have led to non-optimal PCR amplification. Titration and optimization are time consuming, and we had several fragments to amplify with different primer pairs, therefore we turned our focus to generation of fragments with the Rapid Direct-One-Tube-RT-PCR, which yielded better results. Nevertheless, two generated fragments resulting from the total RNA extraction followed by RT-PCR method were selected for construction of the full-length clone namely, GVA-1002 and GVA-1005. These two fragments were selected based on their high sequence homology to the published sequence (table 4.2).

4.1.3. dsRNA extraction followed by RT-PCR

A good yield of dsRNA was obtained and used in two-step RT-PCR (figure 4.2). A number of GVA fragments were obtained with this method (figure 4.2), but did not give us consistent results. No fragments obtained with this method were selected for the final full-length GVA construct.

4.1.4. Rapid Direct-One-Tube-RT-PCR

Results showed that the Rapid Direct-One-Tube-RT-PCR method was the best method for generating GVA cDNA fragments, as there was consistent amplification of cDNA fragments, which was reproducible for different primer pairs. This method had two significant advantages when compared to the dsRNA/2-step RT-PCR and total RNA/2-step RT-PCR methods. Firstly, the reaction was less time

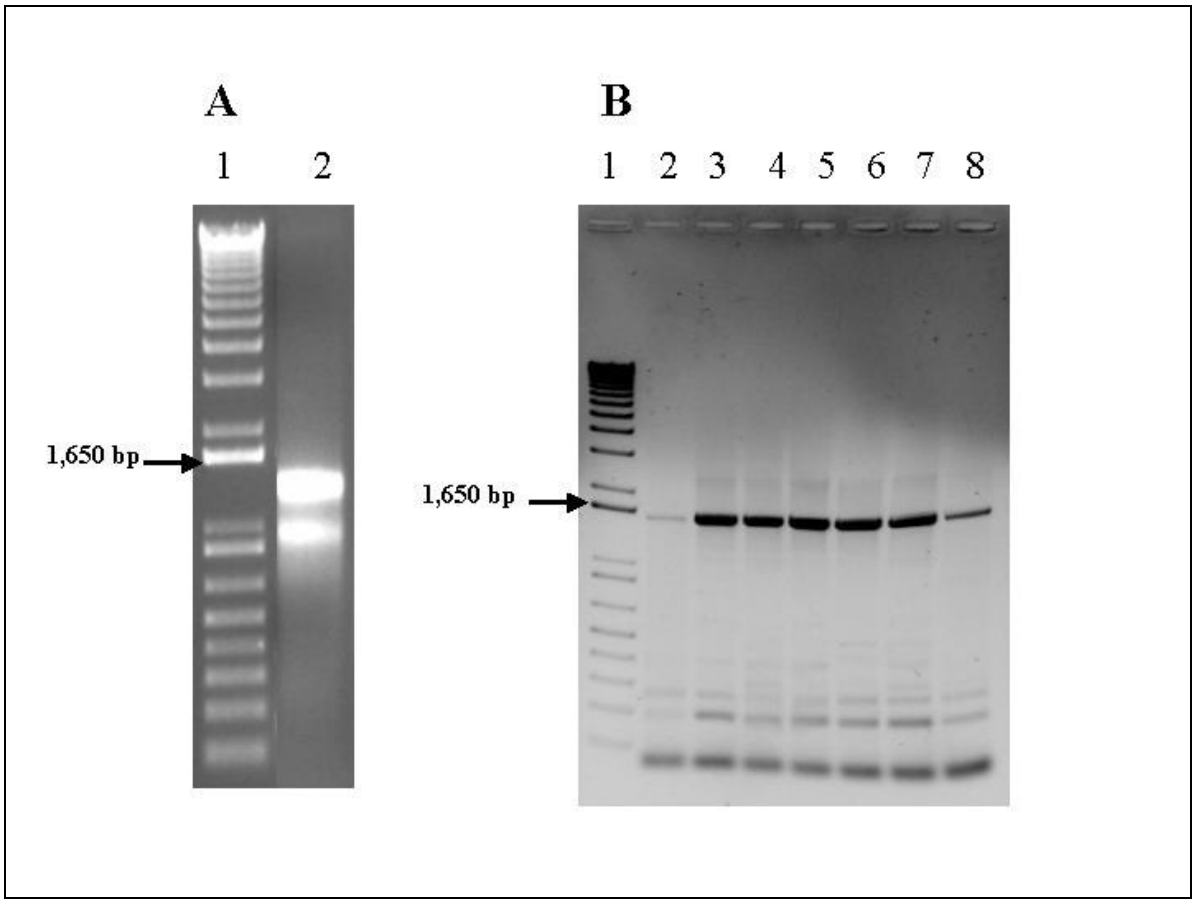


Figure 4.1. Total RNA extraction followed by RT-PCR. **A)** Lane 1: 1kb plus DNA marker. Lane 2: yield of approximately 500 ng total RNA. **B)** Lane 1: 1kb plus DNA marker. Lanes 2-8: RT-PCR product of 1,579 bp generated with primer pair GVA-2793-For/GVA-4370-Rev.

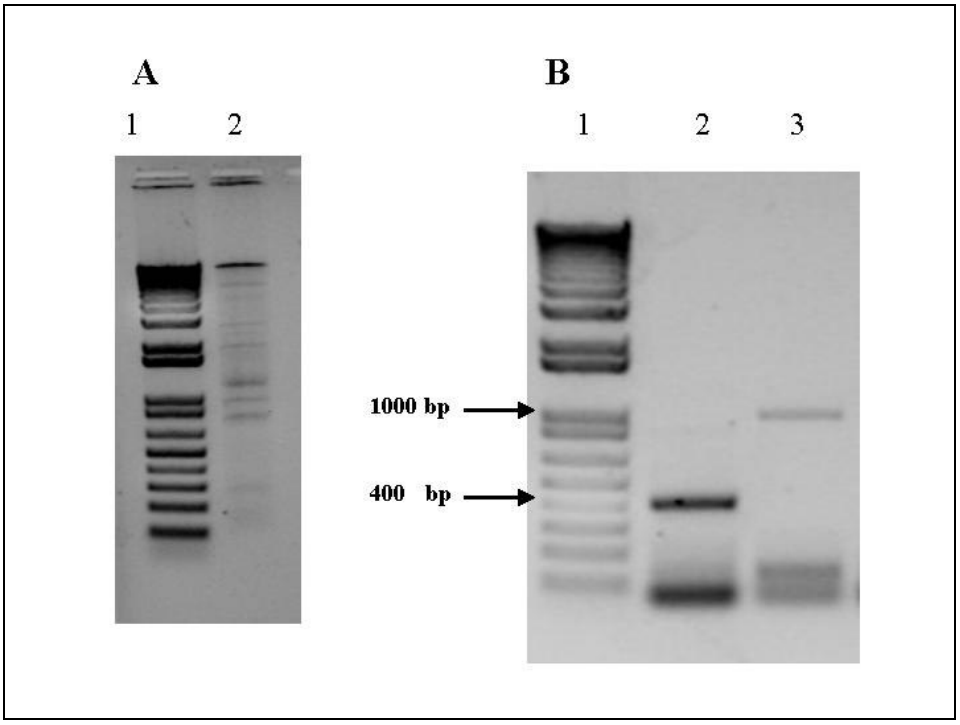


Figure 4.2. dsRNA extraction followed by RT-PCR. **A)** Lane 1: 1kb plus DNA marker. Lane 2: 10 µl dsRNA. **B)** Lane 1: 1kb plus DNA marker. Lane 2: 428 bp fragment generated with GVA-Int-For/GVA-Int-Rev. Lane 3: 1002 bp fragment generated with GVA-Int-For/GVA-2793-Rev.

consuming, as both the reverse transcription and PCR steps could be performed in one buffer, in a single tube. Secondly, the sensitive RNA and laborious dsRNA extraction steps were eliminated as no extraction step was necessary. One drawback of this method is the fact that the *Taq* DNA polymerase enzyme had to be used, instead of a high fidelity enzyme with proofreading capability. This may have decreased the fidelity of amplification and could have led to incorporation of mutations in the genome, which could render the final GVA clone non-infectious. These mutations could conversely also lead to an infectious clone that may induce more severe symptoms than the parental virion RNAs (Hamilton & Baulcombe, 1989; Hayes & Buck, 1990). As mentioned previously (section 2.3.3.2) it has been reported that in some cases infectivity of a full-length clone can be restored by exchanging a specific region of the cDNA with a fragment corresponding to the same region from an independent cDNA clone (Ahlquist *et al.,* 1984). Even though *Taq* DNA polymerase was used in the Rapid Direct-One-Tube-RT-PCR method, it has been reported that PCR with *Taq* DNA polymerase, can be successfully applied to obtain infectious clones (Hayes & Buck, 1990), despite the high error rate of the enzyme (Keohavong & Thilly, 1989). *Taq* DNA polymerase was used in all three above methods for the sake of consistency and all methods were optimized for this enzyme. As mentioned, this method gave the best results and most of the fragments used in this study were obtained using this method. The largest

fragment obtained with this method was 1099 bp, using primer pair GVA-5880-For/GVA-6979-Rev (figure 4.4). A total of ten fragments resulting from this method, were selected for further GVA construct assembly namely, GVA-620, GVA-924, GVA-532, GVA-759, GVA-965, GVA-951, GVA-646, GVA-908, GVA-1099 and GVA-501. In some cases, a faint amplification product could be seen on a gel after the first round of RT-PCR amplification. We found that by excising this band, purifying it, and re-amplifying it, a suitable concentration of PCR product was obtained for cloning (figure 4.4). This was not optimal, because the presence of a second PCR step increased the probability that mutations could be incorporated. Unfortunately we could not obtain a sufficient amount of product using an alternative method.

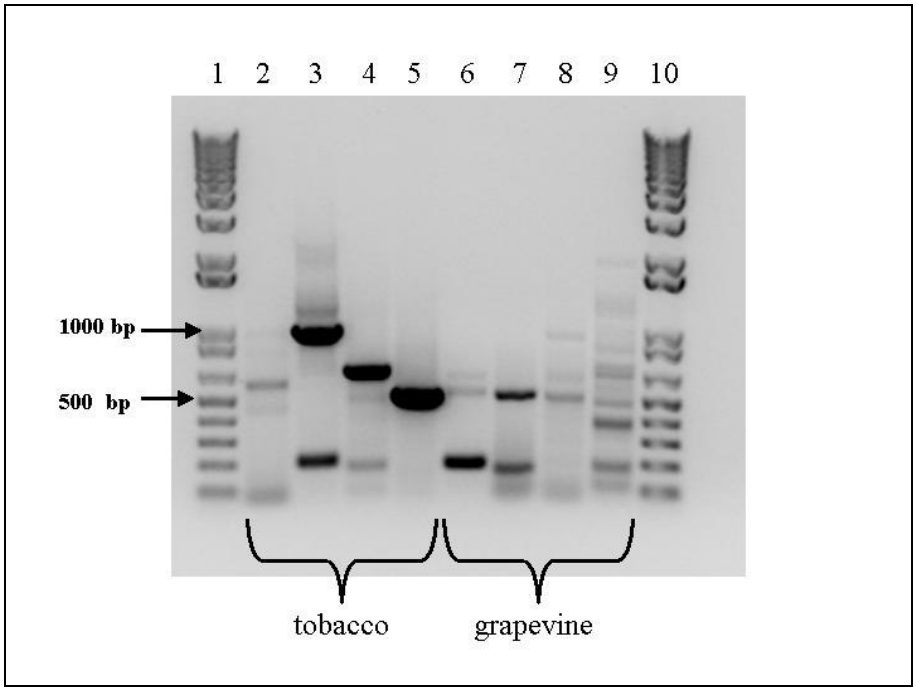


Figure 4.3. Rapid Direct-One-Tube-RT-PCR. Lanes 1 and 10: 1kb plus DNA marker. **Lanes 2 to 5**) fragments generated from *N. benthamiana*. Lane 2: faint RT-PCR product of 588 bp using primer pair GVA-1027-For/GVA-1615-Rev. Lane 3: RT-PCR product of 965 bp using primer pair GVA-2007-For/GVA-2972-Rev. Lane 4: RT-PCR product of 646 bp using primer pair GVA-4821-For/GVA-5467-Rev. Lane 5: RT-PCR product of 501 bp using primer pair GVA-6848-For/GVA-Ext-Rev. **Lanes 6 to 9**) fragments generated from *V. vinifera*. Lane 6: faint non-optimized RT-PCR product of 620 bp using primer pair GVA-Ext-For/GVA-620-Rev. Lane 7: RT-PCR product of 501 bp using primer pair GVA-6848-For/GVA-Ext-Rev. Lane 8: unsuccessful RT-PCR using primer pair GVA-264-For/GVA-1188-Rev. Lane 9: unsuccessful RT-PCR using primer pair GVA-4821-For/GVA-5467-Rev.

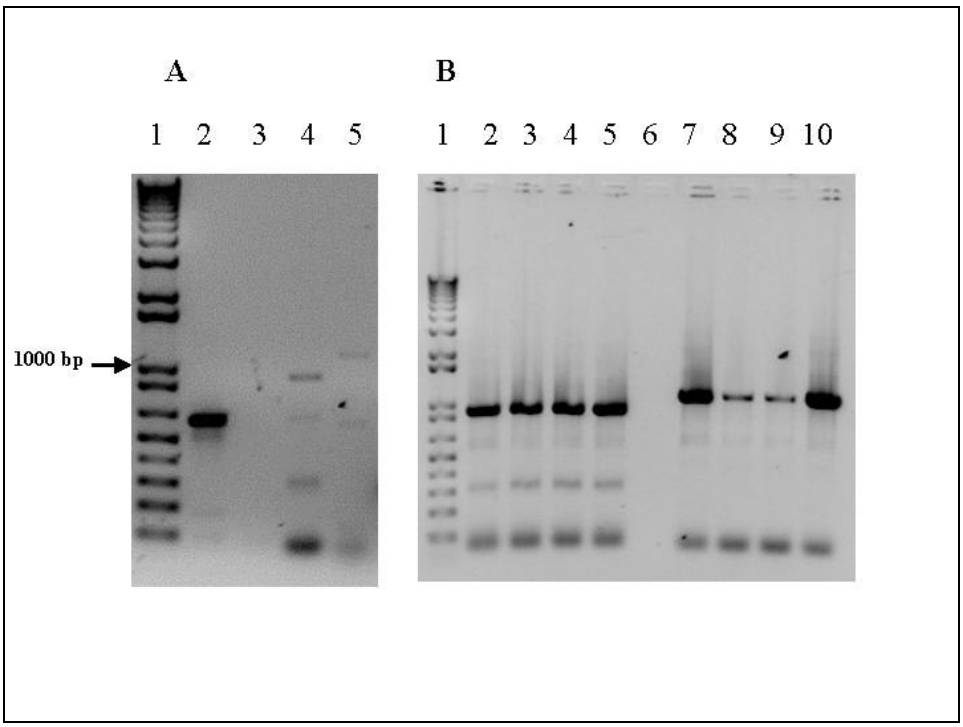


Figure 4.4. A) Rapid Direct-One-Tube-RT-PCR. Lane 1: 1kb plus DNA marker. Lane 2 and 3: not applicable. Lane 4: Faint 908 bp RT-PCR product generated with primer pair GVA-5233-For/GVA-6141-Rev. Lane 5: Faint 1099 bp RT-PCR product generated with primer pair GVA-5880-For/GVA-6979-Rev. B) Re-amplification of faint RT-PCR fragments. Lane 1: 1kb plus DNA marker. Lanes 2-5: re-amplified 908 bp fragment. Lane 6: empty. Lanes 7-10: re-amplified 1099 bp fragment.

4.4. FULL-LENGTH GVA CONSTRUCT ASSEMBLY FROM TWELVE SELECTED FRAGMENTS

As mentioned previously, several fragments were obtained with different techniques (described above) and primer pairs, but twelve fragments (listed in table 4.1) were selected to construct the full-length clone. All fragments were cloned into the pGEM®-T Easy Vector by TA-cloning as described previously (section 3.6) and sequenced (section 3.8). Sequence results [analysed with BioEdit (Version 5.0.9) (Hall, 1999)], revealed that no insertions or deletions were incorporated within any of the twelve clones (sequence results shown in Appendix 3). The presence of point mutations (such as transitions and transversions) could not be analyzed due to the fact that a difference at a single nucleotide position (compared to the published sequence) could not be assigned to either an incorporated mutation or a natural sequence variation (quasispecies).

Table 4.1. Table of 12 selected cDNA fragments for final clone construction.

Selected cDNA fragments for construction of the full-length clone			
Fragment name	*Plant source*	*Primers used* (table 3.1)	*Technique used to obtain*
1) GVA-5'-620	I 3973	GVA-Ext-For/GVA-620-Rev	1-step RT-PCR
2) GVA-924	*N. benthamiana*	GVA-264-For/GVA-1188-Rev	1-step RT-PCR
3) GVA-532	I 3973	GVA-922-For/GVA-1434-Rev	1-step RT-PCR
4) GVA-759	*N. benthamiana*	GVA-1434-For/GVA-2193-Rev	1-step RT-PCR
5) GVA-965	*N. benthamiana*	GVA-2007-For/GVA-2972-Rev	1-step RT-PCR
6) GVA-1002	I 3973	GVA-2793-For/GVA-Int-Rev	Total RNA/RT-PCR
7) GVA-1005	I 3973	GVA-Int-For/GVA-4370-Rev	Total RNA/RT-PCR
8) GVA-951	I 3973	GVA-4110-For/GVA-5061-Rev	1-step RT-PCR
9) GVA-646	*N. benthamiana*	GVA-4821-For/GVA-5467-Rev	1-step RT-PCR
10) GVA-908	I 3973	GVA-5233-For/GVA-6141-Rev	1-step RT-PCR
11) GVA-1099	I 3973	GVA-5880-For/GVA-6979-Rev	1-step RT-PCR
12) GVA-501-3'	*N. benthamiana*	GVA-6848-For/GVA-Ext-Rev	1-step RT-PCR

Fragments were selected based on sequence homology (least number of possible misincorporations by PCR) to GVA sequences in Genbank, and sequence homology within overlaps of adjacent GVA fragments. Fragments with higher sequence homology were preferred to circumvent the possibility of amplifying lethal mutants. Lethal mutants would render the final clone non-infectious. The sequence homology of the selected fragments, to GVA sequences in Genbank (www.ncbi.nlm.nih.gov), is given in table 4.2. Homology values of between 83-96 % were obtained with E-values of mostly 0.0 implying that the obtained GVA fragments are very similar to known published GVA sequences. A theoretic, full-length sequence of the twelve fragments was constructed with BioEdit (Version 5.0.9) (Hall, 1999). The sequence was translated (www.expasy.ch) and the translation products of all five ORFs, examined to confirm that no frameshifts or premature stop codons were present.

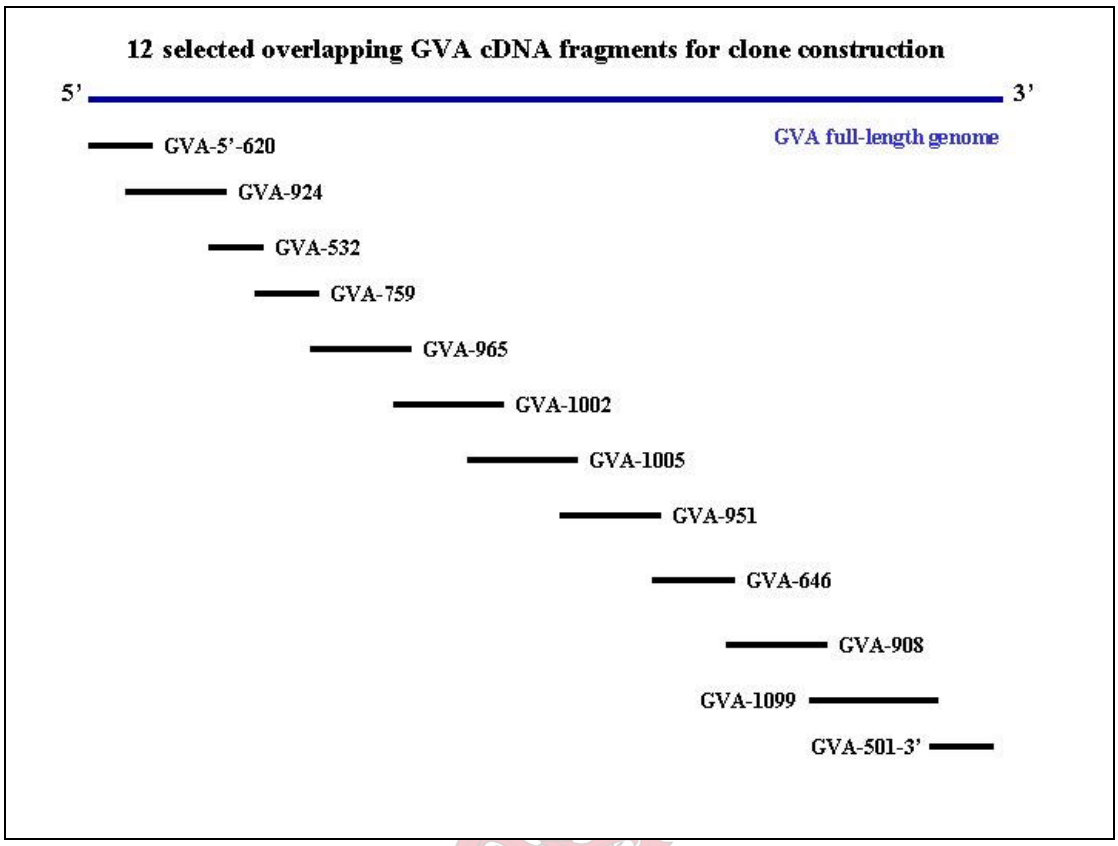


Figure 4.5. The 12 selected cDNA fragments and their relative genomic positions. Twelve overlapping GVA cDNA fragments were selected for final clone assemble based on their sequence homology to known GVA sequences. Fragments were named according to their relative sizes (for example GVA-924 would be approximately 924 bp in size).

We used three approaches to assemble the 12 selected fragments into a final full-length clone: 1) Restriction digestion followed by ligation, 2) Overlap extension PCR and 3) primer overlap extension PCR. A serious consideration was the fact that adjacent fragments did not have a 100% sequence homology within the overlap. This can be attributed to the quasispecies nature of the virus or to incorporated mutations. We attempted to identify unique restriction sites within the overlap that did not cut anywhere else in the two fragments to be joined. The sequence of digestion and ligation was also significant and care had to be taken so that downstream digestions would not cut the clone again. The restriction sites had to be present in the overlaps of both fragments to be joined, and if the site was present anywhere else in the GVA genome, the sequence of digestions had to be considered carefully. We were able to identify unique sites within some overlaps, but in most cases the overlap was too short, or no unique site was present, forcing us to take another approach. We tested overlap extension PCR, which meant that the whole overlap would act as primer for the initial PCR reaction (3.10.2). We were able to join fragments in this way. The fact that we could not calculate the melting temperature of

Table 4.2. Sequence homology of selected cDNA fragments to sequence gi/14056114 in Genbank

Sequence homology of selected cDNA fragments to gi/14056114 in Genbank			
Fragment name	*Clone Name*	*% homology*	*E-value*
1) GVA-5'-620	pGEM-GVA-5'-620	95%	0.0
2) GVA-924	pGEM-GVA-924	87%	0.0
3) GVA-532	pGEM-GVA-532	83%	2e-91
4) GVA-759	pGEM-GVA-759	89%	0.0
5) GVA-965	pGEM-GVA-965	84%	e-158
6) GVA-1002	pGEM-GVA-1002	93%	0.0
7) GVA-1005	pGEM-GVA-1005	93%	0.0
8) GVA-951	pGEM-GVA-951	92%	0.0
9) GVA-646	pGEM-GVA-646	88%	0.0
10) GVA-908	pGEM-GVA-908	93%	0.0
11) GVA-1099	pGEM-GVA-1099	93%	0.0
12) GVA-501-3'	pGEM-GVA-501-3'	96%	0.0

the overlap (some overlaps very long, and two overlaps were not 100% homologous) and that we did not have 100% sequence homology prompted us to shorten the overlap to the length of a ± 20 merprimer with 100% sequence homology and use primer overlap extension PCR (3.10.2). We thus had to shorten the overlaps for primer extension PCR, and had to enlarge them for restriction digestion in order to increase the probability of finding a unique site (4.3.10). Bioline BIOTAQ™ DNA polymerase adds a non-specific A at the 3'-end of PCR fragments, which had to be removed (blunted) before primer extension PCR. This was performed with the Klenow enzyme (Roche). This enzyme is the large fragment of DNA polymerase I that carries the 5'→ 3' polymerase and the 3'→ 5' exonuclease activities of the intact polymerase, but lacks the 5'→ 3' exonuclease activity of the native enzyme. The full-length infectious clone was assembled in 18 steps using the 12 selected cDNA fragments.

4.2.1. Step 1: Joining of fragments GVA-5'-620 and GVA-924 by restriction digestion (fig. 4.6)

Overlapping fragments GVA-5'-620 and GVA-924 were joined by restriction digestion, followed by ligation, after a unique *Sph* I site was identified within the overlap. Clone pGEM-GVA-5'-620 was digested with *Not* I and *Sph* I generating a fragment of approximately 500 bp. pGEM-GVA-924 was

digested with *Sph* I and *Sal* I resulting in a fragment of approximately 750 bp. pGEM®-T Easy Vector was digested with *Not* I and *Sal* I. The two generated fragments were ligated into the digested vector in a three-way ligation. The resulting clone was named pGEM-GVA-1188. The clone was confirmed and the correct orientation was determined by *Sph* I digestion and sequencing.

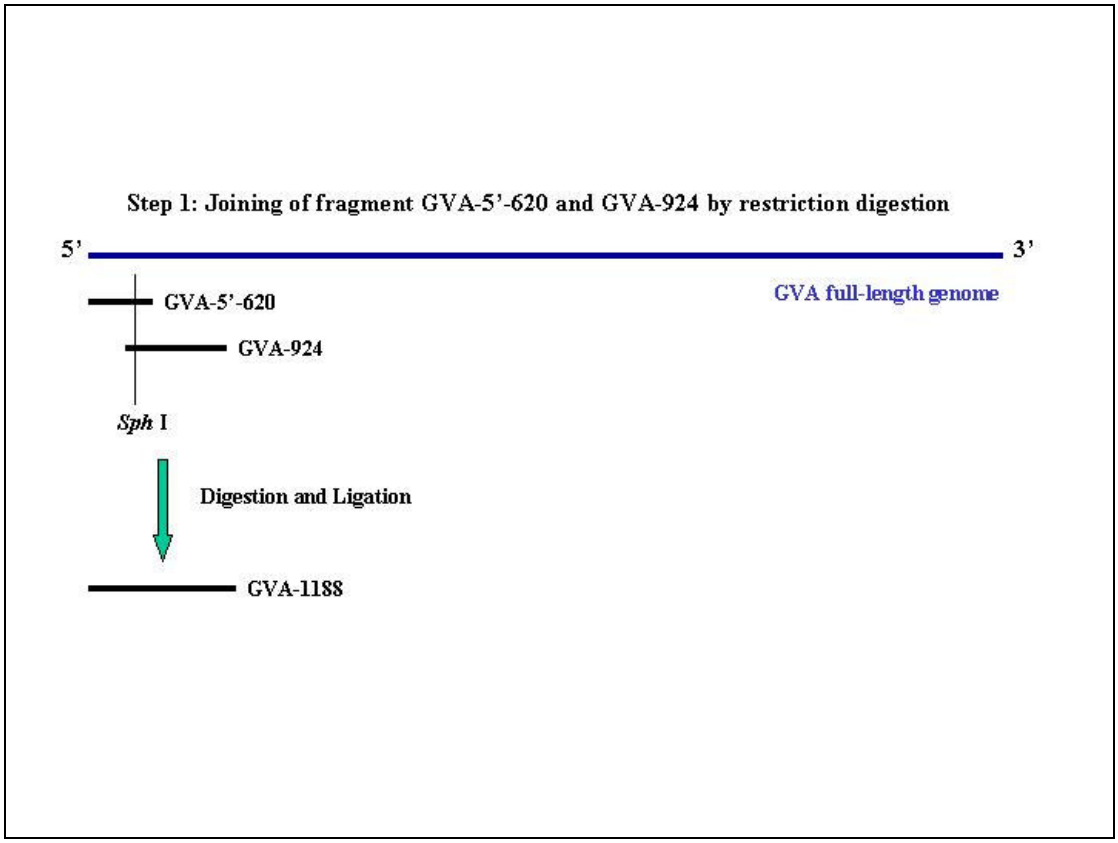


Figure 4.6. Step 1: Joining of fragments GVA-5'-620 and GVA-924 by restriction digestion. Overlapping fragments GVA-5'-620 and GVA-924 were joined by restriction digestion, followed by ligation, after a unique *Sph* I site was identified within the overlap. The resulting clone was named pGEM-GVA-1188.

4.2.2. Step 2: Joining of fragments GVA-1005 and GVA-951 by restriction digestion (fig. 4.7)

Overlapping fragments GVA-1005 and GVA-951 were joined by restriction digestion, followed by ligation, after a unique *Sph* I site was identified within the overlap. The same procedure was used, as described in section 4.3.1. The resulting clone was named pGEM-GVA-1694.

4.2.3. Step 3: Joining of fragments GVA-908 and GVA-1099 by restriction digestion (fig. 4.8)

Overlapping fragments GVA-908 and GVA-1099 were joined by restriction digestion, followed by ligation, after a unique *Eco* RV site was identified within the overlap. The same procedure was used, as described in section 4.3.1. The resulting clone was named pGEM-GVA-1746.

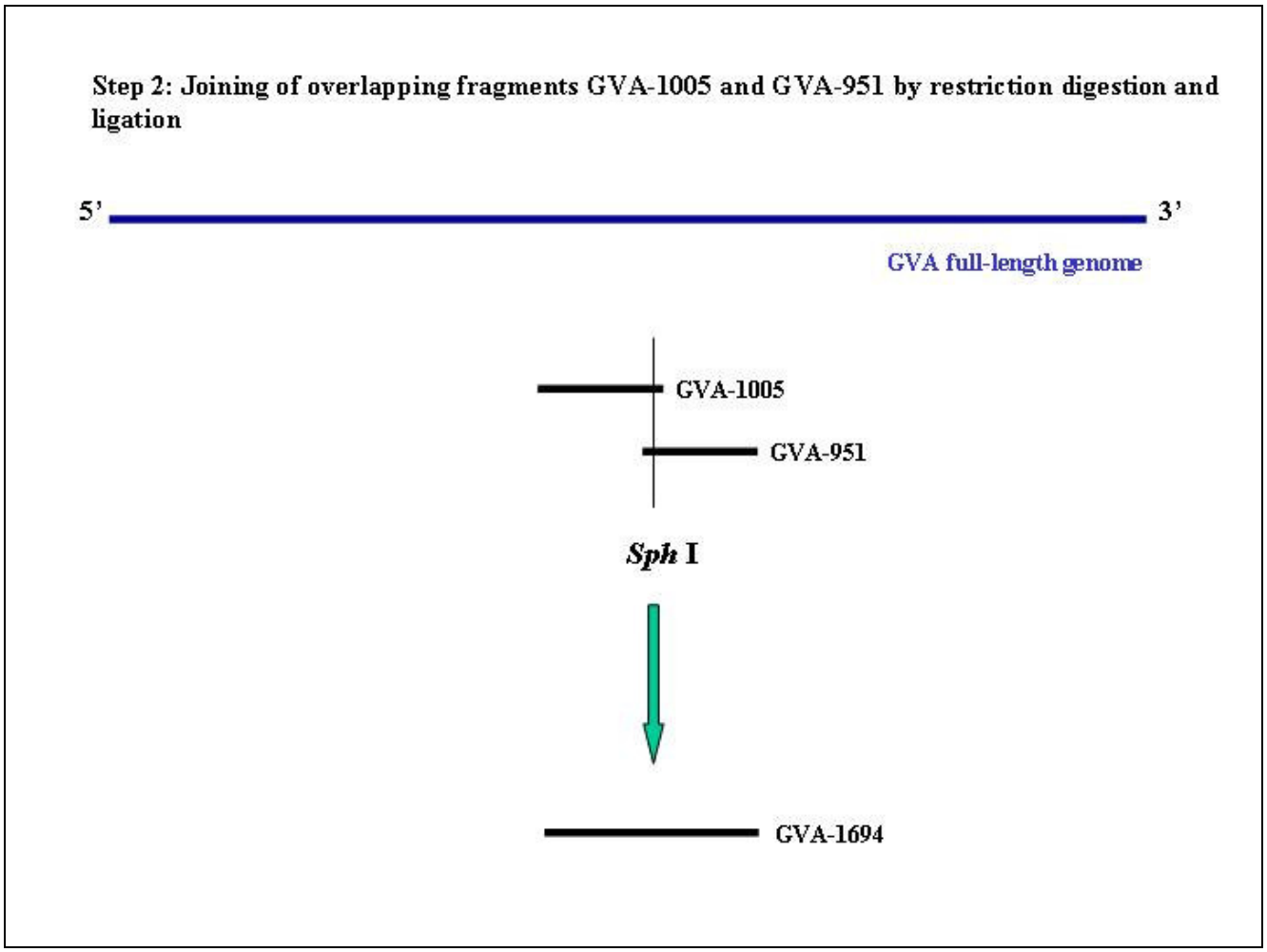


Figure 4.7. Step 2: Joining of fragments GVA-1005 and GVA-951 by restriction digestion. Overlapping fragments GVA-1005 and GVA-951 were joined by restriction digestion, followed by ligation, after a unique *Sph* I site was identified within the overlap. The resulting clone was named pGEM-GVA-1694.

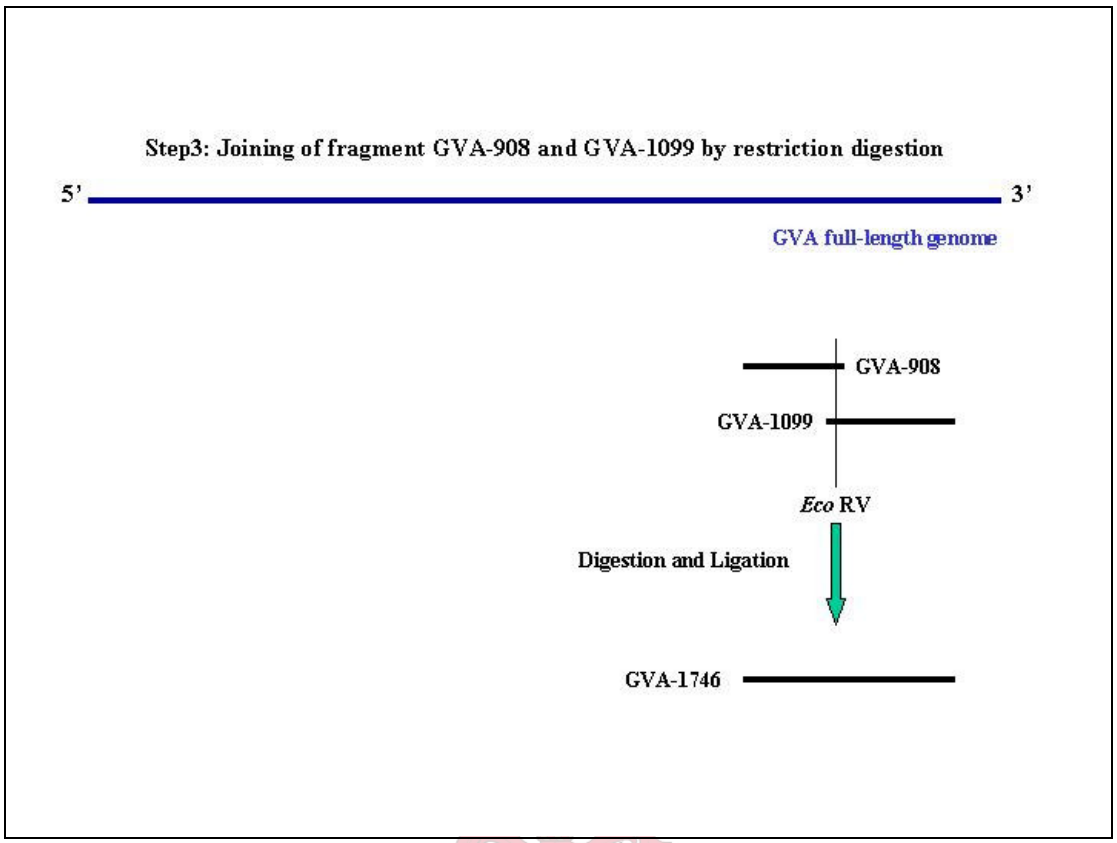


Figure 4.8. Step 3: Joining of fragments GVA-908 and GVA-1099 by restriction digestion. Overlapping fragments GVA-908 and GVA-1099 were joined by restriction digestion, followed by ligation, after a unique *Eco* RV site was identified within the overlap. The resulting clone was named pGEM-GVA-1746.

4.2.4. Step 4: Joining of fragments GVA-532 and GVA-759 by primer overlap extension (fig. 4.9)

No unique restriction site could be identified within this small overlap therefore the two fragments (GVA-532 and GVA-759) had to be joined by primer overlap extension. This was once again not optimal, because it involved another PCR step in which no proofreading enzyme could be used. This was because the method was optimized for *Taq* polymerase and meant that misincorporations could be introduced. Overlapping fragment GVA-532 was generated with primer pair GVA-922-For/GVA-1434-Rev, whereas fragment GVA-759 was generated with primer pair GVA-1434-For/GVA-2193-Rev. By using primer pair GVA-922-For/GVA-2193-Rev, fragment GVA-1271 was generated and the resulting clone was named pGEM-GVA-1271. The correct orientation was determined by colony PCR and the clone integrity was confirmed by sequencing.

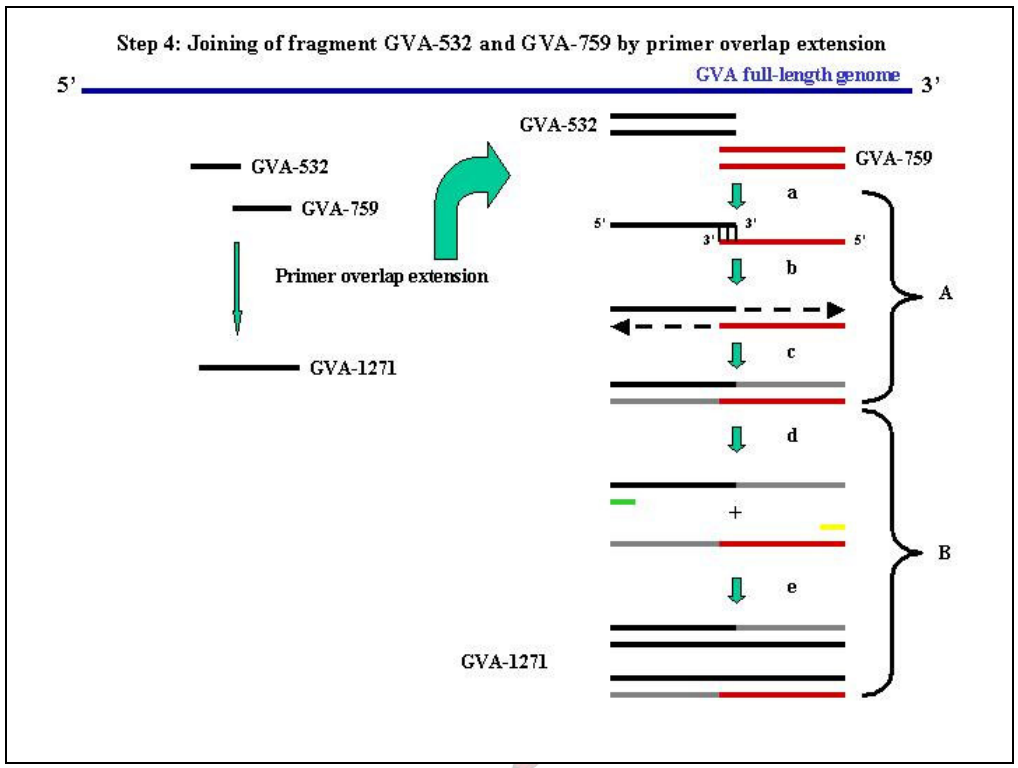


Figure 4.9. Step 4: Joining of fragments GVA-532 and GVA-759 by primer overlap extension. A) First amplification. B) Second amplification. a and b) Denaturing and annealing, c) Extension from 3'-end of overlap, d) Denaturing and annealing of external primers, e)Elongation. Overlapping fragment GVA-532 was generated with primer pair GVA-922-For/GVA-1434-Rev, whereas fragment GVA-759 was generated with primer pair GVA-1434-For/GVA-2193-Rev. By using primer pair GVA-922-For/GVA-2193-Rev, fragment GVA-1271 was generated and the resulting clone was named pGEM-GVA-1271.

4.2.5. Step 5: Joining of fragments GVA-965$_s$ and GVA-1002$_s$ by primer overlap extension (fig. 4.11)

When using primer overlap extension PCR, shorter fragments [(s) = shortened] had to be generated in order to produce an overlap from which extensions can be initiated (figure 3.1 and 4.12). Overlapping fragments GVA-965 and GVA-1002 were joined by primer overlap extension PCR. Fragment GVA-965$_s$ (shortened) was generated with primer pair GVA-2193-For/GVA-2972-Rev, whereas fragment GVA-1002$_s$ was generated with primer pair GVA-2972-For/GVA-3794-Rev. By using primer pair GVA-2193-For/GVA-3794-Rev, fragment GVA-1601 was generated and the resulting clone was named pGEM-GVA-1601. The correct orientation was determined and the clone was confirmed by PCR.

Table 4.3. Shortened (s) fragments for Primer Overlap extension

Shortened fragment	**Approximate New Size**	**Primers used**
GVA-965_s	800 bp	GVA-2193-For/GVA-2972-Rev
GVA-1002_s	842 bp	GVA-2972-For/GVA-3794-Rev
GVA-1746_s	1532 bp	GVA-5467-For/GVA-6979-Rev
GVA-501_s	392 bp	GVA-6979-For/GVA-Ext-Rev
GVA-1271_s	1025 bp	GVA-1188-For/GVA-2193-Rev
GVA-1694_s	1287 bp	GVA-3794-For/GVA-5061-Rev
GVA-646_s	426 bp	GVA-5061-For/GVA-5467-Rev

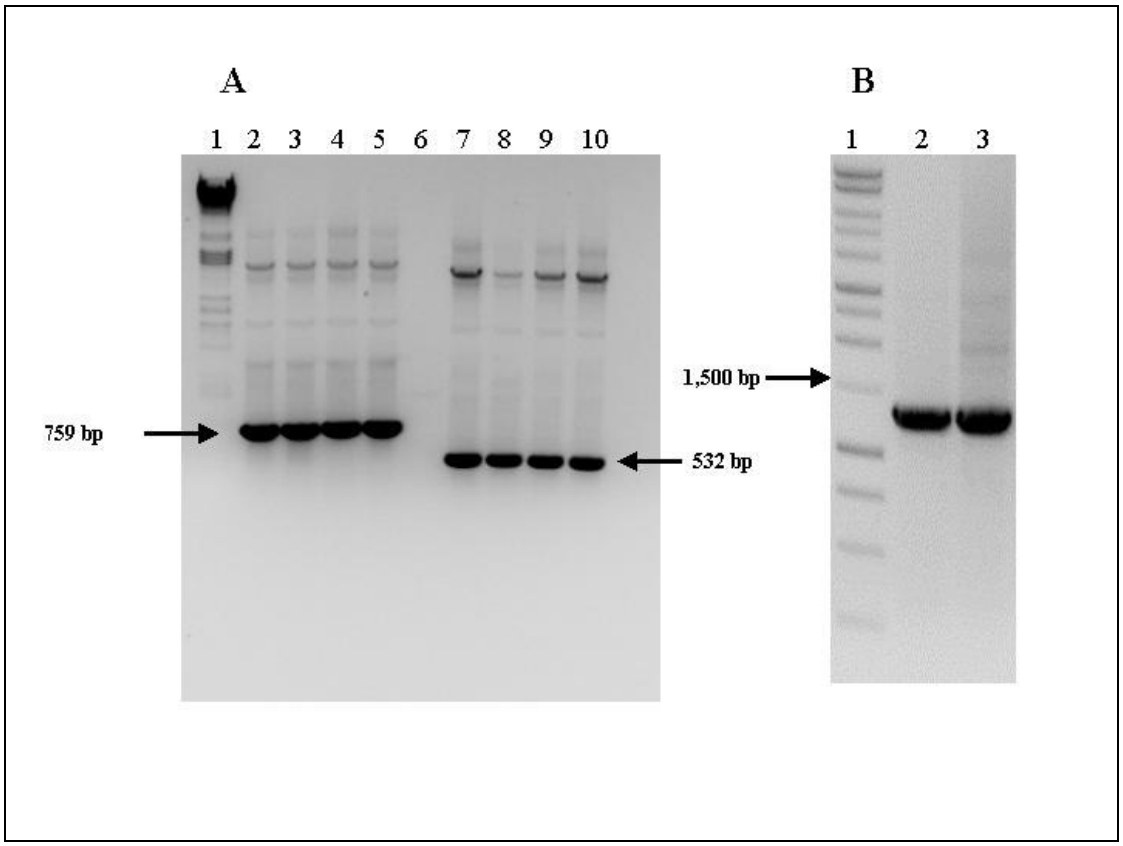


Figure 4.10. Step 4: Joining of fragments GVA-532 and GVA-759 by primer overlap extension A) Initial amplification of fragments. Lane 1: λ *Pst* I marker. Lanes 2-5: 759 bp fragment generated with primer pair GVA-1434-For/GVA-2193-Rev. Lanes 7-10: 532 bp fragment generated with primer pair GVA-922-For/GVA-1434-Rev. B) Primer overlap extension PCR. Lane 1: 1 kb DNA ladder. Lanes 2 and 3: new 1271 bp fragment generated with primer pair GVA-922-For/GVA-2193-Rev.

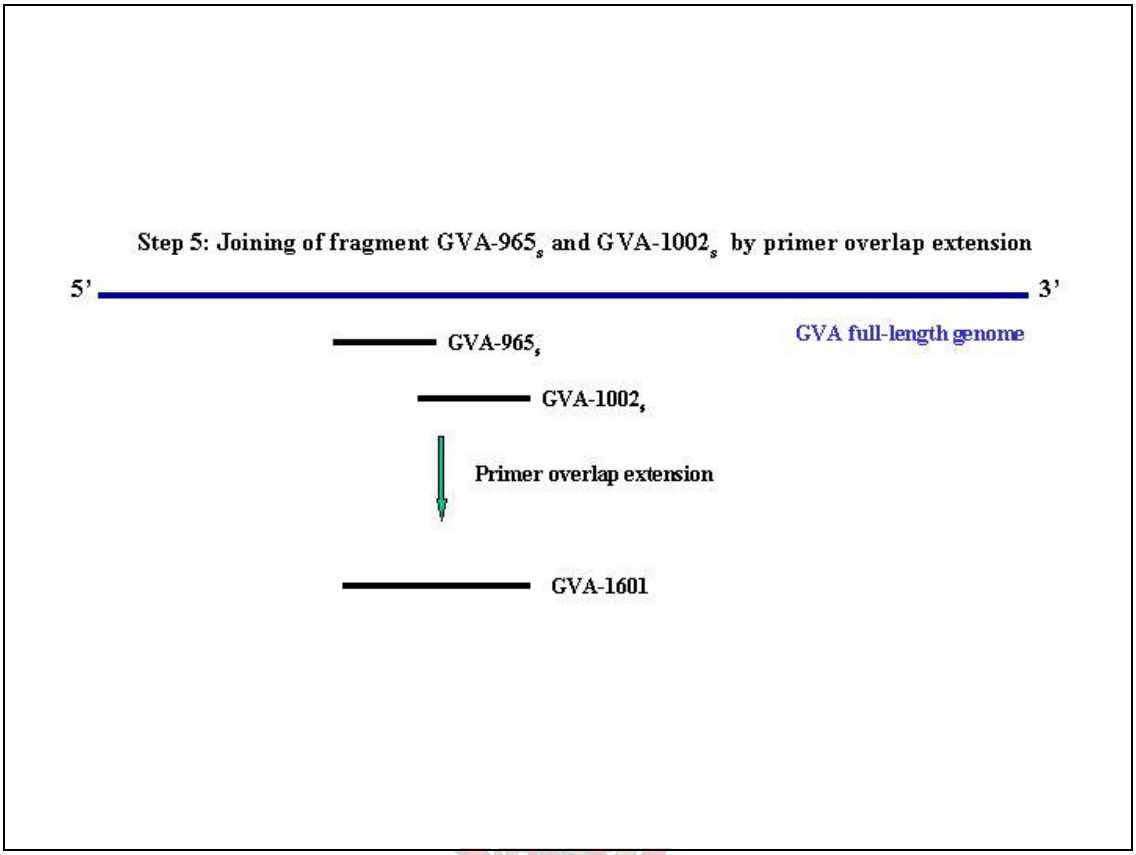


Figure 4.11. Step 5: Joining of fragments GVA-965_s and GVA-1002_s by primer overlap extension. Overlapping fragments GVA-965 and GVA-1002 were joined by primer overlap extension PCR. Fragment GVA-965_s (shortened) was generated with primer pair GVA-2193-For/GVA-2972-Rev, whereas fragment GVA-1002_s was generated with primer pair GVA-2972-For/GVA-3794-Rev. By using primer pair GVA-2193-For/GVA-3794-Rev, fragment GVA-1601 was generated and the resulting clone was named pGEM-GVA-1601.

4.2.6. Step 6: Joining of fragments GVA-1746_s and GVA-501_s by primer overlap extension (fig.4.13)

Overlapping fragments GVA-1746 and GVA-501 were joined by primer overlap extension PCR. Fragment GVA-1746_s was generated with primer pair GVA-5467-For/GVA-6979-Rev, whereas fragment GVA-501_s was generated with primer pair GVA-6979-For/GVA-Ext-Rev. By using primer pair GVA-5467-For/GVA-Ext-Rev, fragment GVA-1884 was generated and the resulting clone was named pGEM-GVA-1884. The correct orientation was determined and the clone was confirmed by PCR.

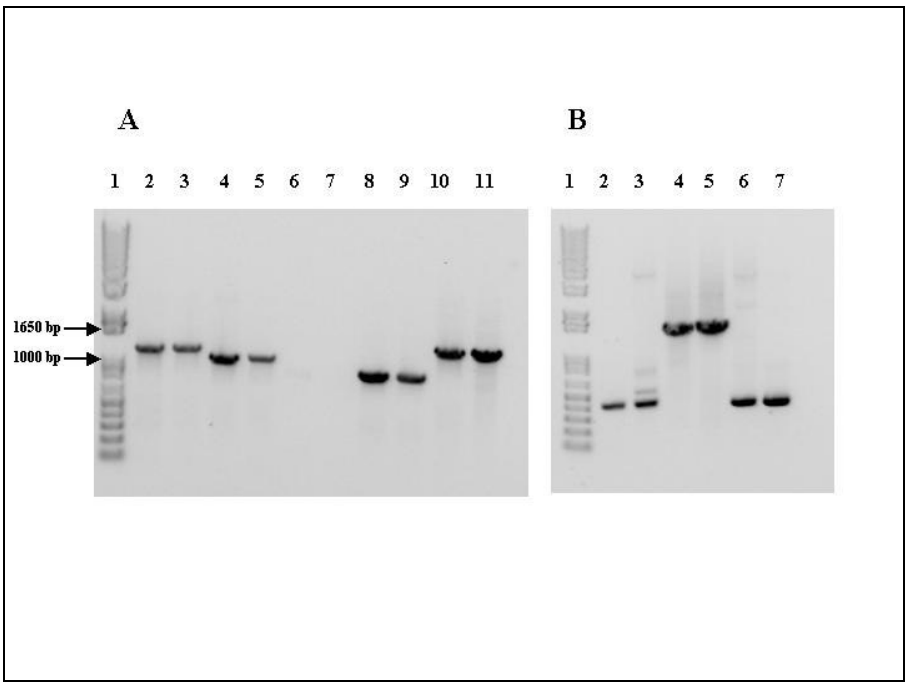


Figure 4.12. Generation of shortened fragments for primer overlap extension PCR. A) Lane 1: 1kb plus DNA marker. Lanes 2 and 3: fragment 1188 generated with primer pair GVA-Ext-For/GVA-1188-Rev. Lanes 4 and 5: fragment 1025 generated with primer pair GVA-1188-For/GVA-2193-Rev. Lanes 6 and 7: PCR failed for fragment 800 with primer pair GVA-2193-For/GVA-2972-Rev, but amplification was obtained in a follow up PCR reaction (results not shown). Lanes 8 and 9: fragment 842 generated with primer pair GVA-2972-For/GVA-3794-Rev. Lanes 10 and 11: fragment 1287 generated with primer pair GVA-3794-For/GVA-5061-Rev. B) Lane 1: 1kb plus DNA marker. Lanes 2 and 3: fragment 426 generated with primer pair GVA-5061-For/GVA-5467-Rev. Lanes 4 and 5: fragment 1532 generated with primer pair GVA-5467-For/GVA-6979-Rev. Lanes 6 and 7: fragment 392 generated with primer pair GVA-6979-For/GVA-Ext-Rev.

4.2.7. Step 7: Joining of fragments GVA-1694$_s$ and GVA-646 $_s$ by primer overlap extension (fig. 4.14)

Overlapping fragments GVA-1694 and GVA-646 were joined by primer overlap extension PCR. Fragment GVA-1694$_s$ was generated with primer pair GVA-3794-For/GVA-5061-Rev, whereas fragment GVA-646$_s$ was generated with primer pair GVA-5061-For/GVA-5467-Rev. By using primer pair GVA-3794-For/GVA-5467-Rev, fragment GVA-1673 was generated and the resulting clone was named pGEM-GVA-1673. The correct orientation was determined and the clone was confirmed by PCR.

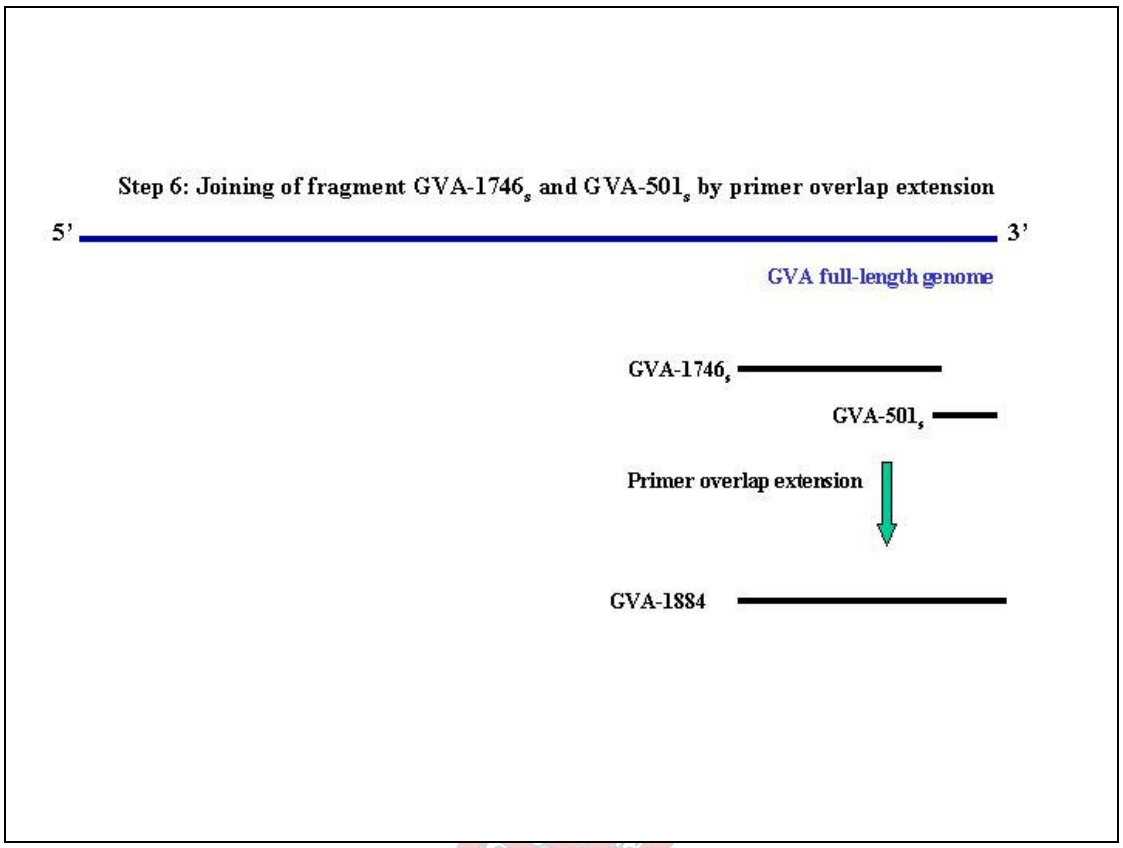


Figure 4.13. Step 6: Joining of fragments GVA-1746$_s$ and GVA-501$_s$ by primer overlap extension. Overlapping fragments GVA-1746 and GVA-501 were joined by primer overlap extension PCR. Fragment GVA-1746$_s$ was generated with primer pair GVA-5467-For/GVA-6979-Rev, whereas fragment GVA-501$_s$ was generated with primer pair GVA-6979-For/GVA-Ext-Rev. By using primer pair GVA-5467-For/GVA-Ext-Rev, fragment GVA-1884 was generated and the resulting clone was named pGEM-GVA-1884.

4.2.8. Step 8: Joining of fragments GVA-1188 and GVA-1271$_s$ by primer overlap extension (fig. 4.15)

Overlapping fragments GVA-1188 and GVA-1271 were joined by primer overlap extension PCR. Fragment GVA-1188 was generated with primer pair GVA-Ext-For/GVA-1188-Rev, whereas fragment GVA-1271$_s$ was generated with primer pair GVA-1188-For/GVA-2193-Rev. By using primer pair GVA-Ext-For/GVA-2193-Rev, fragment GVA-2193 was generated and the resulting clone was named pGEM-GVA-2193. The correct orientation was determined and the clone was confirmed by PCR.

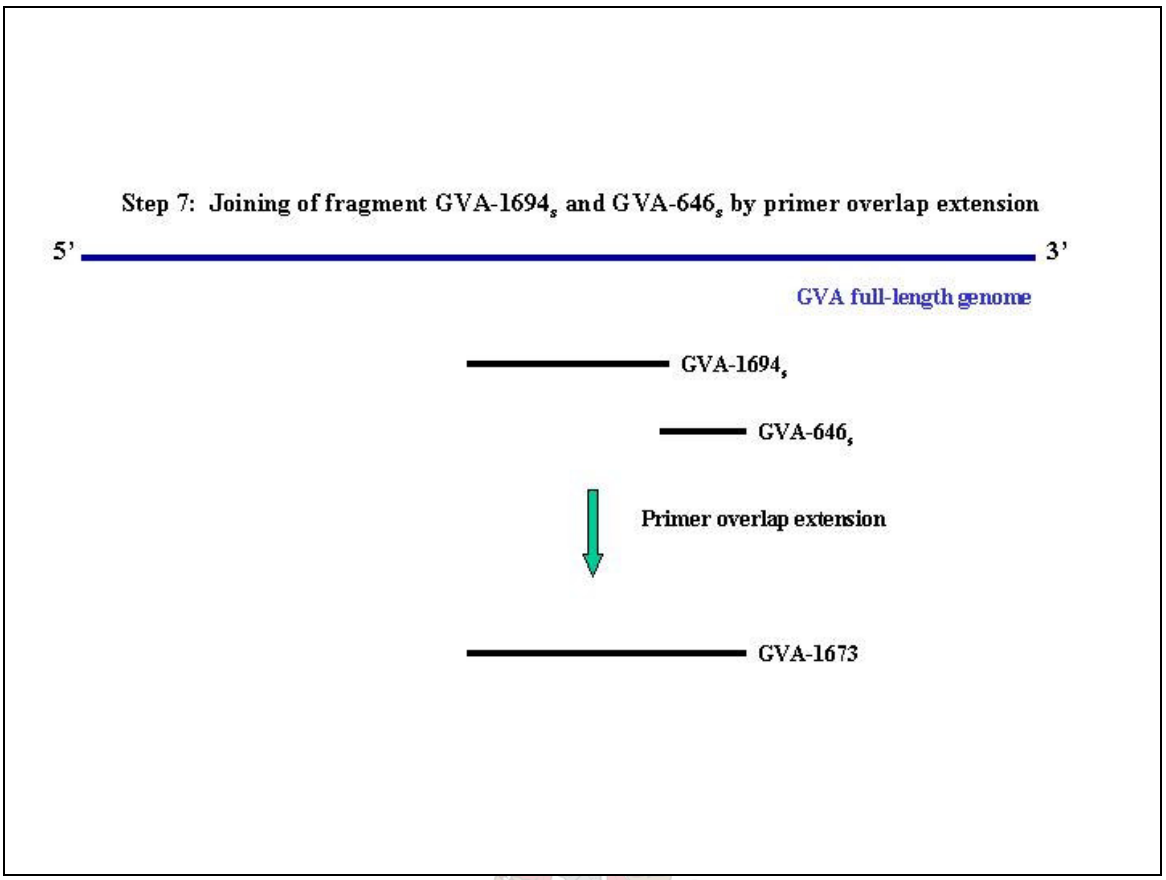


Figure 4.14. Step 7: Joining of fragments GVA-1694$_s$ and GVA-646$_s$ by primer overlap extension. Overlapping fragments GVA-1694 and GVA-646 were joined by primer overlap extension PCR. Fragment GVA-1694$_s$ was generated with primer pair GVA-3794-For/GVA-5061-Rev, whereas fragment GVA-646$_s$ was generated with primer pair GVA-5061-For/GVA-5467-Rev. By using primer pair GVA-3794-For/GVA-5467-Rev, fragment GVA-1673 was generated and the resulting clone was named pGEM-GVA-1673.

4.2.9. Step 9: Joining of fragments GVA-646$_s$ and GVA-1884 by primer overlap extension (fig. 4.17)

Overlapping fragments GVA-646 and GVA-1884 were joined by primer overlap extension PCR. Fragment GVA-646$_s$ was generated with primer pair GVA-5061-For/GVA-5467-Rev, whereas fragment GVA-1884 was generated with primer pair GVA-5467-For/GVA-Ext-Rev. By using primer pair GVA-5061-For/GVA-Ext-Rev, fragment GVA-2308 was generated and the resulting clone was named pGEM-GVA-2308. The correct orientation was determined and the clone was confirmed by PCR.

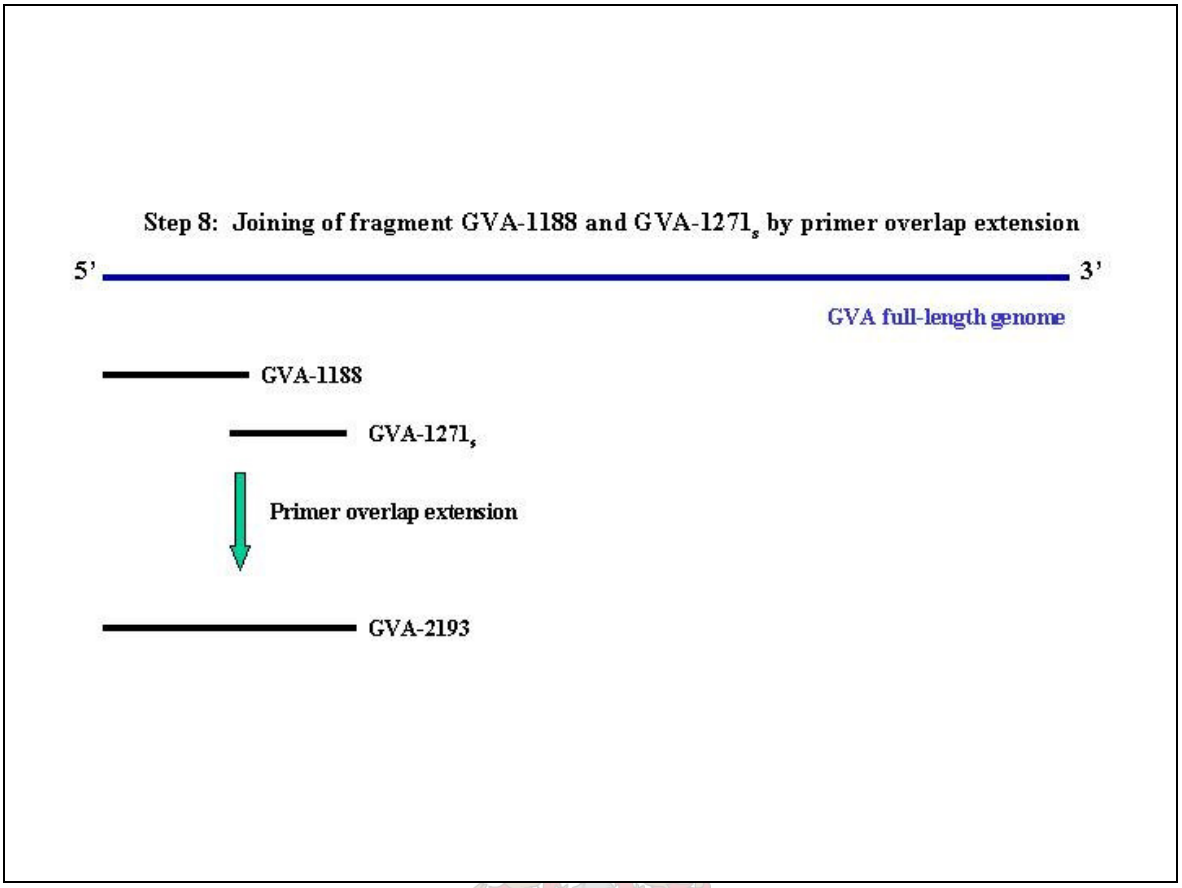


Figure 4.15. Step 8: Joining of fragments GVA-1188 and GVA-1271_s by primer overlap extension. Overlapping fragments GVA-1188 and GVA-1271 were joined by primer overlap extension PCR. Fragment GVA-1188 was generated with primer pair GVA-Ext-For/GVA-1188-Rev, whereas fragment GVA-1271_s was generated with primer pair GVA-1188-For/GVA-2193-Rev. By using primer pair GVA-Ext-For/GVA-2193-Rev, fragment GVA-2193 was generated and the resulting clone was named pGEM-GVA-2193.

4.2.10. Step 10: Generation of fragments GVA-1558 and GVA-1418 by primer overlap extension (fig. 4.19)

At this point 4 overlapping GVA cDNA fragments spanning the whole GVA genome had been obtained (fig. 4.18). Fragments proved to be too large in size to be joined by primer overlap extension. The remaining fragments thus had to be joined by digestion and ligation. With the existing overlaps, it was difficult to identify unique restriction sites that were not present anywhere else within the two fragments to be joined. To overcome this problem, the overlapping regions had to be enlarged. This was performed, by generating two fragments with primer extension PCR. The fragments were named GVA-1558 and GVA-1418.

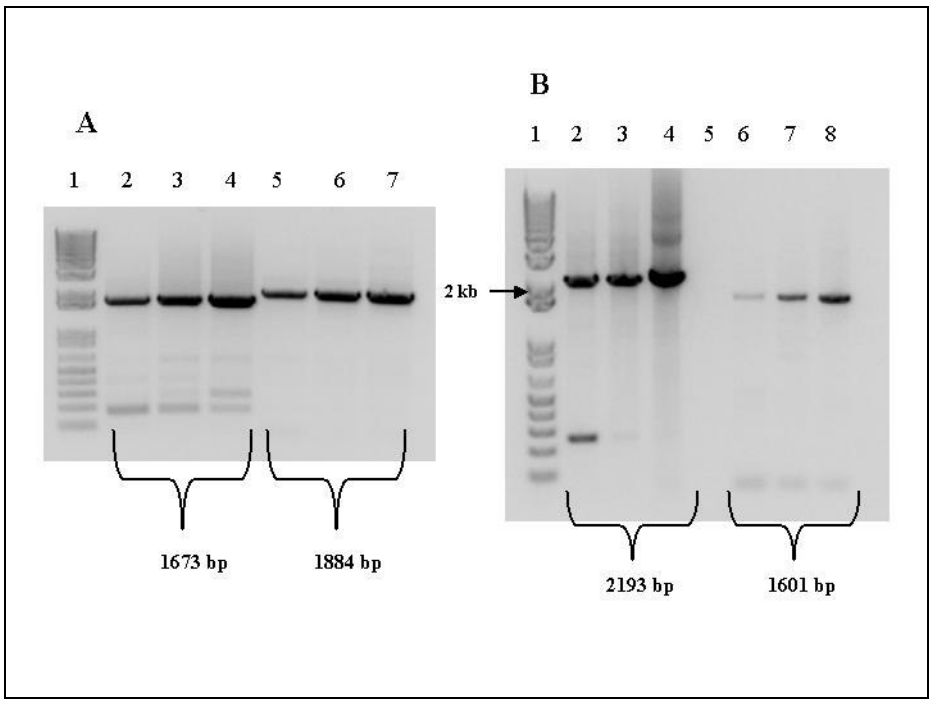


Figure 4.16. Steps 5 to 8: Fragments generated by primer overlap extension PCR. A) Lane 1: 1kb plus DNA marker. Lanes 2 to 4: fragment GVA-1673 generated in step 7. Lanes 5 to 7: fragment GVA-1884 generated in step 6. B) Lane 1: 1kb plus DNA marker. Lanes 2 to 4: fragment GVA-2193 generated in step 8. Lanes 6 to 8: fragment GVA-1601 generated in step 5.

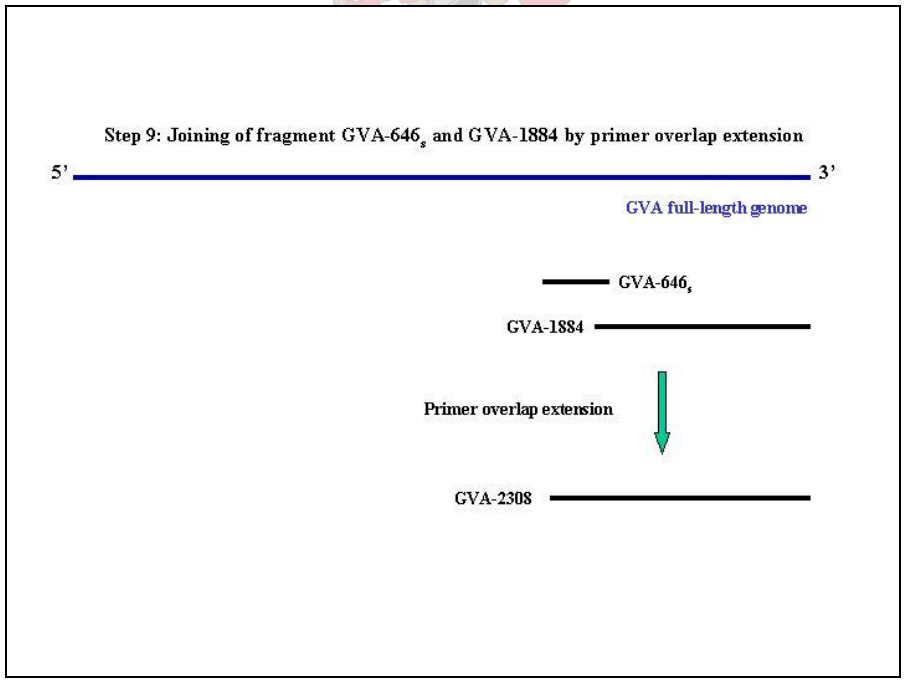


Figure 4.17. Step 9: Joining of fragments GVA-646_s and GVA-1884 by primer overlap extension. Overlapping fragments GVA-646 and GVA-1884 were joined by primer overlap extension PCR. Fragment GVA-646_s was generated with primer pair GVA-5061-For/GVA-5467-Rev, whereas fragment GVA-1884 was generated with primer pair GVA-5467-For/GVA-Ext-Rev. By using primer pair GVA-5061-For/GVA-Ext-Rev, fragment GVA-2308 was generated and the resulting clone was named pGEM-GVA-2308.

To generate fragment GVA-1558, overlapping fragments GVA-759 and GVA-965 were joined by primer overlap extension PCR. Fragment GVA-759 was generated with primer pair GVA-1434-For/GVA-2193-Rev, whereas fragment GVA-965_s was generated with primer pair GVA-2193-For/GVA-2972-Rev. By using primer pair GVA-1434-For/GVA-2972-Rev, fragment GVA-1558 was generated and the resulting clone was named pGEM-GVA-1558. The correct orientation was determined and the clone was confirmed by PCR

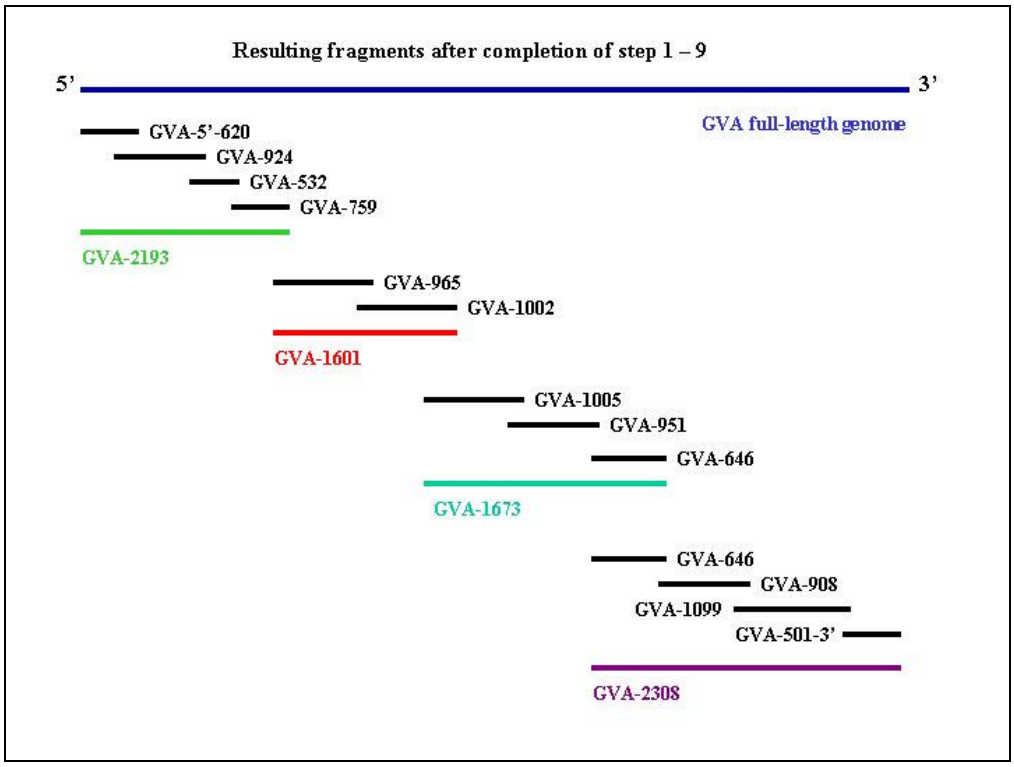


Figure 4.18. Resulting fragments after completion of steps 1 to 9. The resulting fragments are shown in colour, and the fragments from which they were generated are shown in black above the respective coloured fragments.

To generate fragment GVA-1418, overlapping fragments GVA-1002 and GVA-1694 were joined by primer overlap extension PCR. Fragment GVA-1002_s was generated with primer pair GVA-2972-For/GVA-3794-Rev, whereas fragment GVA-1694_s was generated with primer pair GVA-3794-For/GVA-4370-Rev. By using primer pair GVA-2972-For/GVA-4370-Rev, fragment GVA-1418 was generated and the resulting clone was named pGEM-GVA-1418. The correct orientation was determined and the clone was confirmed by PCR.

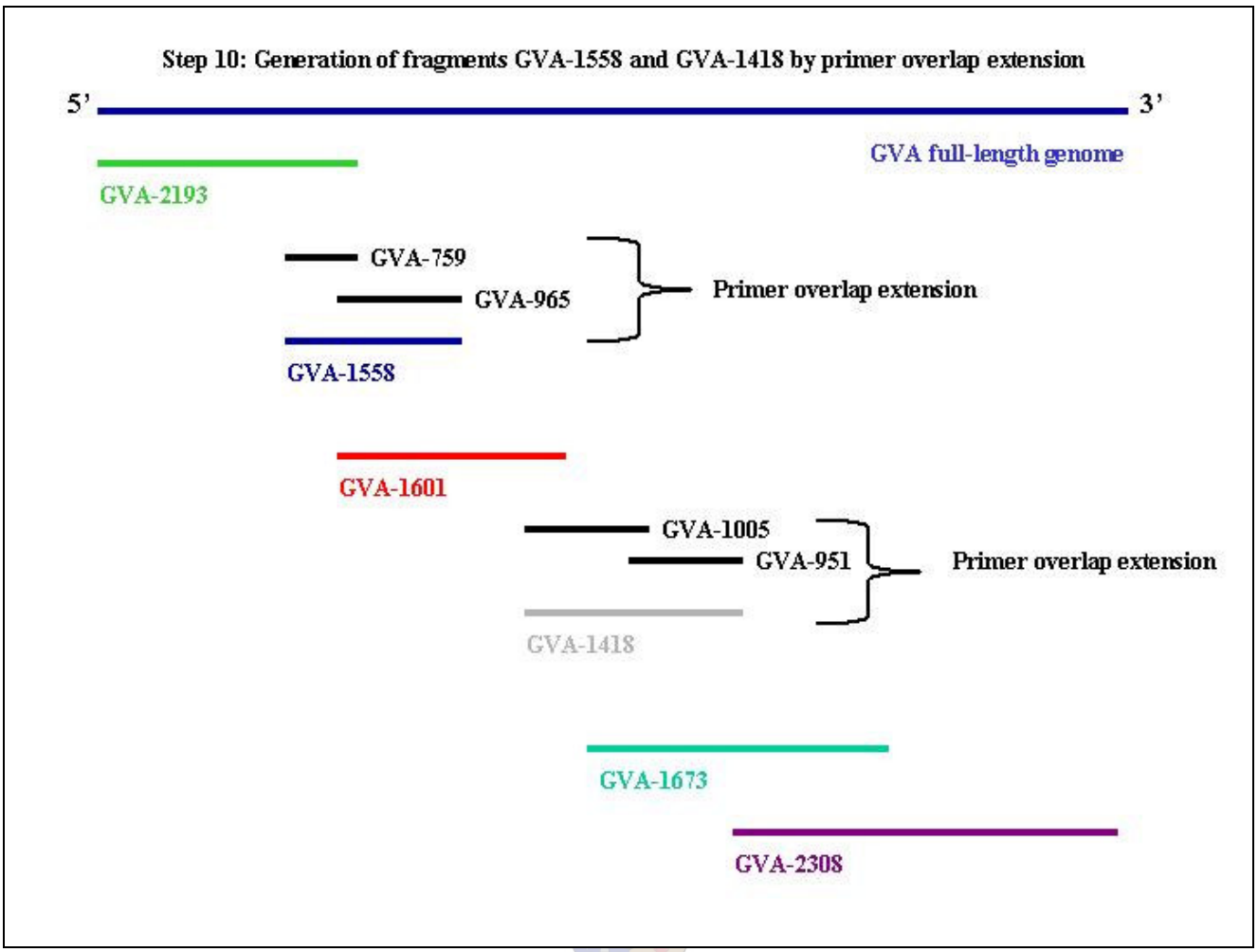


Figure 4.19. Step 10: Generation of fragments GVA-1558 and GVA-1418 by primer overlap extension. To generate fragment GVA-1558, overlapping fragments GVA-759 and GVA-965 were joined by primer overlap extension PCR. Fragment GVA-759 was generated with primer pair GVA-1434-For/GVA-2193-Rev, whereas fragment GVA-965_s was generated with primer pair GVA-2193-For/GVA-2972-Rev. By using primer pair GVA-1434-For/GVA-2972-Rev, fragment GVA-1558 was generated and the resulting clone was named pGEM-GVA-1558. To generate fragment GVA-1418, overlapping fragments GVA-1002 and GVA-1694 were joined by primer overlap extension PCR. Fragment GVA-1002_s was generated with primer pair GVA-2972-For/GVA-3794-Rev, whereas fragment GVA-1694_s was generated with primer pair GVA-3794-For/GVA-4370-Rev. By using primer pair GVA-2972-For/GVA-4370-Rev, fragment GVA-1418 was generated and the resulting clone was named pGEM-GVA-1418.

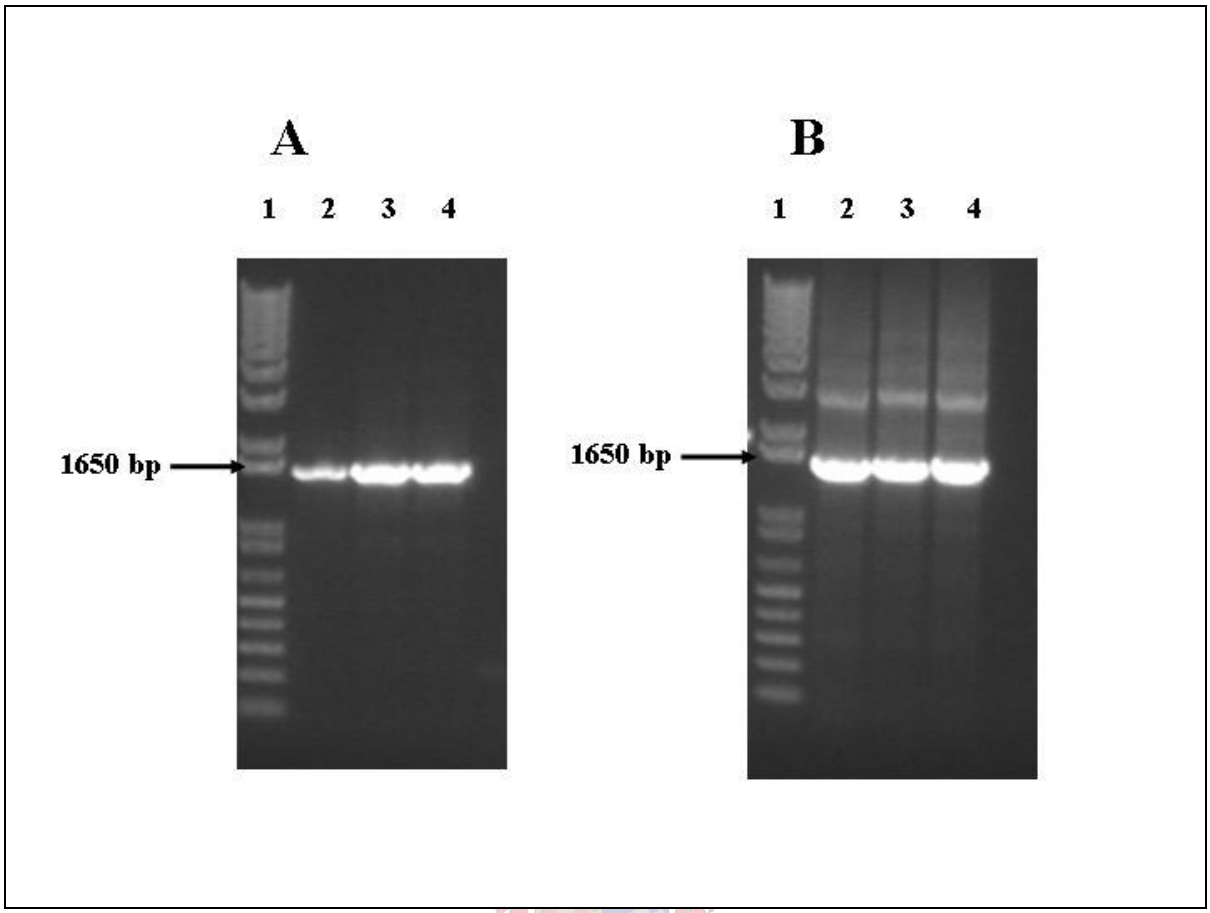


Figure 4.20. Generation of fragments GVA-1558 and GVA-1418 by primer overlap extension PCR. A) Fragment GVA-1558. Lane 1: 1kb plus DNA marker. Lanes 2 to 4: Fragment GVA-1558. B) Fragment GVA-1418. Lane 1: 1kb plus DNA marker. Lanes 2 to 4: Fragment GVA-1418.

4.2.11. Step 11: Joining of fragments GVA-2193 and GVA-1558 by restriction digestion (fig. 4.23)

At this point, six overlapping GVA cDNA fragments spanning the entire genome (except the poly-A tail), with overlaps large enough to identify unique restriction sites for digestion and construction of the clone had been generated. The six GVA cDNA fragments (figure 4.19) were confirmed with PCR (figure 4.21 and table 4.4). The overlaps between adjacent fragments within the entire clone were analyzed, unique restriction sites were identified and a new strategy for full-length GVA clone assembly was designed (fig. 4.22).

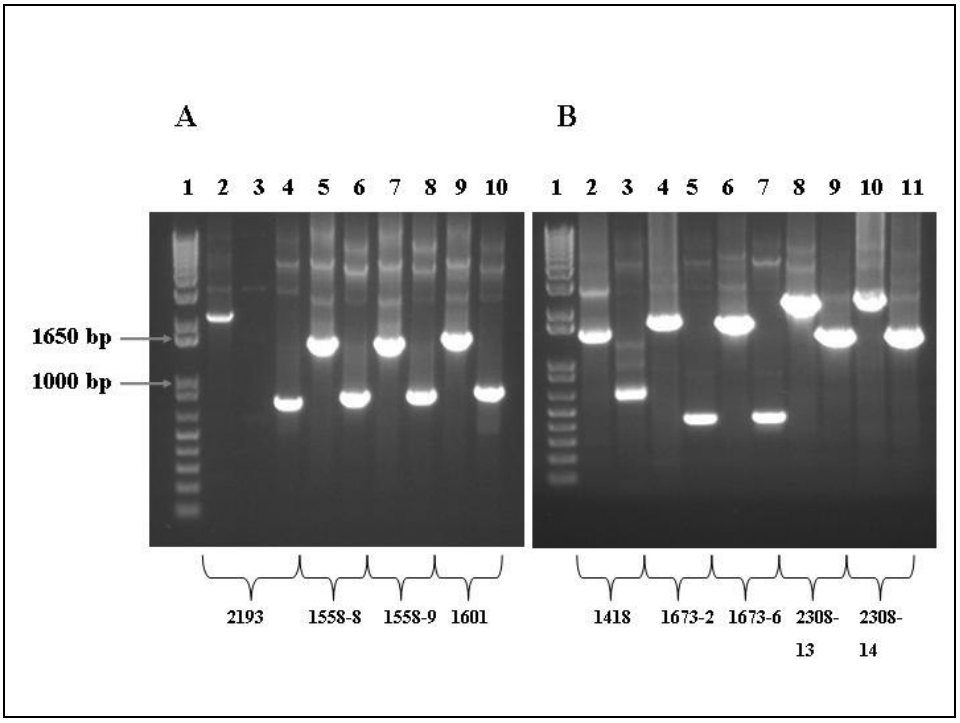


Figure 4.21. Confirmation of GVA clones by PCR. A) Lane 1: 1kb plus DNA marker. Lanes 2 to 4: clone pGEM-GVA-2193. Lanes 5 and 6: clone pGEM-GVA-1558-8. Lanes 7 and 8: clone pGEM-GVA-1558-9. Lanes 9 and 10: clone pGEM-GVA-1601. B) A) Lane 1: 1kb plus DNA marker. Lanes 2 and 3: clone pGEM-GVA-1418. Lanes 4 and 5: clone pGEM-GVA-1673-2. Lanes 6 and 7: clone pGEM-GVA-1673-6. Lanes 8 and 9: clone pGEM-GVA-2308-13. Lanes 10 and 11: clone pGEM-GVA-2308-14.

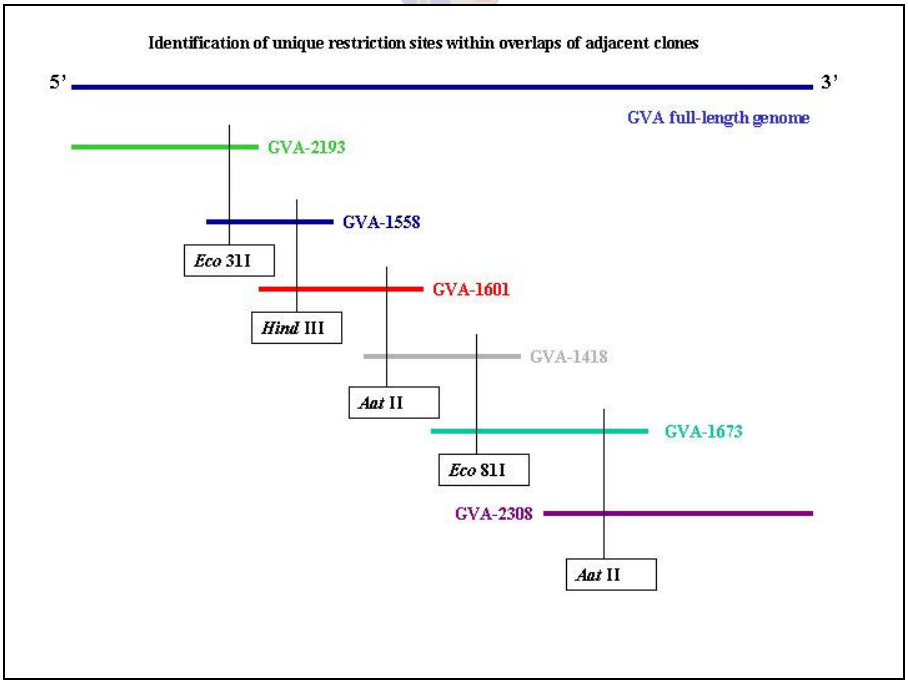


Figure 4.22. Graphic representation of the identified unique restriction sites (for final construct assembly) within overlaps of six adjacent clones.

A unique *Eco* 31I restriction site was identified within the overlap between fragment GVA-2193 and fragment GVA-1558. This enzyme is sensitive to dam methylation, so both fragments had to be transformed into a dam negative host strain. We used *E.coli* GM41 in this case. Clone pGEM-GVA-2193-GM41 was digested with *Not* I and *Eco* 31I, generating a fragment of approximately 1800 bp in size. Clone pGEM-GVA-1558-GM41 was digested with *Eco* 31I and *Sal* I generating a fragment of approximately 1250 bp. pGEM®-T Easy Vector was digested with *Not* I and *Sal* I. The two generated fragments were ligated into the digested vector in a three-way ligation. The resulting clone was named pGEM-GVA-2193/1558. The clone was confirmed by *Not* I and *Eco* 31I digestions.

Table 4.4. PCR confirmation of six GVA clones generated during steps 1 - 10 (fig. 4.21)

PCR confirmation of six GVA clones generated during steps 1 - 10			
Template	**Gel lane (fig. 4.21)**	**Primers used**	**Expected fragment size (bp)**
pGEM-GVA-2193	A2	GVA-Ext-For/GVA-2193-Rev	2193
	A3	GVA-Ext-For/GVA-620-Rev	620
	A4	GVA-1434-For/GVA-2193-Rev	759
pGEM-GVA- 1558-8	A5	GVA-1434-For/GVA-2972-Rev	1538
	A6	GVA-2193-FN/GVA-2972-Rev	779
pGEM-GVA-1558-9	A7	GVA-1434-For/GVA-2972-Rev	1538
	A8	GVA-2193-FN/GVA-2972-Rev	779
pGEM-GVA- 1601	A9	GVA-2193-For/GVA-Int-Rev	1601
	A10	GVA-2972-FN/GVA-Int-Rev	822
pGEM-GVA-1418	B2	GVA-2972-FN/GVA-4370-Rev	1398
	B3	GVA-3794-FN/GVA-4370-Rev	576
pGEM-GVA-1673-2	B4	GVA-3794-FN/GVA-5467-Rev	1673
	B5	GVA-5061-FN/GVA-5467-Rev	406
pGEM-GVA-1673-6	B6	GVA-3794-FN/GVA-5467-Rev	1673
	B7	GVA-5061-FN/GVA-5467-Rev	406
pGEM-GVA-2308-13	B8	GVA-5061-FN/GVA-Ext-Rev	2290
	B9	GVA-5880-For/GVA-Ext-Rev	1471
pGEM-GVA-2308-14	B10	GVA-5061-FN/GVA-Ext-Rev	2290
	B11	GVA-5880-For/GVA-Ext-Rev	1471

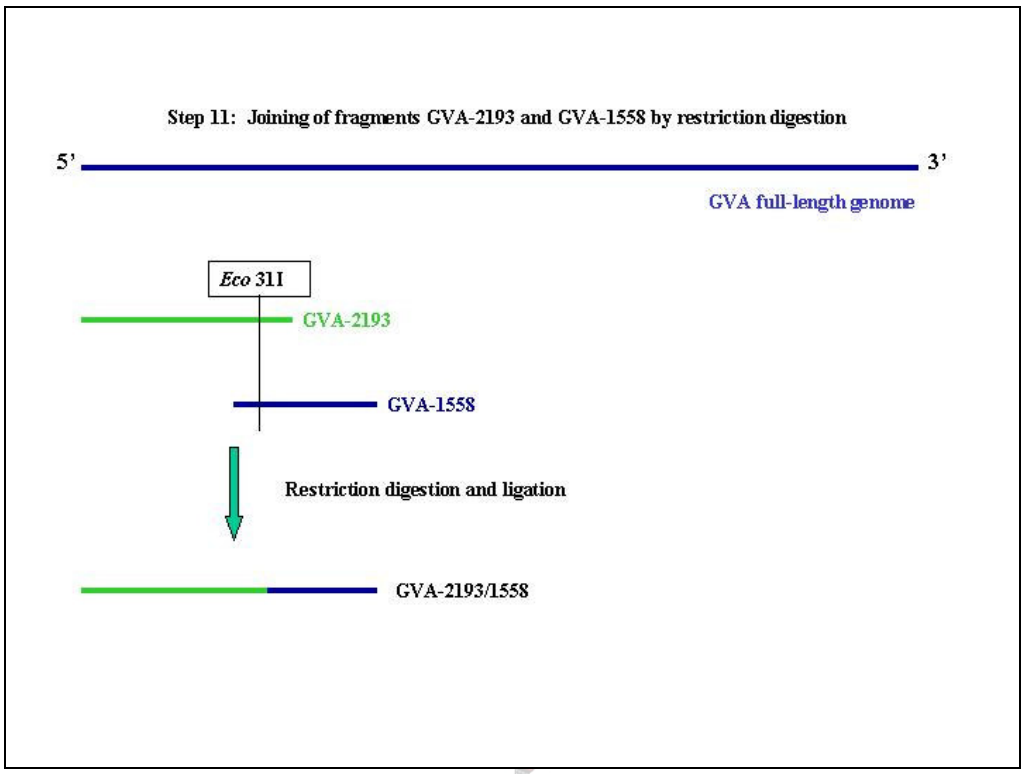


Figure 4.23. Step 11: Joining of fragments GVA-2193 and GVA-1558 by restriction digestion. Overlapping fragments GVA-2193 and GVA-1558 were joined by restriction digestion, followed by ligation, after a unique *Eco* 31I site was identified within the overlap. The resulting clone was named pGEM-GVA-2193/1558.

4.2.12. Step 12: Joining of fragments GVA-1418 and GVA-1673 by restriction digestion (fig. 4.24)

A unique *Eco* 81I site was identified within the overlap between fragment GVA-1418 and fragment GVA-1673. Clone pGEM-GVA-1418 was digested with *Not* I and *Eco* 81I, generating a fragment of approximately 1050 bp in size. Clone pGEM-GVA-1673 was digested with *Eco* 81I and *Sal* I generating a fragment of approximately 1500 bp. pGEM®-T Easy Vector was digested with *Not* I and *Sal* I. The two generated fragments were ligated into the digested vector in a three-way ligation. The resulting clone was named pGEM-GVA-1418/1673. The clone was confirmed by *Aat* II digestion.

4.2.13. Step 13: Joining of fragments GVA-2193/1558 and GVA-1601 by restriction digestion (fig. 4.25)

A unique *Hind* III site was identified within the overlap between fragment GVA-2193/1558 and fragment GVA-1601. In the previous ligation steps, a three-way ligation was used to assemble two adjacent clones. It was realized that by digesting one clone (host) with the unique identified enzyme

(within the overlap) and the vector specific enzyme (at the 3'-end of the insert) (producing vector, discarding insert), and by digesting the second clone with the same enzymes (producing the insert,

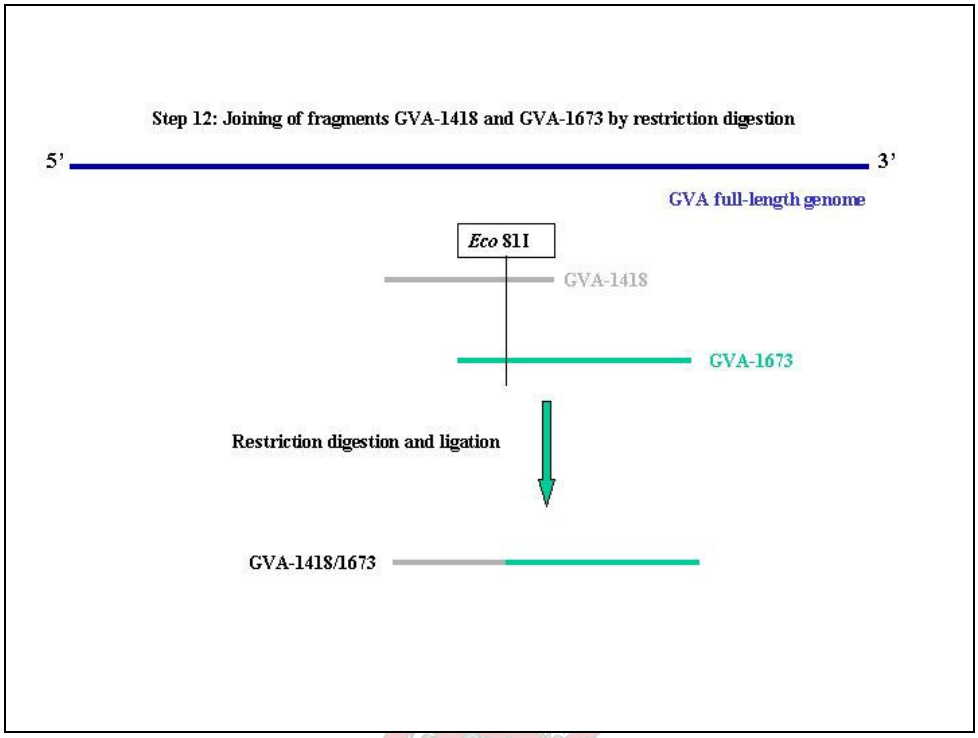


Figure 4.24. Step 12: Joining of fragments GVA-1418 and GVA-1673 by restriction digestion. Overlapping fragments GVA-1418 and GVA-1673 were joined by restriction digestion, followed by ligation, after a unique *Eco* 81I site was identified within the overlap. The resulting clone was named pGEM-GVA-1418/1673.

discarding the vector), a two-way ligation could be accomplished. This was performed by ligating the produced insert into the produced vector, thus significantly simplifying the ligation step. Clone pGEM-GVA-2193/1558 was digested with *Sal* I and *Hind* III, generating a 5.8 kb vector and a fragment of approximately 300 bp in size. Clone pGEM-GVA-1601 was digested with *Hind* III and *Sal* I generating a fragment of approximately 1000 bp. The generated 1000 bp fragment was ligated into the digested 5.8 kb vector in a two-way ligation. The resulting clone was named pGEM-GVA-3794. The clone was confirmed by *Not* I/*Hind* III and *Hind* III/*Sal* I digestions.

4.2.14. Step 14: Incorporation of the 3'-poly-A tail (fig. 4.26)

The poly-A tail was incorporated at the 3'-end of the genome. This was performed with normal PCR using primer pair GVA-5061-For / GVA-Ext-Rev-poly-A (table 3.1) and clone pGEM-GVA-2308 as template. The resulting clone was named pGEM-GVA-2308-polyA. The clone was sequenced to confirm the incorporation of the 40 A's at the 3'-end (fig. 4.27).

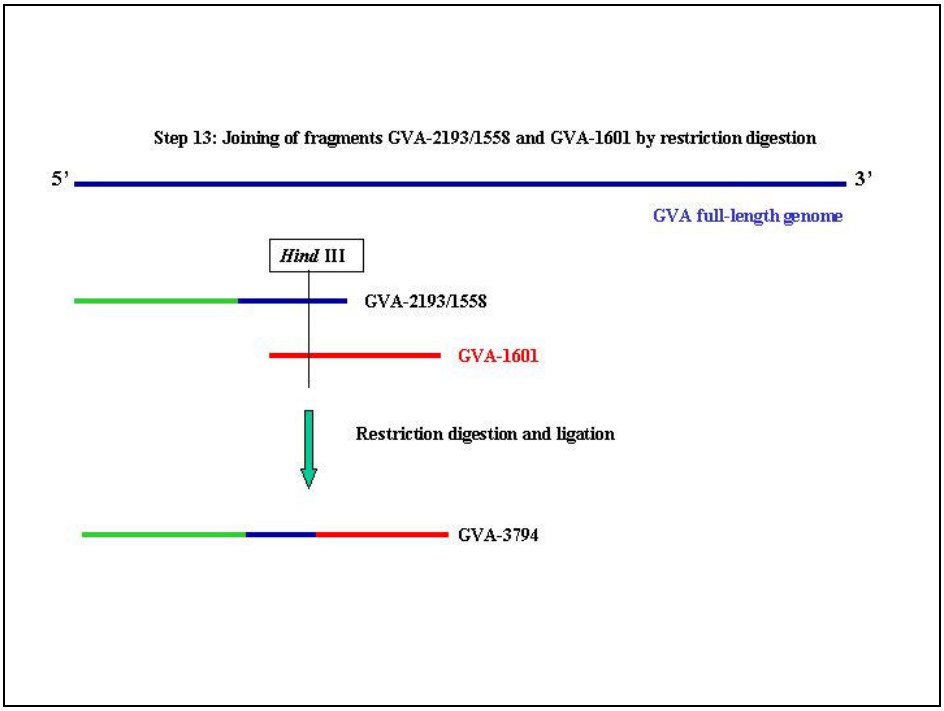


Figure 4.25. Step 13: Joining of fragments GVA-2193/1558 and GVA-1601 by restriction digestion. Overlapping fragments GVA-2193/1558 and GVA-1601 were joined by restriction digestion, followed by ligation, after a unique *Hind* III site was identified within the overlap. The resulting clone was named pGEM-GVA-3794.

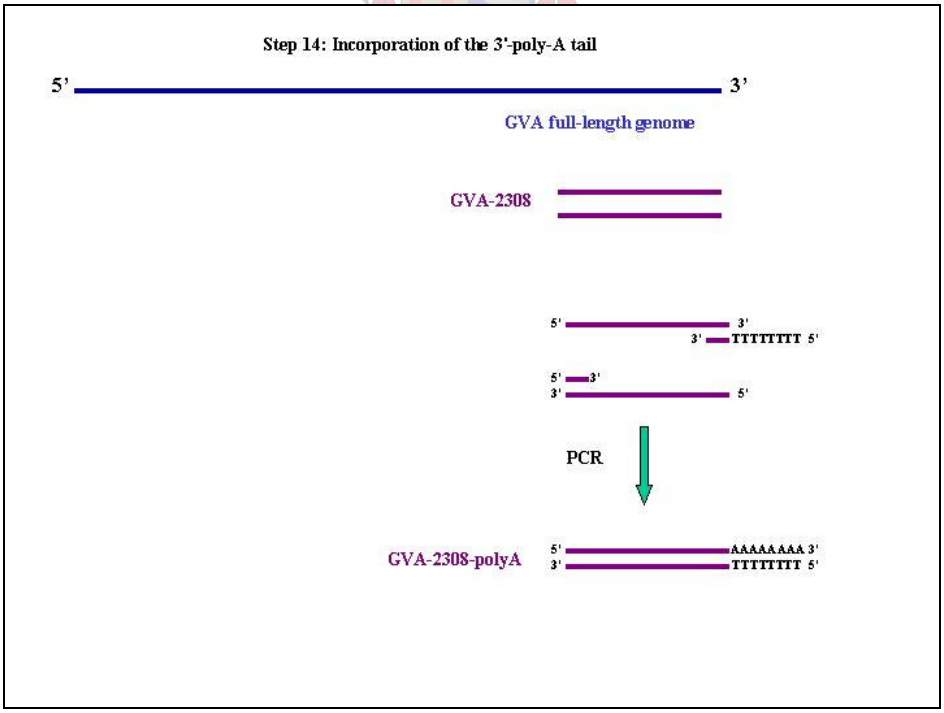


Figure 4.26. Step 14: Incorporation of the 3'-poly-A tail. The poly-A tail was incorporated with normal PCR using primer pair GVA-5061-For / GVA-Ext-Rev-poly-A and clone pGEM-GVA-2308 as template. The resulting clone was named pGEM-GVA-2308-polyA.

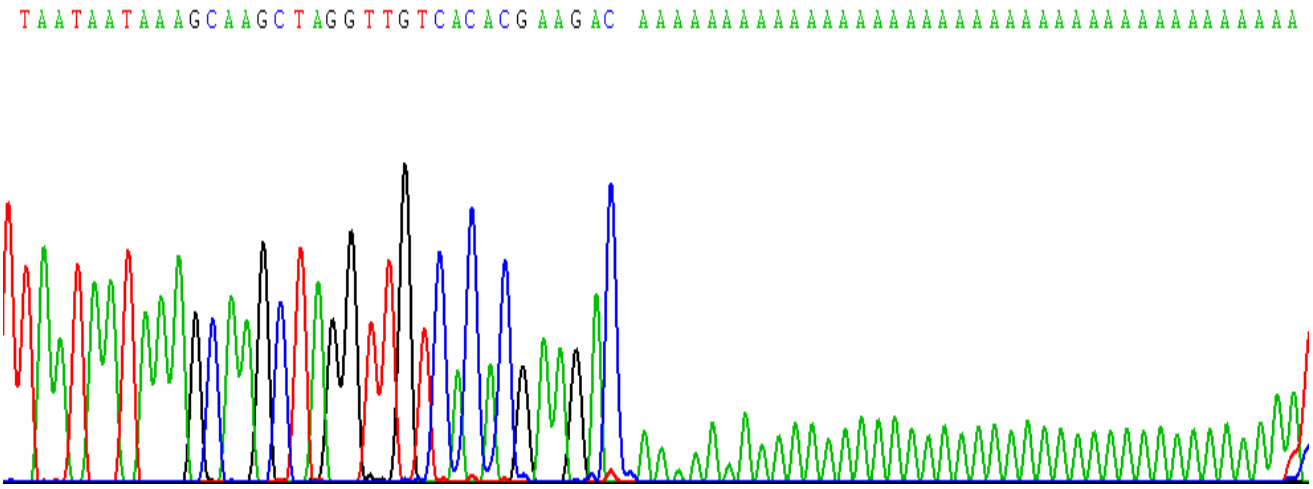


Figure 4.27. Sequence chromatogram showing the incorporation of the poly-A tail (40 A's) at the 3'-end of the viral genome. Clone pGEM-GVA-2308-polyA was sequenced to confirm the incorporation of the 40 A's at the 3'-end.

4.2.15. Step 15: Incorporation of the 5'-T7 promoter

We attempted to incorporate a T7 promoter at the 5'-end of the GVA genome using clone pGEM-GVA-3794. This was performed with the Geiser technique (Geiser *et al.,* 2001) (figure 3.3). The resulting clone was named pGEM-GVA-T7-3794. The incorporation of the T7 promoter was confirmed by PCR. Subsequent sequencing results revealed that the T7 promoter was not incorporated and that there was a 17 bp deletion at the 5'-end of the viral genome. We obtained a false positive PCR result even though the negative control was negative. This was only realized after step 16 to 18 had been completed. The strategy that we followed to rectify this problem, is presented in section 4.3.

4.2.16. Step 16: Cloning of fragment GVA-T7-3794 into pBluescript®II SK (+)

Fragment GVA-T7-3794 was cloned into pBluescript®II SK (+) for further manipulations, because of the absence of an *Aat* II site in this vector and the fact that the T7 promoter was situated at the 3'-end of the multiple cloning site (MCS). Therefore pBluescript®II SK (+) was selected as the final vector for the full-length infectious clone. Clone pGEM-GVA-T7-3794 was digested with *Not* I, and the resulting 3,8 kb fragment was cloned into *Not* I linearized pBluescript®II SK (+). The resulting clone was named pSK-GVA-T7-3794. The orientation was determined and the clone confirmed by *Aat* II/*Sal* I digestion.

4.2.17. Step 17: Cloning of *Aat* II/*Sal* I digested pGEM-GVA-2308-polyA into *Aat* II/*Sal* I digested pSK-GVA-T7-3794

Two *Aat* II sites were identified within the GVA genome, cutting at position 3205 and position 5185. Therefore step 16 was performed to eliminate the *Aat* II site within the vector (pGEM®-T Easy Vector contained an *Aat* II site). Clone pGEM-GVA-2308-polyA was digested with *Aat* II and *Sal* I to generate a fragment of approximately 2.3 kb in size. This fragment was cloned into *Aat* II and *Sal* I digested pSK-GVA-T7-3794. The resulting clone was named pSK-GVA-T7-5'-3'. The orientation was determined and the clone confirmed by *Aat* II/*Sal* I digestion.

4.2.18. Step 18: Cloning of *Aat* II digested pGEM-GVA-1418/1673 into *Aat* II linearized pSK-GVA-T7-5'-3'

The final step in the construction of the full-length clone, was the linearization of pSK-GVA-T7-5'-3' with *Aat* II and the digestion of clone pGEM-GVA-1418/1673 with *Aat* II, generating a fragment of 1980 bp. This fragment was cloned into the linearized pSK-GVA-T7-5'-3'. Nine positive clones were obtained from a total of 58 colonies screened. This was confirmed by PCR and restriction digestion. Restriction digestion determined the correct orientation of three clones. Sequence results indicated that a further two clones were in the correct orientation. The resulting 5 final clones were named pSK-GVA-IC2, pSK-GVA-IC3, pSK-GVA-IC5, pSK-GVA-IC6, pSK-GVA-IC8.

After performing *in vitro* transcription (section 4.4) on all five full-length clones, no RNA was obtained. A good yield of RNA was obtained with the control template. We sequenced the clones at the 5'-end and results revealed that the T7 promoter was never incorporated in step 15 (section 4.2.15), and that all five clones had a 17 bp deletion at the 5'-end of the viral genome. This problem could have been prevented if we sequenced the clones directly after step 15. PCR results in step 15 confirmed (which turned out to be false) the incorporation of the promoter and we continued with our assembly as if the T7 had been incorporated. To circumvent any such problems that might arise in the future, we decided to confirm by sequencing after each individual step from here on, even though it was very expensive.

4.3. CORRECTION OF MUTATIONS IN FULL-LENGTH CLONES

4.3.1. Incorporation of the T7 promoter with Expand High Fidelity PCR

As was mentioned previously in section 4.2.15, steps 16 to 18 had been completed when we sequenced the final clones and realized that the T7 promoter was never incorporated and that there was a 17 bp deletion at the 5'-end of the viral genome. We decided to rectify this by incorporating the T7 promoter with Expand High Fidelity PCR using primer pair Geiser-For/GVA-2972-Rev and pGEM-GVA-2193/1558 as template. The resulting product was cloned into pGEM®-T Easy Vector and sequenced. Sequence results revealed the incorporation of the T7 promoter (fig. 4.28). This clone was named pGEM-T7-2972-3.

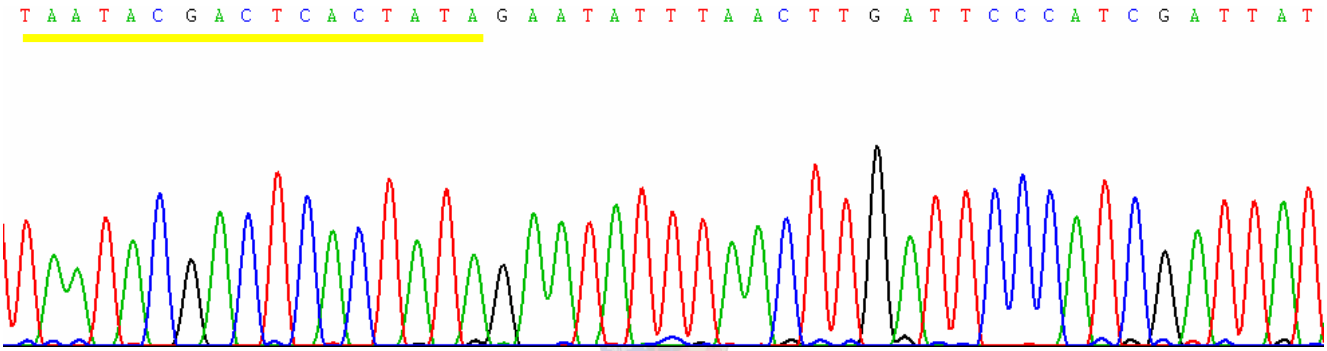


Figure 4.28. Sequence chromatogram showing the incorporation of the T7 promoter (underlined in yellow) at the 5'-end of the viral genome.

4.3.2. Cloning of *Ksp* I and *Nar* I digested T7-2972-3 into *Ksp* I and *Nar* I digested pSK-GVA-IC2, pSK-GVA-IC3, pSK-GVA-IC5, pSK-GVA-IC6, and pSK-GVA-IC8.

We had to devise a strategy to correct the five full-length clones (containing the deletion – 4.2.18) - pSK-GVA-IC2, pSK-GVA-IC3, pSK-GVA-IC5, pSK-GVA-IC6, and pSK-GVA-IC8. We identified a unique *Nar* I restriction site (at position 2612) within the new clone (pGEM-T7-2972-3 - containing the incorporated T7 promoter) and the five full-length clones. Both the pGEM®-T Easy Vector and the pBluescript®II SK (+/-) contained a *Ksp* I (*Sac* II) site upstream of the GVA genome, and this enzyme did not cut anywhere in the GVA genome. By digesting the pGEM-T7-2972-3 clone with *Nar* I and *Ksp* I (*Sac* II) a fragment of approximately 2650 bp was produced, which could be cloned into each of the *Nar* I and *Ksp* I (*Sac* II) digested full-length clones. The deletion was thus substituted with a correct

fragment, containing the T7 promoter. Unfortunately, after screening several colonies, we only obtained one clone with this method. This clone was named pSK-GVA-IC2/T7-2972-3, resulting from using pSK-GVA-IC2 as vector. When sequence results confirmed the presence of the T7 promoter in this clone, it was sequenced again with 20 primers spanning the whole genome. Full-length sequencing revealed that there were two significant mutations, among several transitions and transversions.

The first significant mutation was an insertion of an extra G in a "slippery site" of 6 G's (GGGGGG) within the genome at position 1380 to 1385 which produced a premature stop codon in ORF 1 of the genome. After sequencing of intermediate clones, we found that this mutation was incorporated during a standard PCR step during full-length clone construction. As this was a PCR mutation and could potentially render the final clone non-infectious, it had to be corrected before transcription and infection of *N. benthamiana.* A similar mutation was found by Satyanarayana *et al.*, (2003). Whilst constructing a full-length clone of *Citrus tristeza* virus (CTV), genus *Closterovirus*, family *Closteroviridae*, they incorporated a fortuitous mutation during a PCR step in a "slippery sequence" of 6 U's (UUUUUU) in ORF 1a of the viral genome. This mutation (an extra inserted uridinylate) produced a frame shift, which resulted in a premature stop codon. These authors reported that this mutated clone was less toxic to *E. coli*, and that *in vitro* produced RNA transcripts of this clone infected *N. benthamiana* protoplasts. They also reported that the mutated transcripts infected only 2-3% as many protoplasts as the wild type virus, and that, interestingly, the mutation was repaired in the viral progeny. Slippery sequences can thus be a natural survival mechanism of RNA viruses, because of their error-prone replication. When they corrected this mutation, they found that the corrected clone was much more toxic to *E. coli* (producing colonies after 3 days of incubation), but resulting *in vitro* transcripts were more infectious to protoplasts. There could be similar "slippery sites" in the GVA genome and it could be that similar reparation can occur in the GVA cDNA clone.

The second mutation was a deletion of one G at position 6959 (in ORF 4, coding for the coat protein) near the 3'-end of the virus. Once again after sequencing of intermediate clones, we found that this mutation was incorporated during an overlap extension PCR step. This deletion was immediately adjacent to the 5'-end of the GVA-6979-FN primer sequence used during a primer extension step. This mutation had to be corrected, because a mutation in the coat protein can restrict viral movement. Both these incorporated mutations could have been picked up and corrected at an earlier stage. We had several fragments and several intermediate steps during construction and it would have been ideal to sequence all intermediate clones completely, but we decided to construct the whole full-length clone

and to correct any significant mutations (*i.e.* those that could potentially produce a frameshift) at a later stage with mutagenesis or by substitution with a correct cDNA fragment. Point mutations (such as transitions and transversions) were not corrected, because as mentioned previously, we could not attribute any such mutation to normal diversity or to a PCR introduced misincorporation.

Full-length sequencing revealed that the GVA genome was 2 bases longer compared to the published sequence. Our clone consisted of 7351 base pairs. This was not due to an insertion, because alignment of our clone, with several GVA sequences in Genbank did not show the incorporation of foreign bases at any position. There was always one submitted sequence that contained a homologous region compared to our sequence. The five ORFs of our clone were positioned at: ORF 1) 87-5210, ORF 2) 5179-5712, ORF 3) 5654-6490, ORF 4) 6414-7010, and ORF 5) 7011-7283. ORFs were translated (www.expasy.ch) and aligned to submitted GVA protein sequences from Genbank (Appendix 4).

4.3.3. Correction of the mutation at position 1380 - 1385

4.3.3.1. PCR-based site-directed mutagenesis

To correct the mutation at position 1380 - 1385, PCR-based site-directed mutagenesis was performed on clone pGEM-T7-2972-3 using primer pair GVA-5'-mutagen-F/ GVA-5'-mutagen-R as described in section 3.13. Primer dimers were observed in all reactions, as expected for complementary primers. The reactions, using 8 ng and 16 ng of template, produced approximately 375 ng of new mutated vector. No product was visible in the 1 ng and 4 ng reactions (fig. 4.29). *Dpn* I digestions confirmed that the visible product was not original template DNA (fig. 4.30). Only the 8 ng and 16 ng reactions were transformed, resulting in many colonies. Four colonies were sequenced to confirm the incorporation of the mutation. All four colonies contained the incorporated mutation. Figure 4.29 shows the original 7 G's compared to the corrected 6 G's. The four corrected clones were named pGEM-GVA-mutagen-1, pGEM-GVA-mutagen-2, pGEM-GVA-mutagen-3, and pGEM-GVA-mutagen-4.

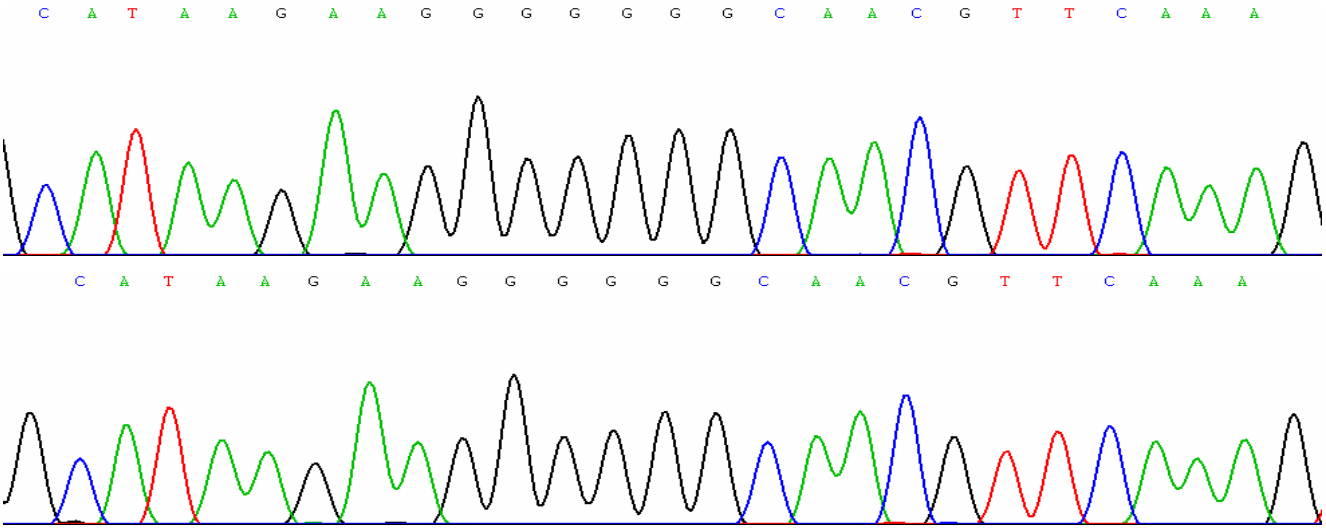


Figure 4.29. Sequence chromatogram of corrected clones (compared to mutated clones) generated with PCR-based site-directed mutagenesis. The original 7 G's (in "slippery site" which had to be corrected) are shown in the top chromatogram and the resulting corrected incorporated mutation (6 G's) in the bottom chromatogram.

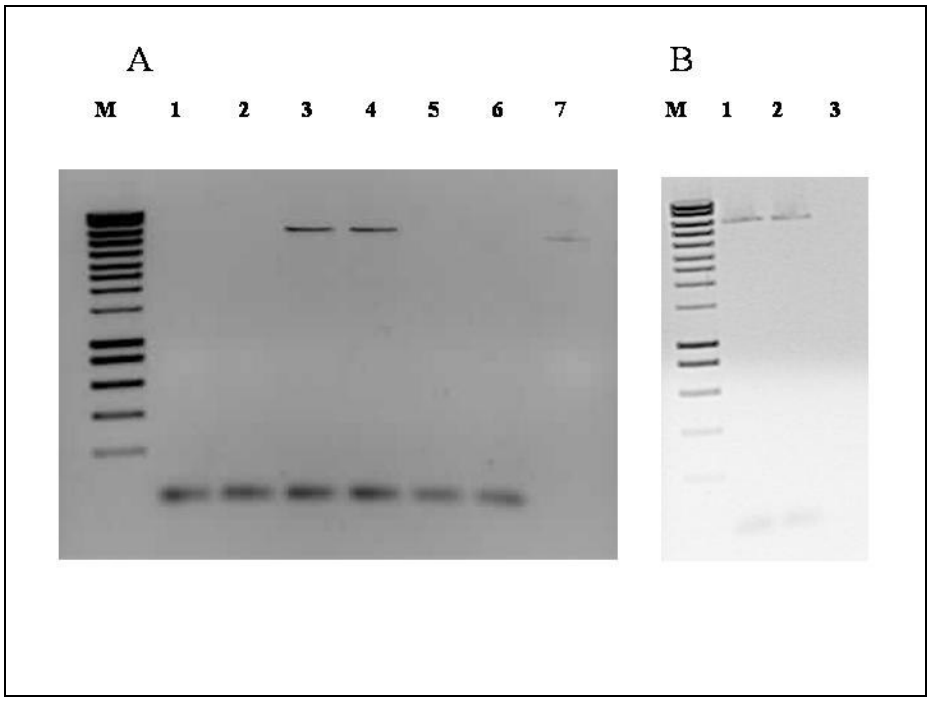


Figure 4.30. PCR-based site-directed mutagenesis. A) Expand High Fidelity PCR. Lane M: DNA Hyperladder I. Lane 1: 1 ng template reaction showing no product. Lane 2: 4 ng template reaction no product. Lane 3: 8 ng template reaction showing 7,5 ng/µl product. Lane 4: 16 ng template reaction showing 7,5 ng/µl product. Lanes 5 and 6: Mutagenesis using *Pfu* Turbo showing no product (method not described). Lane 7: Original template vector DNA. B) *Dpn* I digestion. Lane M: DNA Hyperladder I. Lane 1: digested 8 ng reaction. Lane 2: digested 16 ng reaction. Lane 3: 30 ng of original template DNA digested with *Dpn* I as positive control for digestion.

4.3.4. Cloning of *Ksp* I and *Nar* I digested pGEM-mutagen-1 and pGEM-mutagen-2 into *Ksp* I and *Nar* I digested pSK-GVA-IC2/T7-2972-3.

The same strategy described in section 4.3.2, was used in this cloning step. The only difference was that pSK-GVA-IC2/T7-2972-3 was used as vector and pGEM-GVA-mutagen-1 and pGEM-GVA-mutagen-2 as inserts. We only used pGEM-GVA-mutagen-1 and pGEM-GVA-mutagen-2 as inserts for this cloning step, because sequencing results revealed a perfect homology between the four clones (pGEM-GVA-mutagen-1, pGEM-GVA-mutagen-2, pGEM-GVA-mutagen-3, and pGEM-GVA-mutagen-4). Four positive clones were obtained and sequencing results confirmed the corrected 6 G's. The four clones were named pSK-GVA-mutagen-1, pSK-GVA-mutagen-2, pSK-GVA-mutagen-3, and pSK-GVA-mutagen-4.

4.3.5. Correction of the mutation at position 6959

This mutation was a deletion of one G at position 6959 (in ORF 4, coding for the coat protein) near the 3'-end of the virus. It produces a frameshift, altering the last 16 amino acids of the coat protein. This frameshift alters the position of the stop codon, adding 22 extra amino acids to the coat protein (Appendix 4). We attempted to correct this mutation by substituting the mutated region with a corresponding correct fragment. We found it difficult to identify unique restriction sites spanning the mutated region. Two unique sites (spanning the mutation), *Ase* I and *Bsi* WI, were identified. We planned to digest clone pGEM-GVA-501-3' with both enzymes, to produce a fragment of approximately 350 bp. This fragment would then have been cloned into *Ase* I -, *Bsi* WI-digested pSK-GVA-mutagen-1. Unfortunately, digestion results revealed the presence of more than one site for each enzyme. These results could not be explained, because the sequence, from which the sites were identified, did not reveal the presence of more than one enzyme site. We excluded the possibility of the extra digestion being attributed to star activity of the enzyme, because digestions were performed at optimal reaction conditions. These results forced us to consider another approach. We could not identify any other unique restriction site to follow the above approach, so we considered PCR-based site-directed mutagenesis. The size of the template (approximately 10 kb) was of concern, and was also the reason why we did not attempt mutagenesis as the primary approach. Unfortunately, no other option was available and we had to attempt correcting the mutation with mutagenesis. We designed two primers, GVA-3'-mutagen-F/ GVA-3'-mutagen-R (table 3.1). PCR-based site-directed mutagenesis was performed on clones pSK-GVA-mutagen-1, pSK-GVA-mutagen-2, pSK-GVA-mutagen-3, and

pSK-GVA-mutagen-4 using primer pair GVA-3'-mutagen-F/ GVA-3'-mutagen-R as described in section 3.13. We achieved the same results, as presented in section 4.3.3.1, the only difference was the fact that only a few colonies resulted per plate. Minipreps were performed on three colonies per plate. We struggled to extract a useful amount of plasmid. Sequencing of these clones revealed that the mutation was not corrected in any one of them. We speculate that the full-length clone of GVA might be toxic to *E. coli*, because every occasion when we were close to the final construct, we found it difficult to propagate the resulting clones. We always seem to have very few colonies, which result in no, or insufficient plasmid concentration after extraction. This was also reported by Saldarelli *et al.,* (2000). Despite the use of several combinations of *E. coli* strains and plasmids, they could not obtain infectious clones of GVA under the influence of a T7 promoter in *E. coli.*

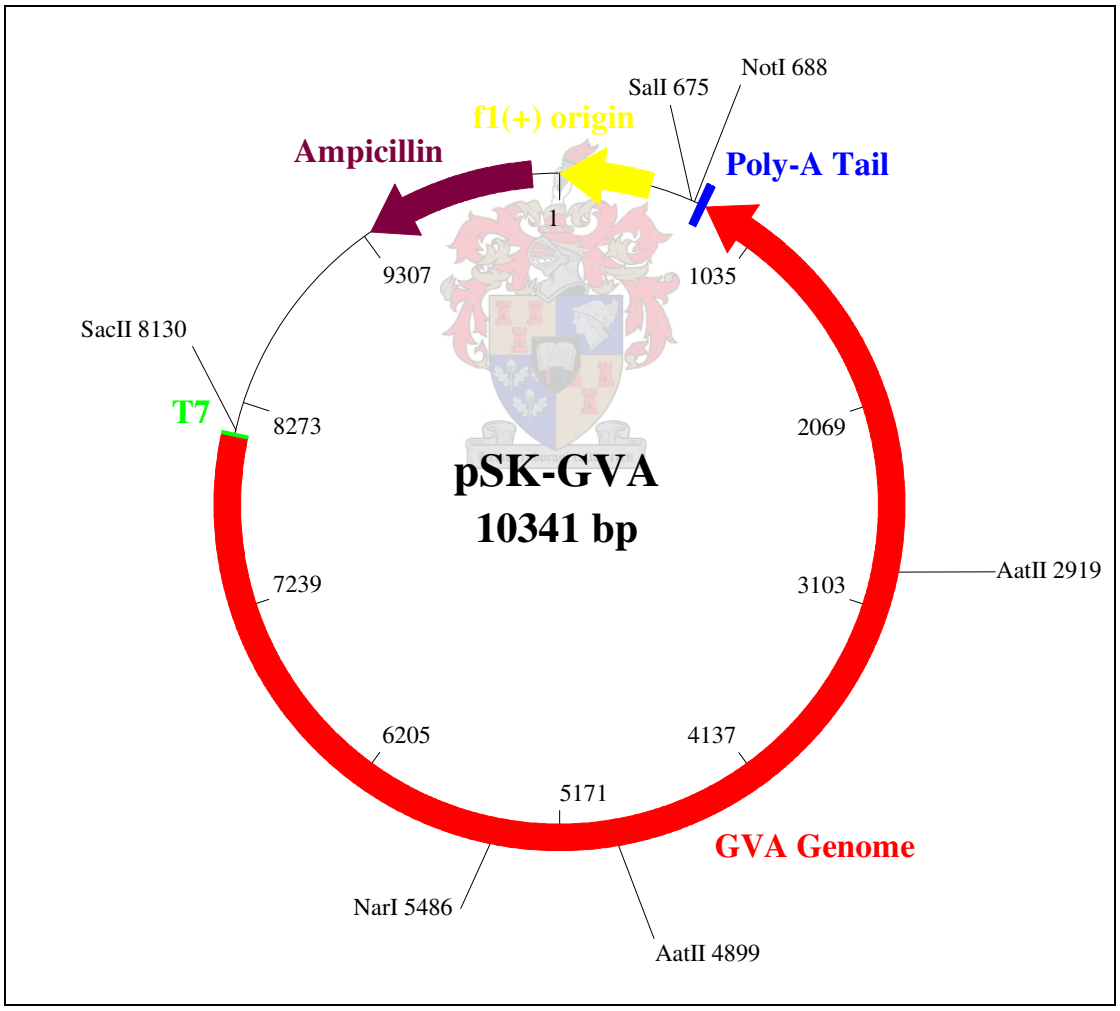


Figure 4.31. pSK-GVA, a graphical vector map of the GVA-clone in pBluescript®II SK (+). The vector map was constructed from the pSK-GVA-mutagen-3 sequence. As mentioned previously, this clone contains a single base pair deletion at position 6959 in the GVA genome resulting in a total size of 10 341 bp. The size of a mutation-free GVA clone in pBluescript®II SK (+) will be 10 342 bp.

4.4. *IN VITRO* RNA TRANSCRIPTION

Even though we could not correct the mutation in the coat protein, we decided to perform *in vitro* transcription on pSK-GVA-mutagen-3, pSK-GVA-mutagen-4 (4.3.4.), and pSK-GVA-IC2/T7-2972-3 (4.3.2.). All three clones contained the mutation at position 6959, and only pSK-GVA-IC2/T7-2972-3 contained the mutation at position 1380-1385. The size of a mutation-free GVA clone in pBluescript®II SK (+) would be 10 342 bp. pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4 sported a single base pair deletion each, changing their size to 10 341 bp (figure 4.31). The size of pSK-GVA-IC2/T7-2972-3 is 10 341 bp as well. This clone contained two mutations, a single base pair insertion, and a single base pair deletion. Approximately 60 µg *in vitro* RNA was synthesized from 1 µg linearized template DNA (of each clone) containing the T7 promoter.

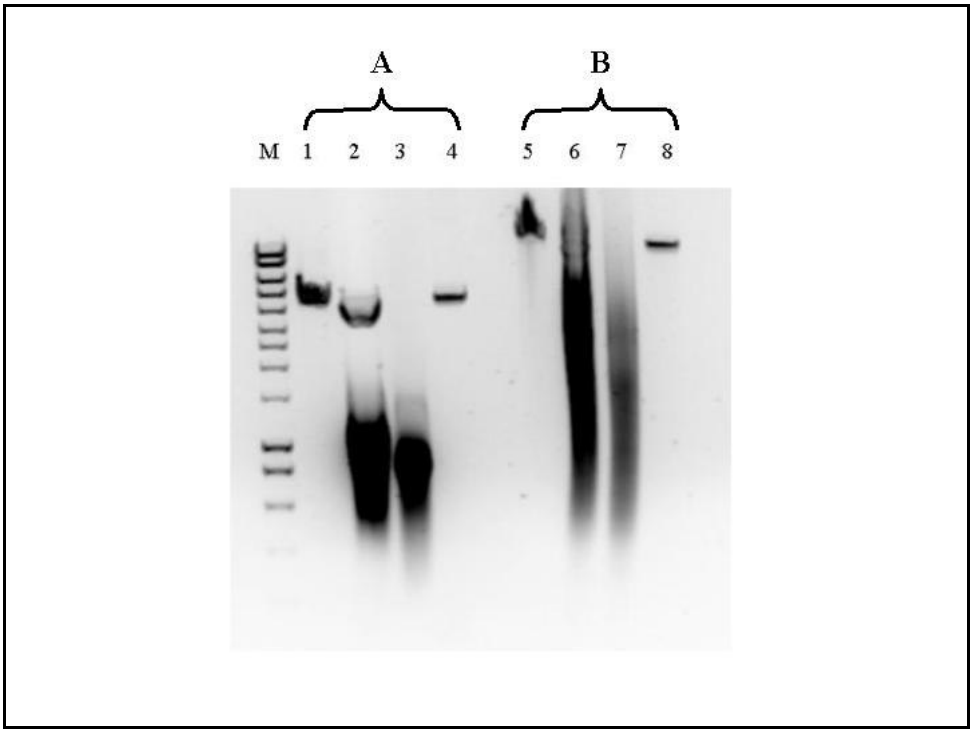


Figure 4.32. *In vitro* RNA transcription. Lane M: DNA Hyperladder I. **A**) Positive control template reaction. **B**) GVA full-length cDNA clone template. Lanes 1 and 5: Linearized template DNA. Lanes 2 and 6: 3 µl of the *in vitro* transcription reaction. Lanes 3 and 7: 3 µl transcription reaction digested with DNase I showing only *in vitro* RNA. Lanes 4 and 8: 3 µl transcription reaction digested with RNase A showing original template DNA.

4.5. INFECTION OF *N. BENTHAMIANA*

In vitro synthesized RNA, produced from pSK-GVA-mutagen-3, pSK-GVA-mutagen-4 and pSK-GVA-IC2/T7-2972-3 were used to infect *N. benthamiana* plants. Theoretically, the mutations could

render all three clones non-infectious. Previous studies had shown that a full-length clone with a mutation in a slippery site can be infectious and furthermore, the viral progeny can be devoid of this mutation (Satyanarayana *et al,* 2003). To test this, we used *in vitro* RNA, produced from clone GVA-IC2/T7-2972-3. As mentioned, this clone sported the second mutation as well. Ideally, to test the theory mentioned above, you would want a clone sporting only the mutation in the "slippery site". Unfortunately this was not possible. Studies have shown that the coat protein is involved in several stages of the viral cycle including virion assembly, cell to cell movement through the plasmodesmata, long distance movement through the phloem, and is necessary for aphid transmission and systemic infection (Ullah *et al.*, 2003). This differs from virus to virus. Galiakparov *et al.,* (2003) showed that two introduced frameshift mutations in ORF 4 (CP) of GVA, restricted viral movement. The CP seems to be essential for infectivity. The fact that we had a mutated CP meant that the chances for infectivity were slim. We thought that we might be able to show infectivity, because only the C-terminal of the CP was affected. Previous studies have shown that neither N-terminal, nor C-terminal extensions to the CP are required for virion formation and systemic infection of *Soil-borne wheat mosaic virus* (SBWMV), genus Furovirus, family unallocated, in wheat plants (Yamamiya and Shirako, 2000). For potyviruses, domains of the CP involved in various functions have been identified, and studies have shown that constructs expressing CPs without the N-terminal or the C-terminal are capable of particle formation and cell to cell spread (Ullah *et al.*, 2003). This could be the case for GVA as well. To test infectivity, we used *in vitro* RNA, produced from clones pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4 to infect *N. benthamiana* plants.

4.5.1. Symptom development in infected plants

At 21 days the infected plants were observed for the development of symptoms. No typical GVA viral symptoms were observed in the positive GVA control plant or any of the plants infected with the three respective constructed clones. A slight discoloration was observed on a few leaves of the respective infected plants.

4.5.2. RT-PCR screening of infected plants

Even though no symptoms were observed, we screened all infected plants by Rapid direct-one-tube RT-PCR as described in section 3.2.5. Preliminary RT-PCR results confirmed the presence of GVA in infected and apical leaves of infected plants (figure 4.33 & fig 4.34). The fact that no viral symptoms

were observed might be an indication that the growth room conditions were not optimal for symptom development. It also suggests that the absence of viral symptoms does not necessarily mean that the plant is virus-free. What is interesting is the fact that it seems that the constructed GVA clones were capable of viral movement and systemic infection, even though mutations were present. These preliminary results show the adaptability of viruses and their ability to survive with their error-prone replication mechanisms. These results are still inconclusive and several experiments (repeat of the above, serological detection of CP and MP and electronmicroscopy) still have to be performed to confirm these suggestions.

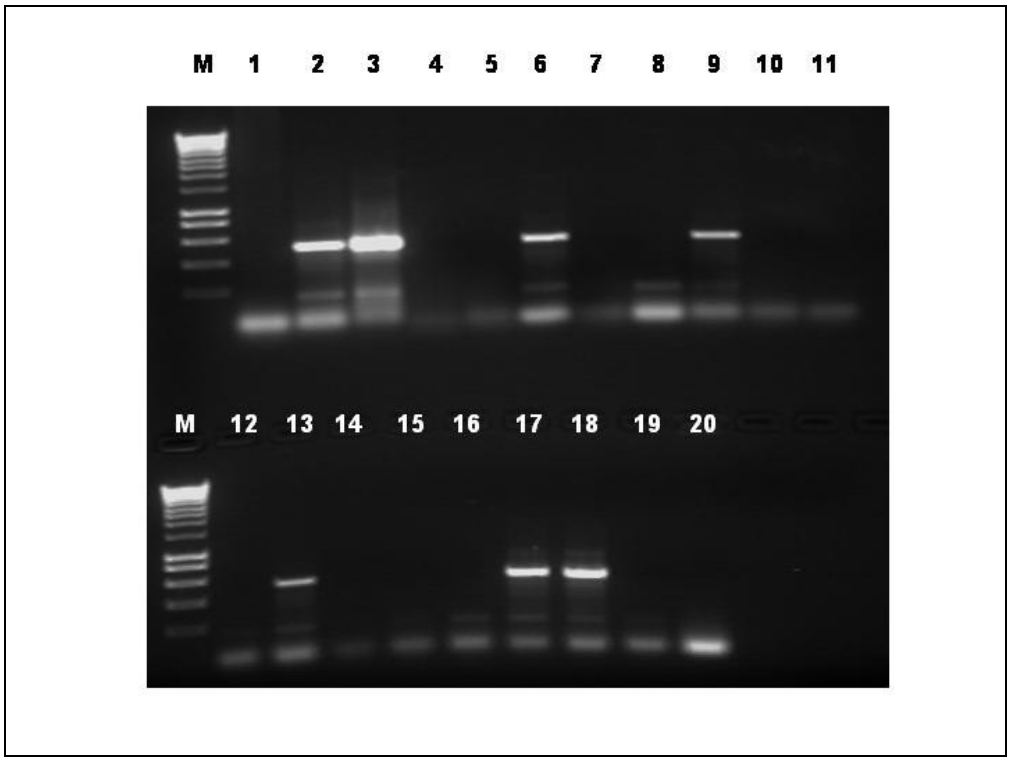


Figure 4.33. RT-PCR results of the screening of *N. benthamiana* plants for the presence of GVA. Lanes M: DNA Hyperladder I. Lanes 1 and 4: Negative H_2O controls. Lanes 2 and 3: Positive GVA control. Lanes 5 to 8: GVA infected plant (live virus) positive control. Lanes 9 to 12: plant infected with *in vitro* RNA from pSK-GVA-mutagen-3. Lanes 13 to 16: plant infected with *in vitro* RNA from GVA-IC2/T7-2972-3. Lanes 17 to 20: plant infected with *in vitro* RNA from pSK-GVA-mutagen-4.

4.6. GENERAL DISCUSSION

The classic technique of constructing an infectious clone of an RNA virus is by purifying virus particles followed by the extraction of viral RNA. A full-length infectious clone can, in theory, be obtained after a single reverse transcription reaction, followed by an optimized long PCR step. This

means only one PCR, which reduces the incorporation of mutations drastically. By utilizing enzymes with proofreading capability, the fidelity of the reverse transcription and PCR reactions are further

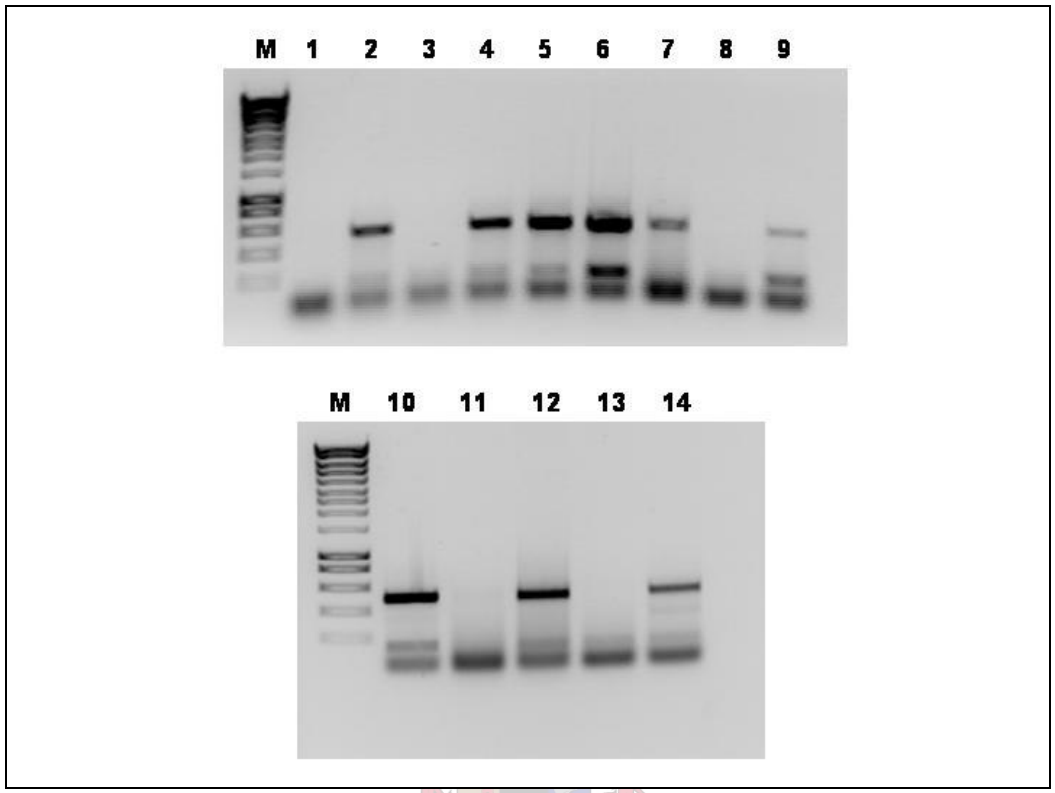


Figure 4.34. RT-PCR results of the screening of *N. benthamiana* plants for the presence of GVA. Lanes M: DNA Hyperladder I. Lane 1: Negative H_2O control. **Lanes 2 to 5)** plant infected with *in vitro* RNA from pSK-GVA-mutagen-3 (each lane is an RT-PCR reaction performed on a specific leaf). **Lanes 6 to 9)** plant infected with *in vitro* RNA from pSK-GVA-mutagen-4. Lane 6: Apical leaf. Lane 7: Infected leaf. **Lanes 10 to 14)** plant infected with *in vitro* RNA from GVA-IC2/T7-2972-3. Lane 10: Infected leaf. Lane 14: Apical leaf.

improved. An added benefit is the fact that only two primers are required. Unfortunately, we were not equipped to follow this approach. As mentioned previously, we investigated three methods to obtain GVA fragments by RT-PCR. Use of the two-step RT-PCR methods facilitated the use of high-fidelity polymerases, because the reverse transcription and PCR steps are separated. When using the one-step RT-PCR method, no high fidelity enzymes could be used, because of the fact that both the reverse transcriptase and polymerase were optimized to work in 10x *Taq* DNA polymerase buffer. This system was not compatible with a proofreading enzyme. Out of all three tested methods, this method was preferred and led to the best results. Therefore a major setback in the construction of the GVA clone was the fact that we could not use proofreading enzymes in all PCR steps, as we had planned. This could have led to the incorporation of mutations (such as transitions, transversions, deletions and

insertions), which could have been prevented. This was exacerbated by the fact that the clone had to be constructed in more than one step, leading to incorporation of more mutations in each round of PCR. These mutations could ultimately render the final clone non-infectious. Regardless, we proceeded with our approach because RNA viral polymerases are naturally error-prone and studies showed that infectious clones could be obtained by using normal *Taq* DNA polymerase (Hayes and Buck, 1990). Sequencing of overlapping GVA clones revealed the heterogeneity and quasispecies of GVA, as no clones showed a 100% sequence identity within the overlap. A sequence difference of approximately 13% was observed on average, within GVA overlaps. The fact that these differences could have been brought about by misincorporations during RT-PCR is not excluded. As mentioned previously, we had to construct the clone in different steps by joining adjacent overlapping clones with either restriction digestion or primer extension PCR. The presence of overlaps sporting a sequence difference, made it more difficult to join overlapping clones, because when a unique restriction site was identified in one overlap, it could be absent in the second overlap. When using primer extension PCR, an overlap with exact sequence similarity is needed. Initially we tried to join these overlaps with overlap extension PCR, even though the overlaps had a sequence difference. We managed to join a few clones in this way, but we had mixed results. The fact that we could not calculate an exact melting temperature for each overlap, and that each overlap between different clones differed in length made this approach problematic. This problem was circumvented by producing larger overlaps between adjacent clones (step 10, section 4.3.10). A full-lengh clone of GVA was constructed in several steps and cloned into pBluescript®II SK (+). This clone was named GVA-IC2/T7-2972-3. After full-length sequencing of the clone, two significant mutations were found. Both these mutations produced a frameshift. The first mutation was a single base pair insertion in a slippery site of 6 G's at position 1380 – 1385 in ORF 1. This mutation was corrected by PCR-based site-directed mutagenesis, which resulted in pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4. The second mutation was a single base pair deletion at position 6965 in ORF 4. This was of concern, because this ORF coded for the coat protein. Several techniques were approached and tested to correct this mutation. The first approach we followed was to identify unique restriction sites spanning the mutation, and substituting this fragment with a normal wild type fragment. Unfortunately, this approach was unsuccessful, because the identification of unique sites proved to be difficult. The second approach we followed was PCR-based site-directed mutagenesis, but again this was unsuccessful. The template DNA used in the amplification proved to be too large, approximately 10 kb. An added disadvantage was that, because of the large size, PCR errors could be incorporated. Even though the second mutation could not be corrected, we decided to perform *in vitro* transcriptions followed by subsequent infections of *N. benthamiana* plants. Two clones, (hosting the

mutation at position 6959) pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4 and one clone (hosting both mutations), GVA-IC2/T7-2972-3 were selected for transcriptions and infections. *In vitro* RNA was produced from all three clones and used to mechanically infect *N. benthamiana* plants. After 21 days no significant visual symptoms were observed in plants infected with wild type GVA or in plants infected with *in vitro* RNA. Rapid direct-one-tube RT-PCR was performed on infected plants. Prelimary results showed the presence of virus in infected leaves and apical leaves of infected plants, suggesting systemic infection and viral movement. This find was very interesting, because it suggests that all constructed clones were capable of movement, irrespective of whether only one or both mutations were present. It also suggested that the mutations in the constructed clones might somehow be corrected in the viral progeny, and that the clones are somehow capable of movement. The effect of the two mutations in the constructed clones is currently under investigation.

Chapter 5: Conclusion

New emerging diseases are presenting a problem in the Western Cape (and elsewhere) and have a negative impact on vineyards. GVA (suggested to be the second most important virus in South African vineyards) is thought to be involved in one such disease named Shiraz disease. The aim of this study was to construct a full-length infectious cDNA clone of GVA, which would provide a tool for studying GVA's involvement in the disease and to unravel the aetiology of Shiraz disease.

A full-length GVA cDNA clone named pSK-GVA-IC2/T7-2972-3 was constructed in several steps. A T7 promoter was ligated to the 5'-end of the 7 351 base pair GVA genome to facilitate *in vitro* transcription. At the 3'-end a 40 bp poly-A tail was incorporated. The cDNA clone was cloned into pBluescript®II SK (+) (Stratagene). Full-length sequencing of pSK-GVA-IC2/T7-2972-3 revealed the incorporation of two significant frameshift mutations. The first mutation was a single base pair insertion in a "slippery site" of six G's in ORF1. This mutation was corrected with PCR-based site-directed mutagenesis resulting in two clones pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4. The second mutation could not be corrected after several approaches were taken. This mutation was a single base pair deletion at position 6959 in the coat protein, which potentially could influence or even restrict viral movement and systemic infection if not corrected. Sequencing also revealed several transitions and transversions, which could not conclusively be attributed to either misincorporations or natural diversity after multiple alignments with submitted GVA sequences. All five ORFs of pSK-GVA-IC2/T7-2972-3 were translated (www.expasy.ch) and aligned to submitted GVA translation products, to investigate the consequence of the two introduced mutations. The first mutation produced a premature stopcodon in ORF1, and the second a frameshift in ORF4 (changing the position of the stopcodon) altering the last 16 amino acids of the wild type coat protein, and adding an extra 22 amino acids. ORF2, ORF3 and ORF5 were fairly similar to submitted GVA protein sequences (ORF protein alignments of pSK-GVA-mutagen-3 are given in Appendix 4).

Three resulting clones pSK-GVA-IC2/T7-2972-3 (containing both mutations), pSK-GVA-mutagen-3 (containing 6959 mutation) and pSK-GVA-mutagen-4 (containing 6959 mutation) were transcribed to produce *in vitro* RNA. *N. benthamiana* plants were mechanically infected and screened with RT-PCR for the presence of virus, at 21 days post-inoculation. Preliminary RT-PCR results showed the presence of GVA in the infected leaves and apical leaves of infected plants, even though no visible viral

symptoms were observed. This was also true and similar for plants infected with wild type GVA. This suggests that the three constructed clones are somehow capable of systemic infection, irrespective of whether one or both mutations are present. These preliminary results are still inconclusive and several post-infection experiments will have to be performed to confirm the infectious nature, and systemic infection of the clones. Follow-up infections will be performed, using infected leaves to infect new *N. benthamiana* plants to prove Koch's postulates. The effect of both mutations will be tested thoroughly, to confirm whether the viral progeny are devoid of the mutation (corrected) or whether they are capable of movement with the mutation.

Nevertheless, three GVA cDNA clones (hosting mutations), seemingly capable of systemic infection, were constructed, which could provide us with a framework for further molecular manipulations of the GVA genome. The mutations can be corrected to establish a mutation-free infectious GVA clone. Areas of the GVA genome within the constructed framework of the cDNA clones can be substituted with wild type GVA fragments, which could make the clones more infectious. This study has laid the foundation for investigations into the GVA genome and the established GVA cDNA clones can also provide us with a tool to do mutational analysis and to study genome organization and gene expression of GVA. After most or all of these puzzles have been unravelled, this knowledge could be applied to study the aetiology of Shiraz disease and GVA's involvement in the disease. This could greatly benefit the grapevine community of the Western Cape and could help us to protect that cherished, ruby coloured drink called wine.

Appendix 1:

Steps taken in construction of the full-lenth GVA cDNA clones

Steps taken in construction of the full-lenth GVA cDNA clones are presented in two tables on the following pages.

Table A1.1. Steps 1 -18 in construction of the full-length clones.

Step	Fragment name	Clone name	Initial fragments/clones	Overlapping Primer	Technique	RE site
1	GVA-1188	pGEM-GVA-1188	GVA-620/GVA-924		Digestion/Ligation	*Sph* I
2	GVA-1694	pGEM-GVA-1694	GVA-1005/GVA-951		Digestion/Ligation	*Sph* I
3	GVA-1746	pGEM-GVA-1746	GVA-908/GVA-1099		Digestion/Ligation	*Eco* RV
4	GVA-1271	pGEM-GVA-1271	GVA-532/GVA-759	GVA-1434-Rev	Primer overlap Extension	
5	GVA-1601	pGEM-GVA-1601	$GVA\text{-}965_s$/$GVA\text{-}1002_s$	GVA-2972-Rev	Primer overlap Extension	
6	GVA-1884	pGEM-GVA-1884	$GVA\text{-}1746_s$/$GVA\text{-}501_s$	GVA-6979-Rev	Primer overlap Extension	
7	GVA-1673	pGEM-GVA-1673	$GVA\text{-}1694_s$/$GVA\text{-}646_s$	GVA-5061-Rev	Primer overlap Extension	
8	GVA-2193	pGEM-GVA-2193	$GVA\text{-}1188$/$GVA\text{-}1271_s$	GVA-1188-Rev	Primer overlap Extension	
9	GVA-2308	pGEM-GVA-2308	$GVA\text{-}646_s$/GVA-1884	5 GVA-467-Rev	Primer overlap Extension	
10a	GVA-1558	pGEM-GVA-1558	GVA-759/$GVA\text{-}965_s$	GVA-2193-Rev	Primer overlap Extension	
10b	GVA-1418	pGEM-GVA-1418	$GVA\text{-}1002_s$/GVA-1694	GVA-Int-Rev	Primer overlap Extension	
11	GVA-2193/1558	pGEM-GVA-2193/1558	GVA-2193/GVA-1558		Digestion/Ligation	*Eco* 31I
12	GVA1418/1673	pGEM-GVA1418/1673	GVA-1418/GVA-1673		Digestion/Ligation	*Eco* 81I
13	GVA-3794	pGEM-GVA-3794	GVA-2193/1558//GVA-1601		Digestion/Ligation	*Hind* III
14	GVA-2308-polyA	pGEM-GVA-2308-polyA	GVA-2308	Incorporation of poly A tail by PCR with primer pair 5061F / ER-poly-A (see primer table)		
15	GVA-T7-3794	pGEM-GVA-T7-3794	GVA-3794	Incorporation of T7 promoter with Geiser Technique		
16	GVA_{pSK}-T7-3794	pSK-GVA-T7-3794	pGEM-T7-3794	Cloning of T7-3794 into pSK using Not I		
17	GVA_{pSK} -T7-5'-3'	pSK-GVA-T7-5'-3'	pSK-T7-3794 pGEM-2308-polyA	Cloning of *Aat* II/ *Sal* I digested pGEM-2308-polyA into *Aat* II/ *Sal* I digested pSK-T7-3794		
18	pSK-T7-GVA	pSK-GVA-IC	pSK-5'-3' pGEM-1418/1673	Cloning of *Aat* II digested pGEM-1418/1673 into *Aat* II linearized pSK-5'-3'		

Table A1.2. Steps 19-23 in construction of the full-length clones.

Correction of mutations in full-length clones					
Step	**Method**	**Clones used**	**Primers used**	**Enzymes used**	**Resulting clones**
19	Incorporation of the T7 promoter with Expand High Fidelity PCR	pGEM-GVA-2193/1558	Geiser-For GVA-2972-Rev		pGEM-T7-2972-3
20	Cloning of *Ksp* I and *Nar* I digested T7-2972-3 into *Ksp* I and *Nar* I digested pSK-GVA-IC2, pSK-GVA-IC3, pSK-GVA-IC5, pSK-GVA-IC6, and pSK-GVA-IC8.	**Vectors:** pSK-GVA-IC2 pSK-GVA-IC3 pSK-GVA-IC5 pSK-GVA-IC6 pSK-GVA-IC8 **Insert:** pGEM-T7-2972-3		*Nar* I *Ksp* I	pSK-GVA-IC2/T7-2972-3
21	PCR-based site-directed mutagenesis	pGEM-T7-2972-3	GVA-mutagen-F GVA-mutagen-R	*Dpn* I	pGEM-GVA-mutagen-1 pGEM-GVA-mutagen-2 pGEM-GVA-mutagen-3 pGEM-GVAmutagen-4
22	Cloning of *Ksp* I and *Nar* I digested pGEM-GVA-mutagen-1 and pGEM-GVA-mutagen-2 into *Ksp* I and *Nar* I digested pSK-GVA-IC2/T7-2972-3.	**Vector:** pSK-GVA-IC2/T7-2972-3 **Inserts:** pGEM-GVA-mutagen-1 pGEM-GVA-mutagen-2		*Nar* I *Ksp* I	pSK-GVA-mutagen-1 pSK-GVA-mutagen-2 pSK-GVA-mutagen-3 pSK-GVA-mutagen-4
23	pSK-GVA-IC2/T7-2972-3, pSK-GVA-mutagen-3 and pSK-GVA-mutagen-4 selected for transcription and infection				

Appendix 2: Vectors used in this study

pGEM®-T Easy Vector (Promega)

pGEM®-T Easy Vector (Promega) (including X-Gal (Promega) and IPTG) was used according to manufacturer's instructions in Promega technical manual no. 042 (diagram obtained from Promega technical manual no. 042)

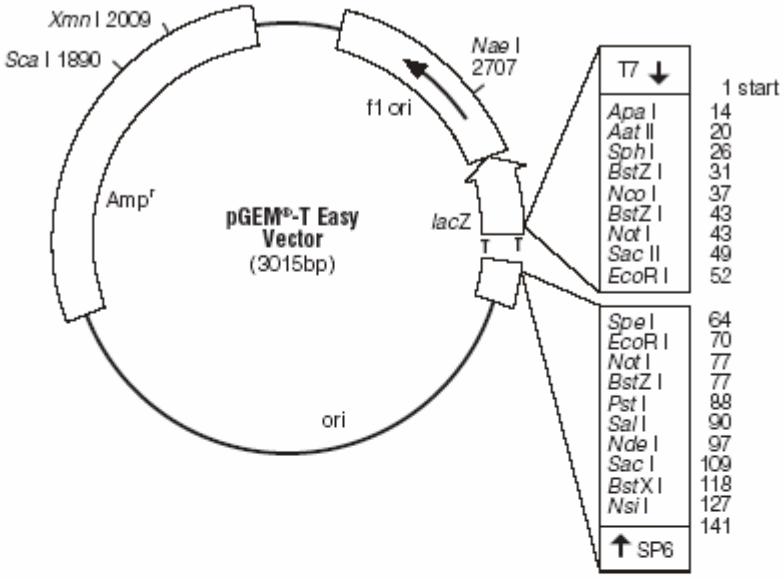


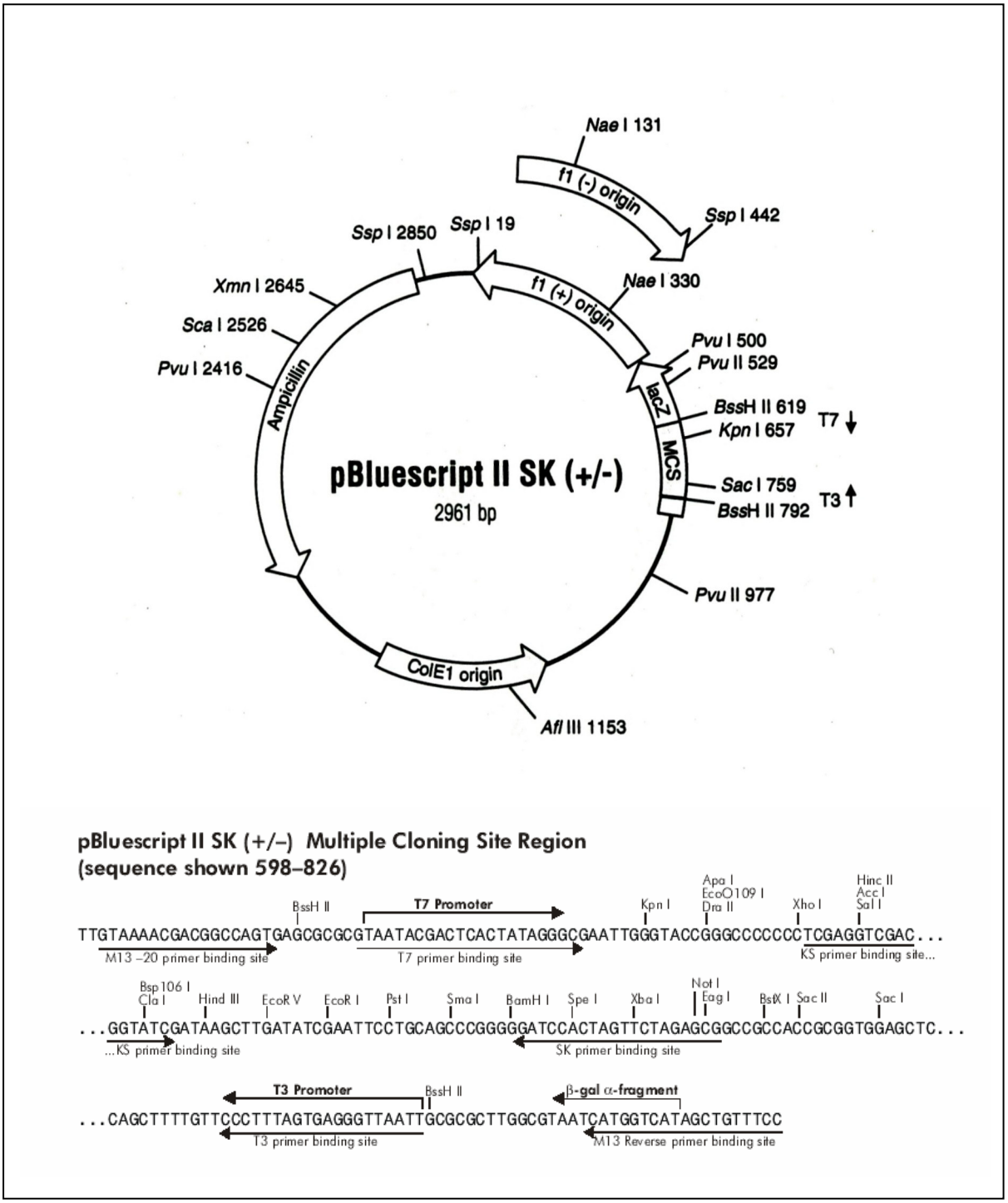
Figure 3. pGEM®-T Easy Vector circle map and sequence reference points.

pGEM®-T Easy Vector Sequence reference points:

T7 RNA Polymerase transcription initiation site	1
SP6 RNA Polymerase transcription initiation site	141
T7 RNA Polymerase promoter (–17 to +3)	2999–3
SP6 RNA Polymerase promoter (–17 to +3)	139–158
multiple cloning region	10–128
lacZ start codon	180
lac operon sequences	2836–2996, 166–395
lac operator	200–216
β-lactamase coding region	1337–2197
phage f1 region	2380–2835
binding site of pUC/M13 Forward Sequencing Primer	2956–2972
binding site of pUC/M13 Reverse Sequencing Primer	176–192

pBluescript®II SK (+) (Stratagene)

pBluescript®II SK (+) (Stratagene) was used according to manufacturer's instructions on website http://www.stratagene.com/manuals/212206.pdf. (diagram obtained from website http://www.stratagene.com/manuals/212206.pdf.).



Appendix 3: Sequences

gi|27502539|gb|AF007415.2| GVA isolate PA3, complete genome

```
   1 GAATA TTTAA CTTGA TTCCC ATTGA TTATA AGTGA ACATA CGTTA CCAGC
  51 ACCCA CGCTC TGCAA AGTAC CTGTA TTTTC ACCGC CATGT CAATA TCAGT
 101 ATCCT CCCAA CGTGT AGCAG TTTCC AACCT CTACA CTAAC GGATC TGAAG
 151 AATCA GTTAA AGCCA TAAAG GAGTT AAAGA GCAAA CGGTT ATTGG AAACC
 201 GAAAC CAGGT TAGAT GGGCT TTTTG ATTAC TACAT CCCAG ACACC TTGAG
 251 AGAAA TACTC TCTGG TTATG GTATG GAGTT CTCAG TCCAC TCTTT CCAGG
 301 GGCAC GCTCA CCCCG TAAGT AAAAT GATTG AGAAT CATAT GCTGT ATAGA
 351 GTAGC GCCGA ACTAT TTTTC TAACA ATACC CTTGT AGTTA GCTGT AAAGA
 401 AGGAA AAATC AAGCG CCTAA GATTA AAGAA CGCAG GTAAT AGGAA TTTGA
 451 ACTTC ACACA ATACA ATAGA TTGGT ACACG CCAAG GATCA TCACA GATAC
 501 GAGAA TGCGT TCAGA GAGCT AGACG TTGGC AATCT GACAA ATCTC ATAAA
 551 CAAGG AATCC CAAAG CGAGT GCATC TTCAT TCATG ACGAA GTTCA ATATT
 601 GGAGC CTAGA CGAAA TGCAG AGATT CTTAG GTAGC TTGTC TAAGG TAGAT
 651 AGAAT AGTGT ACAGC ATCAT ATACC CAGCG GAGGT AGAAG CGGGG TATTC
 701 ACAGA GCCTG TTCCC CGAGG CATAT ACTTT TGATT TAAAA GAGGG GAGGC
 751 TCATA TGGTA CCCTG ATGGT AAGGC CGAGG GCGCC TACAC TCAGC CAATC
 801 AACCC TTGGC TATTA AGGTG CTCAA AAACC GAGGA CTCCA AAGGC AGGTC
 851 ATGGA CTATA ACCAA ATTGC AAACT GTTGG GGCTC ACCAC CTCTT TAGTG
 901 CAATA AGGGG AAGTT ATCTG ACGGA GGAAT CATAC AAATA TGATA ACTTC
 951 ACGAT AATCA ACCCC AATGA TATCC TGAGA GGGAG GAGAG GGAGC AAGCC
1001 GCTAT ACTTG AGAGC CCGTA TGATA AAGCC CACAC TGCTA TATCT CCTGG
1051 CACTA AAGAA GAGTG ACTCG AACTC TGCGG TGGCC AAGCT GCGAA TGCTG
1101 AGCAG CCGCG AAGAG AACAT GGACG AGGCA CTATT CGTGG CTCAA CTGGC
1151 GAAGC AAATC AGGGA CACAG CGCTC TATGA CAAGA TGGGC AATCC AAGTC
1201 TGCGG AGCAT ACTGT CCGAA TCATT CTACG ATGTG GCCGG AAATC TGTTC
1251 ACTAG ATTGT TCAAC AGGCC AGAAT ACGAT GCCAG GTGCT TGGAG AAGTT
1301 CATCA GGGCC TGCGA AACCA CAGAG ATACA CATCG GCAGA AGGTT CATGG
1351 AGGGG GTGCT ACGGG GCTCC AGTTA TCAGG TTCAG AAAGT AGTAG ATTGG
1401 GTTGA GGACG AAGAG GCCAA CGCGC TGAGT GAGGT CAATC TTGTT GAACT
1451 CACGA GAAAT GAAAG AGAGC CACAG CCATA CGGAT CAGAG CAATT GCAAG
1501 GGTCA GGTGC TAGGG TGAGG TTCCC TTTGA GTCGG CTAAT GAGGG CGCAC
1551 GAAAT GGTGC ATGAG GCTCA GTTCG AGGTC AACTT TAATC ACATC AGAAT
1601 AGTGT GCTCC TCTAA AGCCC TGGTG CACTT AAGGC AATAC CTGAT CAGAC
1651 TGGGC AGGTT CAAAT TCATG GAAGG CAGAG CAATA GAGGA CATAC AAGAC
1701 ATTCA AGAGG GGCTT GAACT GGGCC TGATC ACAGA AGAAG AGGCT GAATC
1751 ACGCC TGCTA CCACC GGACA ACATC GCACC CAGCG AAGAA GCCCA AGAAG
1801 AGGTG CAGCA AAGCG GTCCC ATTGA AGAGG TGAGT GAAAA GTTCA TGGGG
1851 GCACG CAAAT TGGTT AAGGA GAACA TTTAC TCTGA AAAAC TGCAC AACAG
1901 AGAGG TGGCC TTCTA CAGCC GGTAC AGCAA GCAGT ACAGT TACAA AGGGG
1951 GCTCA CACCA CTCAC TCGGA TGGGA TAAAG CTTTG GATGA GCTCA TACAA
2001 GAGTT GGGCC TAGAT GAGAG CTATG ACCAC TGTCT CATTC AACGG TACTC
2051 TGAGG GTGGC AGTAT CAATT TCCAC GCCGA TGATG AGCCC TGCTA TCTAC
2101 CAGGA GGAAC CGTGG TGACG GTGAA TTTGA ATGGG GAAGC AATCT TTGAG
2151 TTGAA GGAGA ACACA TCAGG CAAAA CCGAG TCAAA AAGAT TAAAG GATGG
2201 TGATG TGTTC ACAAT GGGCG CAGGA ATGCA GCAAA CCCAT AAACA CAAGG
2251 TCACC TCACT CACTA ACGGG CGATG TAGTA TCACT CTGAG GAACA AGACA
2301 GTGGA CTACG AGGCA AGGAC GGAGG CAATC AATTC CGATT ACGAG GAGGA
2351 TGTAG CAGCA CTGGA CGAGG GAATT GAGTA CCTCC AAAAG AATCA AGGGA
2401 ATATG TGCTC CCTGA AGGCT TTTGC TGAAC ACATG CAATT GAGCA TACCA
2451 AGCGT GATCT CAATA GTAAA TGGCG CATCG CCACA AACAC TGAGA GAAAT
2501 AGAAG GTGGT GGGAT CAGTG TCGCG ACGCT TGTGA ATCTC TCGAG GGCGC
2551 TGGAC TTCCC GATCG CCATT CACGG AGAGA GAGGA TATGC TGAAA CGCCT
2601 GGAAC ATATA GAAGA CTGCA CCTCA AGATC ACAAC AGGGC ACGTG GAACC
2651 ATTCG ATGGA GTCAT AAGTA AGGGC GGTCT TAAGG AGGCC CTGTT GCTGA
```

2701 GCGAT GGCGT TGGAA TAGGT CTATT CAGAG TGGAC AAAAA TAAGG CTGAC
2751 CGCCT TGCCC AAAGT TTCTA CAACG GGAAC ACAGG GGTGC TTCTG GGAAA
2801 GTACA ATAAA GGCAA GATGC ACTCA GAGGG CATAG ATGAA CCAAA GGAGG
2851 TGCTC ACCGC TTTCG GATTT GCAGG TTCAG GGAAA AGTCA CTGGT GTCAA
2901 ACCAT ACTTA AACAC TGTTC CGTAG AGAAA GCCCT CGTTA TCAGC CCACG
2951 GAAGG TACTC AAAGA CGACT GGACC TCAAA GATCT CAAAG AGGCA CAAAG
3001 TGGTG ACATT CGAGG TGGCT TTCAT GGATG ATTAT GGCTG CAAGG ATGTG
3051 GTTAT CGACG AGATC GGGTT GCTAC CTCCT GGGTA CATCG ATTTA GTCAT
3101 TGCTG TGCAT AGACC ACGTA CACTG GTGCT CCTGG GGGAT CCACT GCAAA
3151 GCACC TACCA CAGCA AGCGT GACAA CGTAG TCTTG GAGGC CAGTC AAGAG
3201 GACAT CTTTG CCAGG ATCAA AGGGC AGCTA CCATA CCTCT GCTAC TCCCA
3251 CCGGC TCCCA AAGAA TTGCA GATTA TTCGA GCTAG AGTGC ATGGG AGCGG
3301 AGAGC GAAAA GAGGG TAACC TATCG CTCGA GTCGC CTCAA GGACG AGCCG
3351 GTCAT CTGCG CGACC AGGGC CATGA AGGAG GAAAA AGGGT CAGGC TGGTA
3401 CACAG TGAGC GAGAC CCAAG GGCTA AGCTT CAAAT CGTGT CTCAT TCACC
3451 TTGAT GAGCA TTGGG CGAAG AAAGA GGATG AGGAT GTGAT GGTGG CCCTG
3501 ACAAG ATCAA GAGGG GAAAT AGGCA TAAAC GTCAC CAATG CACTC AAGAA
3551 GACGT TGATC GCGAG CGCAA AAAGT ACGCT GCTCA GGAAG GTACT CAAGG
3601 GGGAG ACATA TAGGC GTTCT GAAAT CGTTG CCATG GCCCG CAAGC ATATA
3651 CCAGA AGTCA CCGTT CTCTT CGAGG AGAGT AGGCT AGCAG AAAAT GTGGA
3701 TTATG AAGCT AGGTT AGCTG GTGAT CCATA CTTGA AGTCA CTGCT CTCAC
3751 TCTAT GATGA GATTG AAATG GACGA GGTGG AAATT GAAGA ACCGG TCGCT
3801 CTGGA ACCGG TTAAA ACCCA TCTCG CATTG AGTGA GAAGA TGAAC GAACT
3851 GGCCC CTTTC GAATT GAAGG CCAAA GAGTA CAGAG AGCAG TACAC GGATG
3901 CTGGA AGGAC TGAGC AGATA GACGA GGAGG GGTAT AGCGG GGAAG AGGCC
3951 AATCC GATGA CCCAC AAGGC TTTGT ACCTA AAACA CACCT CTGAT GACAC
4001 TGCCA CGTTT ATGAT GTCTG TAAGG AAGCG ATTGC GCTTC AGAAA CTATG
4051 AGGCA AACAG AAGGA AGTCC AAGGC ATGTA ATGGC ATCGG AAAGC AGATG
4101 TTCTC GGTCT TCAAG GAGAC ATACC AGTTG AAGGA GATCG ATTCC TTACC
4151 CGATC TGGAG AGGTG CGAAA TGGAA TTCAT GAAGA AACGC GTGGA GAAGA
4201 GCGCG GGTCT AATAG AGCAG CATGC CGGGC GTAGT GACCC AGACT GGCCA
4251 AGCAA CTACC TGAAG ATCTT TCTGA AACAG CAGAC CTGTA CAAAA ATGGA
4301 GAAAC GCGGC GTGGA CGCAA AGGCG GGTCA AACCA TAGCC TGCTT CGCCC
4351 ACTCC GTGCT TTGTA GATTC GGCCC TATAC TCAGA CAGAC AGAAA GGGCA
4401 CTCAG GGAAT TGTTA CCGGA GAATG TAATG ATATA CTCTC AGAAG AACTA
4451 CATGG ACCTG GACAA ATGGG CCAAG ACCTG GGTCG AGAGC ATGAT GGGGA
4501 CGGAT TCCGA TTATG AGGCC TTTGA TCGAT CTCAA GATGA GAAAG TGCTA
4551 GATCT AGAGG TAGAA GTGCT TCGCT TCTTC CTATG GCCGG AGGAT TTAAT
4601 CAGGG AGTAC GAGGA GCTTA AACTG ATGAT GGGGT GCGCA CTGGG TGATC
4651 TAGCC GTCAT GCGGT TTTCG GGCGA GTTCG GGACA TTCTT CTTCA ACACC
4701 ATATG CAACA TGGTA TTTAG CTGCA TGCGC TATCA TATGG ATCGC AACAC
4751 CCCAA TATGT TTCGC TGGGG ATGAC ATGTA TTCAC CTGGG ATCCT GAGGA
4801 CAAAG AATGA TTATG AGGGG ACTCT AGACC AACTG ACTCT GAAAG CGAAA
4851 GTGCA TATAT CTGAA GAGCC ATTGT TCTGT GGTTG GAGGA TGAGC CCCTT
4901 CGGGA TAGTG AAGGA GCCCA ATCTA ATCCT GGATC GCTGG AAGAT CGCGA
4951 AAAGA AGTGG AAATC TGTCA CTCTG TCTCG TCAAC TACGC GATTG AGGCT
5001 AGTTT TGGGT ACAGG TTAAG CGAGT ACCTG TATGA TGTTA ATATA GACGT
5051 CGATG CTCAG CAAGA ACTGA TAAGG GAGAT TGTGG AAAAG AAGCA CCTAC
5101 TGCCG AAAAA GATAG CAGAT CTTTT CAGTG AGGAC GAGTG CGAGG CGCAC
5151 AGCGA TGGCG ACGAG GACTT TTTGA GTAAT GACAT CGCAA AGTTG TACAG
5201 AGTTG GTTGA GTTCT TGAGT CAGGG TGGTG ATAGG TGCAG TTTAG ATAGA
5251 ACGTC ACTTG AGTCC TTGTC GTATA CACAG TGTCT GTACT TATTG TCTGA
5301 TCTTA GAGGC CTAGG ATATC ACAGT ATAGA TAGCA TATTA CATAT ACTGG
5351 AGTGT AGCGA GGCGG AACGT TTCGA AATCT ATCGA ATCTT CCGGA GGCAC
5401 GGAAT CGGGA TTGGA GAGGC TTTGC AGCTA GGCGT GAGAA AAAGT TTGTG
5451 TTACA GTCCG AGAAG CACCG CAGAG ATACT GGACG ACCTC CTGTC CAGGA
5501 TAAGC AGGGG AGGTG CCTTC CTACC AAGTG ATCTG GGTGC AGTGA AAGGA
5551 GAACT TCTGG TTACA TTCCA ACCGT CCAGA TTAAG TGTAG ACCTC TACGC
5601 TAATA CCAAG AAGGT GTGTA CCAGG ACGTG TCAAG GCAAA GGCGA CGCAG

```
5651 GTTAT GTCGC ACGGA GATTC TCAGG CTACA AAGGG CTCGC TCTCA GATCC
5701 CAAGG AAATT AAGAT ATTCA ATGTC AAGAG AAACA CAAAG GATCT CGAGA
5751 CACTG AACAA GAGCC TCCAC AGGGG AGATG TGTAC GACAC GGAGC TAATA
5801 GAGCG TGTGT TTCCA AGACG CACAA AGAAA TGCGT CGTGC ACAAG GAGCT
5851 AGTCG TTCAA GATGG CCGAG TGGAT TGCGA GCTGG ATCTG ATGGA CGGGG
5901 GTTTG GATGA TATAG ATGAG ACAGA ATACC CGCTC TACCA CGTGG GGTGT
5951 ATCGT CATCG CTTTA ATGCC ACATG GTAAG AATCT CCAGG GCAAG GTGGT
6001 TGTCG AGGTG ATAGA TACCC GGCTC AAGGT TGGGT CAGGG CGGAT CTCTA
6051 GGAGC GTGAT GGATA TGTCG AAACC CTTAA GCGCA TGTGC CGACT TCCCG
6101 GGGTA CTTCA TTAGT ACTTC CGATC TGTTG AACGG ATACA CGCTG CACCT
6151 CTCCA TAACC ACAAC TGATC TGCAG TTTGT GGACG GCGTG CACCC ATTCA
6201 GCGTA CAACT CATGA ATATC GGGCG TCTCT GCGGG GATGA CATGA AGACA
6251 AGGTA TGCGG TAGGA GAGGC CTCGA AGATC CTGCA TCAGA ACATC CTAAA
6301 CTCAC AAGGC GATGG CGAGA TGATA CCGCG TGGTG TGCAA GTGCA AAAAG
6351 TACCA GATAC CCTAG TTATG CCAGA AGTTT ATGAG ACAAT AAAGA GACTT
6401 GGGTT GAAGA CAAAT GGCAC ACTAC GCCAA GAGGG TGGAG ATAAG GGCGA
6451 TCATA GAGGA ACTGG TGCTG GCGAA AGCCA CGCCA ACTGA GGACG CATCC
6501 GAAAG TGGCT ACGAC CGAAA TATGT ACCTG AATAC TCTCT TCGGG TACAT
6551 CGCCT TGGTC GGGAC AAGCA AGAAG GCGGT TCATT ATGGG GAGGT AGATA
6601 TAGTA GGACC TAAAG CTAGC AAGAA AACGG GGATC GACCC TCGCG GAAAG
6651 ATGGT TGTAT CGGAG CTGGT CGGCA GAATG CGCAC TCTGA GCGTG GCGGT
6701 TAGTG AGGGA CCTGT AAAAG GTGCA ACCTT GAGGC AGATG TGTGA GCCTT
6751 TCGCA CAGAA CGCCT ATGAC TTCCT CGTGC TTATG GCGGA GATGG GCACC
6801 TATTC ACAGC TAGCC ACAAA GATGA CGAGG TCAGG CTTTA AGGAA CCACA
6851 GGTTA TGTTC GACTT TGCGT CGGGC TTAGA TTTGA AAGCA CTGAC CCTAC
6901 AAGAA GCCAC TGTGA TACAG GCTAT GCACT CACGT CTCTT TCGGA CAGAA
6951 GGTGC AAAGG GGGTA TTCAA CGCTC AATCA TCTGT CGGTG AGCAG GCTGT
7001 CGAGA TATAG ATGGA TGACC CATCG TTTCT CGCGG GTAGG TCCAC GTTTG
7051 CTAAA CGTAG GCGCG CTAGG CGCAT GAATG TGTGT AAGTG TGGTG CTATA
7101 ATGCA CAATA ATGAG GACTG CAAGT CCAGT AGTAT TTCCG GTCAT AAACT
7151 CGATA GATTG CGGTT CGTCA AAGAG GGAAG AGTAA CTCTA ACGGG TGAGA
7201 CTCCT GTTTA CCGGA CTTGG ATCAG ATGGG TTGAG ACTGA ATATC ACATA
7251 TATAT ATTAG AAACC TCAGA CGATG AGAAA TGATC GTCTA TCCTA GCTAA
7301 TAAGT GAATA ACAAC ATAAT GGTAA TGCGA GCTAG GTTGT CATAC TATGA
7351 C
//
```

gi|33590464|gb|AY244516.1| GVA, complete genome

```
   1 GAATA TTTAA CTTGA TTCCC ATTGA TTATA AGTGA ACATA CGTTA CCAGC
  51 ACCCA CGCTC TGCAA AGTAC CTGTA TTTTC ACCGC CATGT CAATA TCAGT
 101 ATCCT CCCAA CGTGT AGCAG TTTCC AACCT CTACA CTAAC GGATC TGAAG
 151 AATCA GTTAA AGCCA TAAAG GAGTT AAAGA GCAAA CGGTT ATTGG AAACC
 201 GAAAC CAGGT TAGAT GGGCT TTTTG ATTAC TACAT CCCAG ACACC TTGAG
 251 AGAAA TACTC TCTGG TTATG GTATG GAGTT CTCAG TCCAC TCTTT CCAGG
 301 GGCAC GCTCA CCCCG TAAGT AAAAT GATTG AGAAT CATAT GCTGT ATAGA
 351 GTAGC GCCGA ACTAT TTTTC TAACA ATACC CTTGT AGTTA GCTGT AAAGA
 401 AGGAA AAATC AAGCG CCTAA GATTA AAGAA CGCAG GTAAT AGGAA TTTGA
 451 ACTTC ACACA ATACA ATAGA TTGGT ACACG CCAAG GATCA TCACA GATAC
 501 GAGAA TGCGT TCAGA GAGCT ~GACG TTGGC A~TCT GACAA ~TCTC ATAAA
 551 CAAGG AATCC CAAAG CGAGT GCATC TTCAT TCATG ACGAA GTTCA ATATT
 601 GGAGC CTAGA CGAAA TGCAG AGATT CTTAG GTAGC TTGTC TAAGG TAGAT
 651 AGAAT AGTGT ACAGC ATCAT ATACC CAGCG GAGGT AGAAG CGGGG TATTC
 701 ACAGA GCCTG TTCCC CGAGG CATAT ACTTT TGATT TAAAA GAGGG GAGGC
 751 TCATA TGGTA CCCTG ATGGT AAGGC CGAGG GCGCC TACAC TCAGC CAATC
 801 AACCC TTGGC TATTA AGGTG CTCAA AAACC GAGGA CTCCA AAGGC AGGTC
 851 ATGGA CTATA ACCAA ATTGC AAACT GTTGG GGCTC ACCAC CTCTT TAGTG
 901 CAATA AGGGG AAGTT ATCTG ACGGA GGAAT CATAC AAATA TGATA ACTTC
 951 ACGAT AATCA ACCCC AATGA TATCC TGAGA GGGAG GAGAG GGAGC AAGCC
```

```
1001 GCTAT ACTTG AGAGC CCGTA TGATA AAGCC CACAC TGCTA TATCT CCTGG
1051 CACTA AAGAA GAGTG ACTCG AACTC TGCGG TGGCC AAGCT GCGAA TGCTG
1101 AGCAG CCGCG AAGAG AACAT GGACG AGGCA CTATT CGTGG CTCAA CTGGC
1151 GAAGC AAATC AGGGA CACAG CGCTC TATGA CAAGA TGGGC AATCC AAGTC
1201 TGCGG AGCAT ACTGT CCGAA TCATT CTACG ATGTG GCCGG AAATC TGTTC
1251 ACTAG ATTGT TCAAC AGGCC AGAAT ACGAT GCCAG GTGCT TGGAG AAGTT
1301 CATCA GGGCC TGCGA AACCA CAGAG ATACA CATCG GCAGA AGGTT CATGG
1351 AGGGG GTGCT ACGGG GCTCC AGTTA TCAGG TTCAG AAAGT AGTAG ATTGG
1401 GTTGA GGACG AAGAG GCCAA CGCGC TGAGT GAGGT CAATC TTGTT GAACT
1451 CACGA GAAAT GAAAG AGAGC CACAG CCATA CGGAT CAGAG CAATT GCAAG
1501 GGTCA GGTGC TAGGG TGAGG TTCCC TTTGA GTCGG CTAAT GAGGG CGCAC
1551 GAAAT GGTGC ATGAG GCTCA GTTCG AGGTC AACTT TAATC ACATC AGAAT
1601 AGTGT GCTCC TCTAA AGCCC TGGTG CACTT AAGGC AATAC CTGAT CAGAC
1651 TGGGC AGGTT CAAAT TCATG GAAGG CAGAG CAATA GAGGA CATAC AAGAC
1701 ATTCA AGAGG GGCTT GAACT GGGCC TGATC ACAGA AGAAG AGGCT GAATC
1751 ACGCC TGCTA CCACC GGACA ACATC GCACC CAGCG AAGAA GCCCA AGAAG
1801 AGGTG CAGCA AAGCG GTCCC ATTGA AGAGG TGAGT GAAAA GTTCA TGGGG
1851 GCACG CAAAT TGGTT AAGGA GAACA TTTAC TCTGA AAAAC TGCAC AACAG
1901 AGAGG TGGCC TTCTA CAGCC GGTAC AGCAA GCAGT ACAGT TACAA AGGGG
1951 GCTCA CACCA CTCAC TCGGA TGGGA TAAAG CTTTG GATGA GCTCA TACAA
2001 GAGTT GGGCC TAGAT GAGAG CTATG ACCAC TGTCT CATTC AACGG TACTC
2051 TGAGG GTGGC AGTAT CAATT TCCAC GCCGA TGATG AGCCC TGCTA TCTAC
2101 CAGGA GGAAC CGTGG TGACG GTGAA TTTGA ATGGG GAAGC AATCT TTGAG
2151 TTGAA GGAGA ACACA TCAGG CAAAA CCGAG TCAAA AAGAT TAAAG GATGG
2201 TGATG TGTTC ACAAT GGGCG CAGGA ATGCA GCAAA CCCAT AAACA CAAGG
2251 TCACC TCACT CACTA ACGGG CGATG TAGTA TCACT CTGAG GAACA AGACA
2301 GTGGA CTACG AGGCA AGGAC GGAGG CAATC AATTC CGATT ACGAG GAGGA
2351 TGTAG CAGCA CTGGA CGAGG GAATT GAGTA CCTCC AAAAG AATCA AGGGA
2401 ATATG TGCTC CCTGA AGGCT TTTGC TGAAC ACATG CAATT GAGCA TACCA
2451 AGCGT GATCT CAATA GTAAA TGGCG CATCG CCACA AACAC TGAGA GAAAT
2501 AGAAG GTGGT GGGAT CAGTG TCGCG ACGCT TGTGA ATCTC TCGAG GGCGC
2551 TGGAC TTCCC GATCG CCATT CACGG AGAGA GAGGA TATGC TGAAA CGCCT
2601 GGAAC ATATA GAAGA CTGCA CCTCA AGATC ACAAC AGGGC ACGTG GAACC
2651 ATTCG ATGGA GTCAT AAGTA AGGGC GGTCT TAAGG AGGCC CTGTT GCTGA
2701 GCGAT GGCGT TGGAA TAGGT CTATT CAGAG TGGAC AAAAA TAAGG CTGAC
2751 CGCCT TGCCC AAAGT TTCTA CAACG GGAAC ACAGG GGTGC TTCTG GGAAA
2801 GTACA ATAAA GGCAA GATGC ACTCA GAGGG CATAG ATGAA CCAAA GGAGG
2851 TGCTC ACCGC TTTCG GATTT GCAGG TTCAG GGAAA AGTCA CTGGT GTCAA
2901 ACCAT ACTTA AACAC TGTTC CGTAG AGAAA GCCCT CGTTA TCAGC CCACG
2951 GAAGG TACTC AAAGA CGACT GGACC TCAAA GATCT CAAAG AGGCA CAAAG
3001 TGGTG ACATT CGAGG TGGCT TTCAT GGATG ATTAT GGCTG CAAGG ATGTG
3051 GTTAT CGACG AGATC GGGTT GCTAC CTCCT GGGTA CATCG ATTTA GTCAT
3101 TGCTG TGCAT AGACC ACGTA CACTG GTGCT CCTGG GGGAT CCACT GCAAA
3151 GCACC TACCA CAGCA AGCGT GACAA CGTAG TCTTG GAGGC CAGTC AAGAG
3201 GACAT CTTTG CCAGG ATCAA AGGGC AGCTA CCATA CCTCT GCTAC TCCCA
3251 CCGGC TCCCA AAGAA TTGCA GATTA TTCGA GCTAG AGTGC ATGGG AGCGG
3301 AGAGC GAAAA GAGGG TAACC TATCG CTCGA GTCGC CTCAA GGACG AGCCG
3351 GTCAT CTGCG CGACC AGGGC CATGA AGGAG GAAAA AGGGT CAGGC TGGTA
3401 CACAG TGAGC GAGAC CCAAG GGCTA AGCTT CAAAT CGTGT CTCAT TCACC
3451 TTGAT GAGCA TTGGG CGAAG AAAGA GGATG AGGAT GTGAT GGTGG CCCTG
3501 ACAAG ATCAA GAGGG GAAAT AGGCA TAAAC GTCAC CAATG CACTC AAGAA
3551 GACGT TGATC GCGAG CGCAA AAAGT ACGCT GCTCA GGAAG GTACT CAAGG
3601 GGGAG ACATA TAGGC GTTCT GAAAT CGTTG CCATG GCCCG CAAGC ATATA
3651 CCAGA AGTCA CCGTT CTCTT CGAGG AGAGT AGGCT AGCAG AAAAT GTGGA
3701 TTATG AAGCT AGGTT AGCTG GTGAT CCATA CTTGA AGTCA CTGCT CTCAC
3751 TCTAT GATGA GATTG AAATG GACGA GGTGG AAATT GAAGA ACCGG TCGCT
3801 CTGGA ACCGG TTAAA ACCCA TCTCG CATTG AGTGA GAAGA TGAAC GAACT
3851 GGCCC CTTTC GAATT GAAGG CCAAA GAGTA CAGAG AGCAG TACAC GGATG
3901 CTGGA AGGAC TGAGC AGATA GACGA GGAGG GGTAT AGCGG GGAAG AGGCC
```

3951 AATCC GATGA CCCAC AAGGC TTTGT ACCTA AGACA CACCT CTGAT GACAC
4001 TGCCA CGTTT ATGAT GTCTG TAAGG AAGCG ATTGC GCTTC AGAAA CTATG
4051 AGGCA AACAG AAGGA AGTAC AAGGC ATGTA ATGGC ATCGG AAAGC AGATG
4101 TTCTC GGTCT TCAAG GAGAC ATACC AGTTG AAGGA GATCG ATTCC CTACC
4151 CGATC TGGAG AGGTG CGAAA TGGAA TTCAT GAAGA AACGC GTGGA GAAGA
4201 GCGCG GGTCT AATAG AGCAG CATGC CGGGC GTAGT GACCC AGACT GGCCA
4251 AGCAA CTACC TGAAG ATCTT TCTGA AACAG CAGAC CTGTA CAAAA ATGGA
4301 GAAAC GCGGC GTGGA CGCAA AGGCG GGTCA AACCA TAGCC TGCTT CGCCC
4351 ACTCC GTGCT TTGTA GATTC GGCCC TATAC TCAGA CAGAC AGAAA GGGCA
4401 CTCAG GGAAT TGTTA CCGGA GAATG TAATG ATATA CTCTC AGAAG AACTA
4451 CATGG ACCTG GACAA ATGGG CCAAG ACCTG GGTCG AGAGC ATGAT GGGGA
4501 CGGAT TCCGA TTATG AGGCC TTTGA TCGAT CTCAA GATGA GAAAG TGCTA
4551 GATCT AGAGG TAGAA GTGCT TCGCT TCTTC CTATG GCCGG AGGAT TTAAT
4601 CAGGG AGTAC GAGGA GCTTA AACTG ATGAT GGGGT GCGCA CTGGG TGATC
4651 TAGCC GTCAT GCGGT TTTCG GGCGA GTTCG GGACA TTCTT CTTCA ACACC
4701 ATATG CAACA TGGTA TTTAG CTGCA TGCGC TATCA TATGG ATCGC AACAC
4751 CCCAA TATGT TTCGC TGGGG ATGAC ATGTA TTCAC CTGGG ATCCT GAGGA
4801 CAAAG AATGA TTATG AGGGG ACTCT AGACC AACTG ACTCT GAAAG CGAAA
4851 GTGCA TATAT CTGAA GAGCC ATTGT TCTGT GGTTG GAGGA TGAGC CCCTT
4901 CGGGA TAGTG AAGGA GCCCA ATCTA ATCCT GGATC GCTGG AAGAT CGCGA
4951 AAAGA AGTGG AAATC TGTCA CTCTG TCTCG TCAAC TACGC GATTG AGGCT
5001 AGTTT TGGGT ACAGG TTAAG CGAGT ACCTG TATGA TGTTA ATATA GACGT
5051 CGATG CTCAG CAAGA ACTGA TAAGG GAGAT TGTGG AAAAG AAGCA CCTAC
5101 TGCCG AAAAA GATAG CAGAT CTTTT CAGTG AGGAC GAGTG CGAGG CGCAC
5151 AGCGA TGGCG ACGAG GACTT TTTGA GTAAT GACAT CGCAA AGTTG TACAG
5201 AGTTG GTTGA GTTCT TGAGT CAGGG TGGTG ATAGG TGCAG TTTAG ATAGA
5251 ACGTC ACTTG AGTCC TTGTC GTATA CACAG TGTCT GTACT TATTG TCTGA
5301 TCTTA GAGGC CTAGG ATATC ACAGT ATAGA TAGCA TATTA CATAT ACTGG
5351 AGTGT AGCGA GGCGG AACGT TTCGA AATCT ATCGA ATCTT CCGGA GGCAC
5401 GGAAT CGGGA TTGGA GAGGC TTTGC AGCTA GGCGT GAGGA AAAGT TTGTG
5451 TTACA GTCCG AGAAG CACCG CAGAG ATACT GGACG ACCTC CTGTC CAGGA
5501 TAAGC AGGGG AGGTG CCTTC CTACC AAGTG ACCTG GGTGC AGTGA AAGGA
5551 GAACT TTTGG TTACA TTCCA ACCGT CCAGA TTAAG TGTAG ACCTC TACGC
5601 TAATA ACAAG AAGGT GTGTA CCAGG ACGTG TCAAG GCAAA GGCGA CGCAG
5651 GTTAT GTCGC ACGGA GATTC TCAGG CTACA AAGGG CTCGC TCTCA GATCC
5701 CAAGG AAATT AAGAT ATTCA ATGTC AAGAG AAACA CAAAG GATCT CGAGA
5751 CACTG AACAA GAGCC TCCAC AGGGG AGATG TGTAC GACAC GGAGC TAATA
5801 GAGCG TGTGT TTCCA AGACG CACAA AGAAA TGCGT CGTGC ACAAG GAGCT
5851 AGTCG TTCAA GATGG CCGAG TGGAT TGCGA GCTGG ATCTG ATGGA CGGGG
5901 GTTTG GATGA TATAG ATGAG ACAGA ATACC CGCTC TACCA CGTGG GGTGT
5951 ATCGT CATCG CTTTA ATGCC ACATG GTAAG AATCT CCAGG GCAAG GTGGT
6001 TGTCG AGGTG ATAGA TACCC GGCTC AAGGT TGGGT CAGGG CGGAT CTCTA
6051 GGAGC GTGAT GGATA TGTCG AAACC CTTAA GCGCA TGTGC CGACT TCCCG
6101 GGGTA CTTCA TTAGT ACTTC CGATC TGTTG AACGG ATACA CGCTG CACCT
6151 CTCCA TAACC ACAAC TGATC TGCAG TTTGT GGACG GCGTG CACCC ATTCA
6201 GCGTA CAACT CATGA ATATC GGGCG TCTCT GCGGG GATGA CATGA AGACA
6251 AGGTA TGCGG TAGGA GAGGC CTCGA AGATC CTGCA TCAGA ACATC CTAAA
6301 CTCAC AAGGC GATGG CGAGA TGATA CCGCG TGGTG TGCAA GTGCA AAAAG
6351 TACCA GATAC CCTAG TTATG CCAGA AGTTT ATGAG ACAAT AAAGA GACTT
6401 GGGTT GAAGA CAAAT GGCAC ACTAC GCCAA GAGGG TGGAG ATAAG GGCGA
6451 TCATA GAGGA ACTGG TGCTG GCGAA AGCCA CGCCA ACTGA GGACG CATCC
6501 GAAAG TGGCT ACGAC CGAAA TATGT ACCTG AATAC TCTCT TCGGG TACAT
6551 CGCCT TGGTC GGGAC AAGCA AGAAG GCGGT TCATT ATGGG GAGGT AGATA
6601 TAGTA GGACC TAAAG CTAGC AAGAA AACGG GGATC GACCC TCGCG GAAAG
6651 ATGGT TGTAT CGGAG CTGGT CGGCA GAATG CGCAC TCTGA GCGTG GCGGT
6701 TAGTG AGGGA CCTGT AAAAG GTGCA ACCTT GAGGC AGATG TGTGA GCCTT
6751 TCGCA CAGAA CGCCT ATGAC TTCCT CGTGC TTATG GCGGA GATGG GCACC
6801 TATTC ACAGC TAGGC CACAA AGATA CGAGG TCAGG CTTTA AGGAA CCACA
6851 GGTTA TGTTC GACTT TGCGT CGGGC TTAGA TTTGA AAGCA CTGAC CCTAC

```
     6901 AAAAA GCCCC TGTGA TACAG GCTAT GCACT CACGT CTCTT TCGGA CAGAA
     6951 GGTGC AAAGG GGGTA TTCAA CGCTC AATCA TCTGT CGGTG AGCAG GCTGT
     7001 CGAGA TATAG ATGGA TGACC CATCG TTTCT CGCGG GTAGG TCCAC GTTTG
     7051 CTAAA CGTAG GCGCG CTAGG CGCAT GAATG TGTGT AAGTG TGGTG CTATA
     7101 ATGCA CAATA ATGAG GACTG CAAGT CCAGT AGTAT TTCCG GTCAT AAACT
     7151 CGATA GATTG CGGTT CGTCA AAGAG GGAAG AGTAA CTCTA ACGGG TGAGA
     7201 CTCCT GTTTA CCGGA CTTGG ATCAG ATGGG TTGAG ACTGA ATATC ACATA
     7251 TATAT ATTAG AAACC TCAGA CGATG AGAAA TGATC GTCTA TCCTA GCTAA
     7301 TAAGT GAATA ACAAC ATAAT GGTAA TGCGA GCTAG GTTGT CATAC TATGC
     7351 CAAAA AAAAA AAAAA AAAAA AAAAA AAAAA A
//
```

gi|20153359|ref|NC_003604.1| GVA, complete genome

```
        1 GAATA TTTAA CTTGA TTCCC ATCGA TTATA AGTGA ACAAA CGTTG CCAGC
       51 ACCCA CGCTC TGCAA AGCAC CTGCT AATTC ACTAC CATGT CGATA TCAGT
      101 ATCCT CCCAA CGTGT CGCAG TCTCC AACCT GTACA CGAAT GGATC TGAGG
      151 AGTCA GTTAA AGCTA TTAAG GAGTT GAAAA GCAAA CGGTT ATTGG AAACC
      201 GAAAC CAGGC TAGAT GGACT ATTTG ATTAC TACAT TCCAG ATACC TTAAG
      251 AGAAA TTCTT ACAGG CTATG GTATG GAGTT CAGTG TCCAC TCTTT CCAAG
      301 GACAT GCTCA CCCCG TAAGT AAGAT GATAG AAAAC CACAT GTTGT ATAGA
      351 GTAGC ACCAA ATTAC TTTTC TAGTA ATACA TTGGT AGTTA GTTGT AAAGA
      401 GAGCA AGATA AAGCG CCTAC GTCTG AAGAA TGCAA ACAAT AGGAA TTTGA
      451 ACTTC ACACA GTACA ATAGA TTGGT GCATG CAAAC CATCA CCACA GATAC
      501 GAAAA CGCAT TCAGA GAACT CGACG TTGGG AATTT GACAA ATCTC ATAAA
      551 CAAGG AAGAT CAGAG TGAAT GCATA TTCAT ACATG ATGAG GTTCA ATACT
      601 GGAGT CTAGA CGAGA TGCAA AGGTT CTTAG GTAGT CTCTC CAAAG TCGAT
      651 AGAGT AGTGT ATAGT ATCAT ATACC CATCA GAGGT CGAGG CTGGA TATTC
      701 ACAAA GCCTA TTTCC AGAGG CGTAC ACTTT TGACC TCAAA GACAG GAGAT
      751 TAGTA TGGTA CCCGG ACGGG AAGGC TGAAG GAGCC TATAC GCAGC CTGTG
      801 AACCC TTGGC TGCTC AGATG CTCAA AAACG GAGGA TTCGA AGGGG CGATC
      851 ATGGA CGATC ACAAA GCTCC AAACC ATCGG AGCTC ATCAC CTCTT TAGTG
      901 CAATA AAGGG CAGCT ACTTG ACGGA GGAAT CGTAC AAGTA CGATA ACTTC
      951 ACGAT CATAA ACCCT AACGA TGTTC TGAAG GGAAA GAGAG GCGGG AAACC
     1001 ACTCT ATCTA CGGGC CCGCA TGATC AAGCC AACAC TTCTG TACCT CTTGG
     1051 CTCTG AAGAA AAGCG ACTCC AACTC CGCAG TCGCA AAGCT AAGGA TGCTG
     1101 AGCAG CCGGG AGGAG AACAT GGACG AGGCC CTATT TGTGG CACAA CTCGC
     1151 AAAGC AGATA AAGGA CACGG CTCTG TATGA CAAGA TGGGG AACCC TAACC
     1201 TGAGG AGTAT ACTGT CTGAA TCCTT CTATG ATATC GCAGG AAGCC TCTTT
     1251 ACTCG CTTAT TCAAT CGCCC CGAGT ATGAT GCGAG GTGCC TGGAG AAATT
     1301 CATAA GGAGC TGTGA AACGA CAGAG ATACA TGTGG AACGG AGATA CATGG
     1351 AAGGC ATAAG AAGAG GGGCC TCATT CAAGG TTCAA AATGT CATGG ACTGG
     1401 GTCGA GGACG ACAGC GCCAA CGCGC TCAGT GAGGT GAACT TCCTC GACAT
     1451 CAGCT GGAAC GATAG AGTGT CAGAG CCCTA TGGCA TCGAA TGCAT TCACG
     1501 GAGAA GGGAG TAGGA TAAGG GTCCC ACTAT CCAGA ATACT CAGGG CCCAC
     1551 GAATT GATTG CAGGG GTCCA AACAG ATGTG GAAAT CAACT TCCCG CGGTA
     1601 TGTGT GCTCC TCTAG AGCCC TCATA CACTT CAGGC AGTAC CTCAT CAAGC
     1651 TTGGG AGGTT CAGTT TCATG GAGAG CAGGG CCATC AAGGA CATCG AAGAC
     1701 ATACA AGCAG GACTC GAAGA GGGGG TGATC ACAGA GGAAG AAGCG GAACT
     1751 GCGCT TATTA CCCAC CACAA AGCCG AAGAT CACAG AAATC CACAT GGATG
     1801 ATGAC ACGCC AGGAA CCAGT GGCGA GAGTG ATGTG GAGAA GTTCA AAAGT
     1851 GTGAG GAGCC TGTGT CGTGA GGAAA TCTAC TCCGA GAAAC TGAAA GGGCG
     1901 CGAGG TGGCG TTCTA TAGCA GGCAC AGCAA GGAGT ACAAA TACAA TGGGG
     1951 GCTCC CATCG TTCCC TCGGT TGGGA TGAGG CTCTA AATGA GCTCA CACAG
     2001 GAACT GGGCC TAGAT GATAG CTACG ATCAT TGTCT GATCC AAAGA TACAC
     2051 GGCAG GGGGC AGTAT AGGAT TCCAC GCAGA TGATG AACCA TGCTA CCTAC
     2101 CAGGG GGTTC CGTCG TGACC GTCAA TTTAC ACGGC GACGC TACTT TTGAG
     2151 GTGAA AGAGA ATCAA TCCGG GAAGA TCGAG AAGAA GGAGC TACAC GACGG
     2201 GGATG TGTAC GTGAT GGGTC CAGGA ATGCA ACAAA CTCAC AAGCA CAGAG
```

2251 TTACC TCTCA CACGG ATGGC CGCTG CAGCA TCACC CTTCG GAACA AAACG
2301 GTAGA TTATG AAGCC AGAAA GGGCG ACGAG GATTC TGAGT ATGAG GAGGA
2351 CAAAG CTGAA CTGGA TGAAG GCATT GACTA CCTAC AGAAA AATCA AGGCA
2401 ACATG TGCTC CCTCA AAGCT TTCGC TGACC ACATG CAGTT GAGCA CCCCA
2451 AGCGT CATAG CCATA GTCAA CGGGG CATCG CCACA AACCC TGAGG GAAAT
2501 AGAGG ATGGC GGGTA CAGTC TGGCT ACACT AGTCA ACCTC TCAAA GGCAC
2551 TCGAT TTTCC AATAG CTATA CACGG CGAGA GGGGG TACGC TGAGA CACCA
2601 GGGTC TTACA GGCGT CTGCA CCTGA AGATC ACATC AGGGC ACGTG GAGCC
2651 TTTCG AGGGC GTGAC AAGCA AAGGG GGCTT TAGGG AGGCC ATGCT GCTTG
2701 GTGAC GGCGT CGGAG TCGGG CATTT CAGGG TGGAC AAAGC CAAGG CCGAC
2751 CGTCT GGCTC AAAGT TTCTA CAATG GAAAC ACGGG GGTTC TGCTC GGCAA
2801 GTACA ATAAA GGCAA GATGC ATACG GGCGA GATAG AGGAG CCAAA AGAGG
2851 TTCTA ACGGC CTTTG GATTT GCAGG ATCAG GGAAA AGTCA CTGGT GCCAA
2901 ACCAT ACTTA AACAC TGTTC CGTGG AGAAA GTGCT TGTGA TTAGC CCAAG
2951 GAAAG TGCTC AGAGA TGATT GGGTA GCAAA AATAT CCAAG AAGCA TAGAG
3001 TGGTA ACTTT CGAGG TGGCA TTCAT GGACG ACTAC GGCTG CAAAG ATATC
3051 GTGAT AGACG AGATC GGTCT TCTAC CACCT GGGTA CATAG ACCTT GTGAT
3101 AGCCG CACAT CAGCC ACGCA CCCTG GTGCT CCTGG GAGAT CCTCT TCAAA
3151 GCACC TACCA CAGTA AACGT GACAA TGTGG TCCTG GAGGC TAGTC AAGAG
3201 GATGT CTTCA ATCGG GTGAG AGGGA AACTA CCATA CTTAT GCTAC TCACA
3251 TAGGT TACCC AGAAA CTGCA AGCTG TTCGA AATAG AATGC ATGGG GGCCG
3301 AGAGC GAGAA GAGGG TGGTG TACCG CTCTA ACAGG CTGAA AGATG AGCCC
3351 ACCAT ATGCG CCACC AGAGC AATGA AGGAA GAGAA GGGGT CCGGG TGGTA
3401 CACCG TCAGT GAGAC GCAAG GTCTG AGCTT CAAAT CATGC CTCAT TTATC
3451 TGGAC GAGCA CTGGG CAAAG AAAGA AGATG AGGAC GTGAT GGTGG CCTTA
3501 ACCCG CTCCA GAGGC GAAAT CGGCA TACAT GTAAC CCCTG CTCTG AAGAA
3551 GAAAC TGATC ACCAA TGCTA AAAGC ACACT GCTGA AGAAA GTACT CAAGG
3601 GCGAA ACCTA CAGAA GATCG GAGAT AGTGG CAATG GTCCG AAAGC ACATA
3651 CCAGA GACCA CAGTG CTGTT TGAGG AAAGC AGACT GGCCG AGACG GTGGA
3701 CTACG AGGCT AGGCT GGCTG GCGAC CCCTA CCTTA AATCC CTTCT GGCCC
3751 TGTAC GACGA GATAG AGATG GAGGA CATAG AGATA GAGGA GCCGG TGACA
3801 CTCGA GCCCA CCAAG ACCCA TTTAG CCCTG AGCAC AAAGA TGAAC GAGCT
3851 TGCAC CATTT GACCT CAAAG CAAAA GAGCA CCGGG AGCAA CACAC AGAGG
3901 CAGGG CGGAC CGAAC AGATA GACGA GAATG GGTAC CAGGG AGAAG TGGGT
3951 GACCC CATGA CCCAC AAGGC GTTGT ACCTG AGGCA TACAT CTGAT GATAC
4001 AGCTA CCTTC ATGAT GTCAG TCAAA AAGAG GTTGC GCTTC AGAAA CTATG
4051 AAGCC AACAG AAGGA AGTAC AAGAC TTGCC ACGGT ATAGG GCACC AAATG
4101 TTCTC GGTGT TCAAG GACAC ATACC AGCTC AAAGA GATTG ACTCA CTGCC
4151 CGAAC TGGAA AGGTG CGAAA TGGAG TTCAT GAAGA AGAGG ATAGA GAAGA
4201 GCACC GGGCT TATCG AAAAG CATGC CGGAA GGAGT GACCC AGACT GGCCG
4251 AGCAA TTACC TCAAA ATATT CCTAA AGCAG CAGAC CTGCA CAAAA ATGGA
4301 AAAGA GGGGG GTGGA TGCTA AGGCG GGACA AACCA TCGCC TGTTT CGCTC
4351 ATTCG GTGCT GTGTA GATTC GGGCC CATTC TGCGT CAAAC TGAGA AAGCG
4401 CTACG GGAGC TCCTG CCCGA GAAGC TCATG ATATA CTCTC AGAAA AAGTA
4451 CATGG ACTTG GATAA ATGGG CTAAG ACGTG GGTAG AGAGC ATGAT GGGGA
4501 CGGAC TCCGA CTACG AGGCA TTCGA CAGAT CACAA GACGA GAAAG TGCTG
4551 GACTT GGAGG TGGAG GTCTT GCGCT TCTTT CTATG GCCCG AAGAT TTAAT
4601 CAGGG AGTAC GAGGA GCTTA AGCTG ATGAT GGGAT GTGCA TTAGG CGACC
4651 TGGCG GTGAT GAGGT TCTCC GGGGA ATTCG GCACC TTCTT CTTCA ACACC
4701 GTGTG CAACA TGGTG TTTAG TTGTA TGCGT TACCA CATAG ACAGG AATAC
4751 CCCGA TGTGC TTTGC GGGGG ATGAT ATGTA TTCGC CAGGC ATTCT GAGAG
4801 TGAAG AAGGA TTACG AAGCC ACTCT CGACC AGTTG ACACT TAAAG CCAAA
4851 GTCCA TATAT CTGAG GAACC CCTCT TCTGT GGGTG GAGGA TGAGC CCATT
4901 CGGAA TAATC AAGGA GCCAA ATCTC ATACT CGACA GGTGG AAGAT AGCGC
4951 TGAGG AGCGG GAATC TATCA CTATG TTTGG TGAAC TATGC GATCG AGGCA
5001 AGCTT TGGGT ACAGG TTAAG TGAGC ACCTG TATGA TGTGA ATATT GATGT
5051 TGATG CACAG CAAGA GCTCG TGAGG GAAAT AGTGA TCAAA AAGCA CTTGC
5101 TACCA AAGAA GATAT CTGAT CTTTT CAGCG AAGAC GAGTG CGAAC GTCAC
5151 AGCGA CGGGG ATGAA GATTT CTTAA GCAAT GACGT CGCGA GATTG TACAG

```
5201 GATTG AGTGA ATTTC TAGGA CATGG TAGTG ATAGC TGTGA TTCAG GTAGT
5251 GGGTC TCTAG GGTCT TTGTC CTATG TACAG TGTGT CAATC TTTTG AGTGA
5301 TCTAA AGAGT TTAGG GTACC AGAGT ATAGA TAGCA TACTA TATAT TTTGG
5351 GTGGT GGCGA GGCGG AACGA TTCGA GATCT ATCGA ATCTT CCGGA GGCAC
5401 GGAAT CGGGA TCGGA GAGGC TTTAC AGCTA GGCGT CAAAA AGAGT TTGTG
5451 CAACA GTCCA AGAAG TCTAC TGGCA ATACT AGATG ACTTG TTGTC TAGAC
5501 TTGGG AGAGG CAGCG CTTTC CTTCC AAGTG ATCTG GGGGC AGTCA AAGGT
5551 GAGCT GGTAG CAACG TTCCA CTCGT CCAGA TTAAG TGTAG ACCTT TACGT
5601 TAATA ACAAG AAGGT TGTCA CGCGA TCAGT GCAAG CGGAA GGTGA TTACA
5651 GCTAT GTCGC AAGAA GGTTC TCTGG GTACA AAGGC CTCAT CCTTC GAGCC
5701 ACAAG ATATT AAGGT CTTTC ACGTC AAAAG GAGCA CCAGA GACCT CGAAA
5751 CTCTC AACAA GAGCC TGCAT CGGGG TGACG TGTAT AACAC AGAGC TCATA
5801 GAGAA GGTTT TTCCA AGGAG AACGA AGAAG TGCGT CATCC ACAAA GACGT
5851 CATCG TTAAG GACGG CCGCG TGGAC TGCGA TCTGG ACATC ATGGA TGAGG
5901 GCCTA GACGA CATAA ATGAA GAGGA ATTCC CACTC TACCA TGTAG GGTGC
5951 ATAGT GGTGG CGCTG ATGCC ACATG GTAAG AACCT TCAAG GCAAG GTGAG
6001 TGTCG AAGTT TTAGA CACCA GGCTG GTAGA CGGGG CATCA AGGAT ATCTA
6051 GGACC CTCAT GGATA TGTCA AAGCC ATTGA GCGCC TGTGC TGACT TCCCG
6101 GGGTA CTTCA TTAGC ACCAG TGATC TGCTG AACGG ATACA CACTG CATCT
6151 CTCGA TAACC ACAAC TGATC TGCAG TTTGT GGACG GAGTA CATCC TTTCA
6201 GTGTT CAACT GATGA GCATA GGGCG ATTCT GCGGT GAAGA CATGA AGACA
6251 AGGTA CGCGA TCACT GAGAC CTCCA AAATG CTGCA CCAGA ACATA TTGAA
6301 CACGG AAGGC GATGG CGAAC TAATC CCAAG GGGCG TGCAG GTTCA GAAAG
6351 TACCG GATAC CCTAG TTATG CCAGA GGTGT TTGAG ACAAT AAAGA AGTTT
6401 GGTTT GAAGA CAAAT GGCAC ACTAC GCCAA GAGGG TAGAG ATAAG GGCGA
6451 TAATA GAAGA GTTGG TGTTG GCGAA AGCCC AACCA ACTGA CGACG CTTCT
6501 GAGAG CGGCT ACGAC CGAAA TATGT ACCTG AACAC TCTCT TCGGG TACAT
6551 CGCCT TGGTC GGGAC AAGCA AAAAG GCGAT TCATT ATGGG GAGGT AGATA
6601 TAGTA GGTCC TAAAG CTAGC AAGAA GACGG GGATC GACCC CAGGG GGAAG
6651 ATGGT GGTGT CGGAG CTGGT CGGCA GGATG CGCAC GCTGA GTGTG GCAGT
6701 AAGCG AGGGA CCCGT CAAGG GCGCA ACTCT GAGGC AGATG TGCGA GCCAT
6751 TCGCA CAGAA CGCCT ACGAT TTCCT CGTAG TGATG GCTGA AATGG GCACA
6801 TACTC ACAGT TAGCC ACCAA GATGA CTAGG TCAGG CTTTA AAGAG CCACA
6851 GGTTA TGTTC GACTT CGCGT CGGGC TTAGA TCTGA AAGCA CTGAC ACTAC
6901 AAGAA GCTAC TGTGA TACAG GCTAT GCACT CTCGC CTCTT TCGTA CGGAA
6951 GGCGC GAAGG GAGTA TTCAA CGCTC AGTCA TCAAT CGGCG AGCAG GCTGT
7001 TGAGA TATAG ATGGA TGACC CATCA TTTCT CACGG GTAGG TCCAC GTTTG
7051 CTAAG CGTAG GCGC~ ~TAGG CGCAT GAATG TGTGT AAGTG TGGTG CAATA
7101 ATGCA CAATA ATAAG GATTG CAAAT CCAGC AGTAT ATCCA GTCAT AAACT
7151 CGACA GATTA CGGTT CGTGA AAGAG GGAAG AGTAG CCTTA ACAGG CGAGA
7201 CTCCT GTTTA CCGAA CTTGG GTCAA ATGGG TTGAG ACTGA GTATC ATATA
7251 TATAT AGTAG AAACC TCAGA CGATG AGGAT TAATC GTCTA TCCTA GCTAA
7301 AAAGT AAAGT ATAAC TTAAT AATAA AGCGA GCTAG GTTGT CACAC GAAGA
7351 C
//
```

GVA–5′–620

```
   1 GAATA TTTAA CTTGA TTCCC ATCGA TTATA AGTGA ACAAA CGTTG CCAGC
  51 ACCCA CGCTC TGCGA AGCAC CTGCT AATTC ACTAC CATGT CGATA TCAGT
 101 ATCCT CCCAA CGTGT CGCAG TCTCC AACCT GTACA CGAAC GGATC TGAGG
 151 AGTCA GTTAA AGCAA TAAAG GAGTT AAAAA GCAAA CGGTT ATTGG AAACC
 201 GAAAC CAGGC TAGAT GGACT ATTTG ATTAC TACAT CCCAG ATACC TTAAG
 251 AGAAA TTCTT ACAGG TTATG GTATG GAGCT CAGTG TCCAT TCTTT CCAAG
 301 GACAT GCTCA CCCCG TAAGT AAGAT GATAG AAAAC CACAT GTTAT ATAGA
 351 GTAGC ACCAA ATTAT TTCTC TAGTA ACACA TTGGT AGTTA GTTGT AAAGA
 401 GAGCA AAATA AAGCG TCTAC GTCTG AAAAA TGCAA ACAAC AGGAA TCTGA
 451 ACTTC ACACA GTACA ACAGA CTGGT GCATG CAAAG GATCA CCACA GATAT
 501 GAAAA TGCAT TCAGA GAACT AGACG TCGGG AATTT GACAA ATCTC ATAAA
```

```
      551 CAAAG AGGAC CAGAG TGAAT GCATA TTCAT ACATG ATGAA GTTCA ATACT
      601 GGAGT CTAGA CGAGA TGCAA
//
```

GVA-924_

```
      251 ~~~~~ ~~~~~ ~~~GG CTATG GTATG GAGTT CAGTG TCCAT TCTTT CCAGG
      301 GACAT GCTCA CCCAG TAAGT AAAAT GATAG AAAAC CACAT GTTGT ACAGA
      351 ATAGC ACCCA ATTAT TTTTC TAGTA ACACA TTGAT AGTTA GTTGC AAAGA
      401 GAGCA AGATT AAGCG TCTAC GTCTC AAAAA TGCGG GCAAT AGGAA TCTAA
      451 ACTTC ACACA GTACA ACAGA CTGGT GCATG CAAAG GATCA CCACA GGTAC
      501 GAAAA TGCAT TTAGA GAGCT CGATG TTGGA AATCT GACAA ATCTG ATCAA
      551 CAAGG AAGAC AAGAG CGAAT GTATA TTCAT ACATG ATGAG GTCCA ATACT
      601 GGAGT TTAGA TGAAA TGCAA AGGTT CCTAG GTAGT TTGTC CAAAG TGGAT
      651 AGGGT AGTGT ATAGC ATCAT ATACC CATCA GAGGT AGAAG CAGGA TATTC
      701 GCAAA GCCTG TTCCC AGAGG CATAC ACGTT CGACC TGAAG GATGG GAGGT
      751 TAGTG TGGTA CCCAG ATGGG AAGGC GGAAG GGGCC TACAC GCAAC CTGTG
      801 AACCC TTGGC TGCTG AGATG CTCAA AGACA GAGGA TTCGA AGGGA CGCCC
      851 ATGGA CGATC ACAAA GCTCC AAACA ATCGG GGCTC ATCAT CTCTT CAGCA
      901 CCATA AAGGG TAGTT ATCTA ACGGA GGAAT CGTAC AAGTA CGATA ACTTC
      951 ACGAT CATAA ACCCT AATGA TATCC TGAAG GGAAG GAGAG GCGGG AAACC
     1001 ACTCT ACTTG AGAGC GCGCA TGATC AAGCC CACCC TACTG TACCT TCTGG
     1051 CTCTG AAGGA AAGCG ACTCA AATTC CGCTG TCGCA AAGTT GAGGA TGCTT
     1101 AGCAG TAGAG AGGAA AACAT GGATG AGGCC CTCTT TGTGG CACAG CTCGC
     1151 CAAGC AGATC AAGGA GACGG CTCTG TATGA CAAGA TGG
//
```

GVA-532

```
      901 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~CGGA GGAAT CGTAC AAGTA CGATA ACTTC
      951 ACGAT CATAA ACCCA AACGA TATTC TGCGA GGGAG AAGAG GCGGG AAGCC
     1001 ACTCT ATTTG AGAGC ACGTA TGATC AAGCC CACCC TACTG TATCT CTTGG
     1051 CCCTG AAGAA AAGCG ACTCA AATTC TGCCG TCGCA AAATT GAGAA TGTTA
     1101 AGCAG CAGGG AAGAG AACAT GGATG AGGCC CTCTT CGTGG CGCAA CTGGC
     1151 AAAGC AAATC AAGGA TACGG CTCTG TATGA CAAAA TGGGA AGCCC AAACC
     1201 TAAGG AGCAT ACTGT CGGAA TCATT CTACG ACATA GCCGG AAGCC TCTTC
     1251 ACTCG GTTGT TCAAC AAACC GGAGT ACGAT GCCAG GTGCC TTGAA AAATT
     1301 CATAC GGAGC TGCGA CACCA CGGAA ATACA TGTGG AGCGG AAATA TATGG
     1351 AGGGC ATAAG AAGGG GGGCA ACGTT CAAAG TGCAA AACGT CATAG ATTGG
     1401 GTGGA AGACG ACAGT GCTAA CACTC TCAGC GAGGT GAACT TCCTC GACAT
     1451 CAG
//
```

GVA-759

```
     1401 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~GT GAACT TCCTC GACAT
     1451 CAGCT GGAAT GATAG AGTGT CTGAG CCCTA TGGCA CTGAG TGCAT CCACA
     1501 GAGAA GGAAT CAGGA CGAGG GTCCC ATTGT CCAGA ATACT CAGGG CCCAC
     1551 GAACT GATTG CAGGG GTCCA AACGG ATGTG GAACT CAACC TTACA CGGTA
     1601 TGTAT GCTCC TCCAG GGCCC TCGTA CACTT CAGGC AGTAC CTCAT CAAGC
     1651 TCGGG AGGTT CAGTT TCATG GAGAG CAGGG CTATC AAGGA CATTG AGGAC
     1701 ATACA ATCAG GGCTC GAAGA GGGGG TGATT ACAGA GGAAG AAGCG GAACT
     1751 GCGCT TACTA CCCAC CATAA AGCTG GAGAC CACAG GAATC CATAA GGATG
     1801 AGAGC ATGCC GGGAA CCAGT GGCGA GAGTG ACATA GAAAA ATTCA AGAGC
     1851 GTGAG GAGCC TGTGC CGCGA GGAAA TCTAC TCTGA GAAAT TGAAA GGGCG
     1901 CGAGG TGGCG TTCTA TAGCA GGCAC GGCAA GGAGT ATAAG TATAA TGGAG
     1951 GCTCC CATCG TTCCC TCGGC TGGGA CAAGG CCTTA GATGA GCTCA CACAA
     2001 GAACT GGGCC TAGAC GATAG CTACG ATCAT TGCCT GATCC AGAGA TACAC
     2051 AGCAG GGGGC AGTAT AGGAT TCCAC GCAGA TGATG AACCA TGCTA CCTCC
```

```
     2101 CAGGG GGCTC CGTCG TAACT GTCAA TTTAT ACGGT GAAGC CACCT TTGAG
     2151 GTAAA AGAGA ATCAA TCCGG GAAGA TCGAG AAGAA GGAGC TAC
//
```

GVA–965

```
     2001 ~~~~~ ~GGCC TAGAT GATAG CTACG ATCAC TGCCT AATCC AAAGG TACAC
     2051 AGCAG GGGGA AGCAT TGGGT ACCAT GCAGA TGATG AACCT TGCTA TCTAC
     2101 CTGGG GGCTC CGTCG TAACC GTCAA CCTGT TCGGG GATGC CACAT TTGAG
     2151 GTCAA GGAGA ACAAA TCCGG AAAGG TCGAC AAAAT AGAGA TGAGT GATGG
     2201 GGATG TATAT GTGAT GGGGC CGGGG ATGCA ACAAA CCCAC AAGCA TAGAG
     2251 TCACC TCCCA TACTG ATGGA CGCAG CAGCA TCACC CTCCG GAATA AAACG
     2301 ATCGA CTACG AAAGT AGAAA GGGTG ACGAT GACTC TGAAT ATGAG GAGGA
     2351 TAAGG CCGAA CTGGA CGAGG GCATT GAGTA CCTAC AGAAA AATCA GGCTA
     2401 ACATG TGCTC CCTAA AGGCA TTCGC GGATC ACATG CAGTT AAGCA TCCCA
     2451 AGCGT GATAT CCATA GTCAA TGGAG CATCC CCTCA AACTC TGAGG GAGAT
     2501 AGAGG ATGGC GGGAT CAGTC TAGCC ACACT CGTCA ATCTT TCCAA GGCTC
     2551 TAGAC TTCCC AATCG CCATA CACGG AGAGA GAGGG TACGC TGAAA CACCG
     2601 GGATC GTACA GGCGC CTGCA CCTAA AGATC ACCTC AGGGC ATGTG GAGCC
     2651 ATTCG AAGGG GTGAC AAGCA AGGGT GGTTT CAGAG AGGCT ATGTT ACTTG
     2701 GTGAT GGTGT TGGCC TTGGG CATTT CAGGG TGGAC AAAAG CAAAG CGGAC
     2751 CGGCT GGCCC AAAGC TTTTA CAACG GGAAC ACTGG GGTGC TACTC GGCAA
     2801 GTACA ACAAA GGGAA GATGC ACACG GGAGA CATAG AAGGG CCAAA GGAGG
     2851 TTCTA ACGGC TTTTG GGTTC GCCGG ATCAG GGAAA AGTCA CTGGT GTCAA
     2901 ACCAT ACTCA AACAC TGTTC CGTAG AAAAA GTGCT TGTGA TCAGC CCCAG
     2951 GAAAG TGCTC AGAGA TGATT GG
//
```

GVA–1002

```
     2751 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~CTC GGCAA
     2801 GTACA ATAAA GGCAA GATGC ATACG GGCGA GATAG AGGGG CCAAA AGAGG
     2851 TTCTA ACGGC CTTTG GATTC GCAGG ATCAG GGAAA AGTCA CTGGT GCCAA
     2901 ACCAT ACTTA AACAC TGTTC CGTGG AGAAA GTGCT CGTGA TTAGC CCTAG
     2951 GAAGG TGCTC AAAGA TGACT GGGTA GCAAA AATAT CCAAG AAGCA TAAAG
     3001 TGGTG ACCTT CGAAG TGGCA TTCAT GGATG ACTAC GGCTG CAAAG ACATC
     3051 GTGAT AGACG AGATT GGTCT CCTAC CACCT GGGTA CATAG ACCTT GTGAT
     3101 AGCCG CGCAT CAGCC ACGCA CCCTG GTGCT CCTGG GAGAC CCTCT TCAAA
     3151 GCACC TACCA TAGTA AACGT GACAA TGTGG TGCTG GAGGC TAGTC AAGAG
     3201 GACGT CTTCA ACAGG GTGAA AGGGA GACTA CCATA CTTAT GCTAT TCACA
     3251 TAGGC TACCC AGAAA CTGCA AGCTG TTCGA GATAG AATGC ATGGG GGCCG
     3301 AGAGC GAGAA GAGGG TGGTG TACCG CTCTA ACAGG CTGAA AGATG AGCCT
     3351 ACCAT ATGCG CCACG AGAGC AATGA AGGAA GAGAA AGGGT CCGGC TGGTA
     3401 CACGG TCAGC GAGAC ACAAG GTCTG AGCTT CAAAA CGTGC CTCAT TTACC
     3451 TAGAC GAACA CTGGG CAAAG AAAGA AGATG AGGAT GTGAT GGTGG CCTTG
     3501 ACTCG CTCCA GAGGT GAGAT AGGCA TACAT GTAAC CCCTG CTCTC AAGAA
     3551 GAATC TGATC GCCAA TGCTA AAAGC ACACT GCTGA AGAAA GTGCT TAAGG
     3601 GCGAG ACCTA CAGAA GGTCA GAGAT AGTGG CAATG GTCCG AAAGC ACATA
     3651 CCAGA GACCA CAGTG CTGTT TGAGG AGAGC AGACT AGCCG AGACA GTGGA
     3701 TTACG AGGCT AGGCT GGCCG GCGAT CCCTC CCTAA AATCC CTTCT GGCCC
     3751 TGTAC GACGA GATTG AGATG GAGGA CATAG AGATA GAGGA GCCG-
//
```

GVA–1005

```
     3351 ~~~~~ ~~~~~ ~~~~~ ~GAGC AATGA AGGAA GAGAA GGGGT CCGGC TGGTA
     3401 CACGG TCAGC GAGAC ACAAG GTCTG AGCTT CAAAA CGTGC CTCAT TTACC
     3451 TAGAC GAACA CTGGG CAAAG AAAGA AGATG AGGAT GTGAT GGTGG CCTTG
     3501 ACTCG CTCCA GAGGT GAGAT AGGCA TACAT GTAAC CCCTG CTCTC AAGAA
```

```
     3551 GAATC TGATC GCCAA TGCTA AAAGC ACACT GCTGA AGAAA GTGCT TAAGG
     3601 GCGAG ACCTA CAGAA GGTCA GAGAT AGTGG CAATG GTCCG AAAGC ACATA
     3651 CCAGA GACCA CAGTG CTGTT TGAGG AGAGC AGACT AGCCG AGACA GTGGA
     3701 TTACG AGGCT AGGCT GGCCG GCGAT CCCTA CCTAA AATCC CTTCT GGCCC
     3751 TGTAC GACGA GATTG AGATG GAGGA CATAG AGATA GAGGA GCCAG TGGTA
     3801 CTCGA ACCTA CCAGG ACCCA TCTAG CCCTG AGCAC AAAGA TGAAT GAACT
     3851 TGCAC CATTT GACCT CAAAG CCAAA GAGCA TCGGG AGCAA CACAC AGAGG
     3901 CAGGG CGAAC CGAAC AGATA GATGA GGATG GGTAC CAGGG TgAAG TGGGC
     3951 GACCC CATGA CCCAC AAGGC GTTGT ACCTG AGGCA CACAT CTGAT GATAC
     4001 AGCTA CCTTC ATGAT GTCAG TCAAA AAGAG GTTGC GCTTC AGGAA CTATG
     4051 AAGCC AACAG GAGAA AGTAC AAGGC TTGCC ACGGT ATAGG GCATC AAATG
     4101 TTCTC GGTGT TCAAG GACAC ATACC AGCTC AAAGA GATTG ACTCA CTGCC
     4151 CGAAC TGGAA AGGTG CGAAA TGGAG TTCAT GAAGA AGAGA ATAGA GAAGA
     4201 GTACC GGGCT TATCG AGAAG CATGC CGGAA GGAGT GATCC TGACT GGCCA
     4251 AGTAA CTACC TCAAG ATATT TCTGA AGCAG CAGAC CTGCA CAAAA ATGGA
     4301 AAAGA GAGGG GTCGA TGCTA AGGCG GGGCA GACCA TCGCT TGTTT TGCTC
     4351 ATTCG GTGCT GTGTA GATTC ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~
//
```

GVA-951

```
     4101 ~~~~~ ~~~~T TCAAG GACAC ATACC AGCTC AAAGA GATTG ACTCA CTGCC
     4151 CGAAC TGGAA AGGTG CGAAA TGGAG TTCAT GAAGA AGAGA ATAGA GAAGA
     4201 GTACC GGGCT TATCG AGAAG CATGC CGGAA GGAGT GATCC TGACT GGCCA
     4251 AGTAG CTACC TCAAG ATATT TCTGA AGCAG CAGAC CTGCA CAAAA ATGGA
     4301 AAAGA GAGGG GTTGA TGCTA AGGCG GGGCA GACCA TCGCT TGTTT TGCTC
     4351 ATTCG GTGCT GTGTA GATTC GGGCC TATTC TGCGT CAAAC TGAGA AGGCA
     4401 CTAAG GGAGC TCCTA CCTGA GAACG TAATG ATATA CTCTC AGAAA AACTA
     4451 CATGG ACTTG GACAA ATGGG CGAAG ACATG GGTCG AAAGC ATGAT GGGAA
     4501 CGGAC TCCGA CTATG AAGCA TTCGA CAGAT CACAA GACGA AAAAG TGCTG
     4551 GACTT GGAGG TAGAG GTCTT GCGCT TCTTC CTGTG GCCTG AAGAT TTAAT
     4601 CAGAG AGTAC GAGGA GCTCA AGCTT ATGAT GGGAT GCGCA TTAGG TGACC
     4651 TGGCG GTGAT GAGGT TCTCC GGGGA ATTCG GCACC TTCTT CTTCA ATACC
     4701 GTGTG CAACA TGGTG TTTAG CTGTA TGCGT TACCA CATGG ACCGG AACAC
     4751 CCCGA TGTGC TTTGC GGGGG ATGAT ATGTA TTCGC CAGGC ATTCT GAGGG
     4801 TAAAG AAGGA TTACG AAGCC ACTCT CGACC AGCTG ACACT CAAGG CGAAG
     4851 GTCCA CATAT CTGAG GAACC CCTCT TCTGT GGGTG GAGGA TGAGC CCATT
     4901 CGGTA TAATC AAGGA GCCAA ACCTC ATACT CGACA GGTGG AAGAT AGCGC
     4951 TGAGG AGCGG GAATC TATCA CTTTG CCTAG TGAAC TACGC AATAG AGGCA
     5001 AGCTT TGGGT ACAGG TTAAG TGAGC ACCTG TATGA TGTGA ATATT GATGT
     5051 TGATG CACAG C~
//
```

GVA-646D

```
     4801 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~CTCT CGACC AGTTG ACACT CAAGG CCAAG
     4851 GTACA TGTGT CTGAA GAACC CCTCT TTTGC GGATG GAGGA TGAGC CCCTT
     4901 CGGGA TTATC AAAGA GCCAA ACCTC ATACT GGACA GATGG AAGAT GGCGC
     4951 TGCGG AGCGG GAACC TATCA CTATG CCTTG TGAAC TACGC CATAG AGGCA
     5001 AGTTT TGGGT ACAGG TTAAG TGAGC ACCTG TATGA TGTGA ATATA GACGT
     5051 TGATG CCCAG CAAGA GCTTG TGAGG GAAAT AGTGA TCAAG AAGCA CCTAC
     5101 TGCCA AAGAA GATAT CTGAT CTTTT CAGCG AGGAC GAGTG CGAAG CTCAC
     5151 AGCGA CGGGG ATGAT GACTT CTTGA GCAAT GACGT CGCAG GATTG TACAG
     5201 GATTG AGTGA ATTCT TAGGG TGTGG TAGTG ATAGC CGTAG TCTAG GTGTA
     5251 GGGTC TTTGG AGTCT TTGTC TTATG TACAG TGCAT CCAAC TTTTG AGTGA
     5301 TCTAA AGAGC CTAGg ATATC ATAGC ATAGA TAGTA TACTA TATAT ACTGG
     5351 GTGGT GGTGA GGCGG AACGA TTTGA GATCT ATCGA ATCTT CCGGA GGCAC
     5401 GGAAT CGGGA TTGGA GAGGC TTTAC AGCTA GGCGT CAAAA AGAGT TTGTG
     5451 CAACA GTCCA AGAAG TC
```

//

GVA-908

```
5201 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~AGC TGTGA TTCAG GTAGT
5251 GGGTC TCTAG GGTCT TTGTC TTATG TACAG TGTGT TAATC TTTTG AGTGA
5301 TCTAA AGAGT TTAGG GTATC AGAGT ATAGA TAGCA TACTA TATAT CTTGG
5351 GTGGT GGTGA GGCGG AACGA TTCGA GATCT ATCGA GTCTT CCGGA GGCAC
5401 GGAAT CGGGA TCGGA GAGGC TTTAC AGCTA GGCGT CAAGA AGAGT TTGTG
5451 TAACA GTCCG AGAAG CCTGC TGGCA ATACT AGACG ACTTG TTGTC TAGAC
5501 TCGGG AGAGG CAGCG CTTTC CTTCC AAGTG ATCTG GGAGC AGTCA AAGGT
5551 GAGCT ACTAG TGACA TTCCA CTCGT CCAGA TTAAG TGTAG ACCTT TACGT
5601 TAATA ACAAG AAGGT CGTCA CGCGA TCGGT GCAAG CGGAA GGTGA TTACA
5651 GCTAT GTCGC AAGAA GGTTC TCTGG GTACA AAGGC CTCGT CCTCC GAGCT
5701 ACAAG ACATT AAGGT CTTTA ACGTC AAAAG GAGCA CCAAG GACCT CGAAA
5751 CTCTC AACAA GACTC TGCAC AGGGG TGACG TGTAC GACAC AGAGC TCATA
5801 GAGAA GGTCT TCCCA AGAAG AACAA AGAAG TGCGT CATCC ACAAG GAGCT
5851 CATTG TCAAG GACGG CCGCG TGGAC TGCGA TCTGG ACATC ATGGA TGAGG
5901 GCCTG GACGA CATAG ATGAA GAGGA ATTCC CACTC TACCA CGTAG GGTGC
5951 ATAGT GGTGG CGCTG ATGCC ACGTG GTAAG AACCT TCAAG GCAAG GTGAG
6001 CGTCG AAGTT TTAGA CACCA GGCTG GTAAA CGGGG CGTCA AGGAT ATCTA
6051 GGACC CTTAT GGATA TGTCA AAGCC ACTGA GCGCC TGTGC TGATT TCCCA
6101 GGGTA CTTCA TAAGC ACCAG TGATC TGCTG AACGG ATACA C~
```

//

GVA-1100

```
5851 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~A TCTGG ACATC ATGGA TGAGG
5901 GCCTG GACAA CATAG ATGAA GAGGA ATTCC CACTC TACCA CGTAG GGTGC
5951 ATAGT GCTGG CACTG ATGCC ACATG GTAAG AACCT CCAAG GCAAG GTGAG
6001 TGTCG AAGTC CTAGA CACCA GGCTG GTAAA CGGGG CGTCA AGGAT ATCCA
6051 GAACC CTCAT GGATA TGTCA AAGCC GTTAA GCGCC TGTGC TGATT TCCCG
6101 GGGTA CTTCA TTAGT ACCAG TGATC TGCTG AATGG GTACA CATTG CACCT
6151 CTCAA TAACC ACGAC TGATC TGCAG TTTGT GGACG GCGTA CACCC CTTCA
6201 GCGTC CAATT GATGA GCATA GGGCG GTTCT GCGGC GAAGA TATGA AGACC
6251 AGGTA TGCAA TCACT GAGGC CTCCA AAGTG CTGCA CCAAA ACATA CTGAA
6301 TGCAG AAGGC GATGG CGAAC TAATC CCAAG GGGCG TGCAG GTCCA GAAAG
6351 TACCG GATAC CCTAG TTATG CCAGA GGTGT TTGAG ACAAT AAAGA AGCTT
6401 GGTTT GAAGA CAAAT GGCAC ACTAC GCCAA GAGGG TCGAG ATAAG GGCGA
6451 TAATA GAAGA GTTGG TGCTG GCGAA AGCCC AGCCA ACTGA CGACG CTTCT
6501 GAGAG CGGCT ACGAC CGAAA TATGT GCCTG AATAC TCTCT TTGGG TACAT
6551 CGCCT TGGTC GGGAC AAGCA AAAAG GCGGT TCATT ATGGG GAGGT AGATA
6601 TAATA GGCCC TAAAG CTAGC AAAAA GACAG GGATA GACCC AAGGG GGAAG
6651 ATGGT CATAT CAGAG CTGGT CGGCA GGATG CGCAC GCTGA GTGTG GCAGT
6701 AAGCG AGGGG CCCGT CAAGG GCGCG ACTCT TAGGC AGATG TGCGA GCCAT
6751 TCGCA CAGAA CGCCT ACGAT TTCCT CGTAG TGATG GCTGA AATGG GCACA
6801 TACTC ACAGT TAGCT ACCAA GATGA CTAGG TCAGG CTTTA AAGAG CCACA
6851 GGTTA TGTTC GACTT CGCGT CGGGC TTAGA TCTGA AAGCA CTGAC ACTAC
6901 AAGAA GCTAC TGTGA TACAG GCTAT GCACT CTCGC CTCTT TCGTA CGGAA
6951 GGTGC AAAGG GAGTA TTCAA CGCTC AGTC~ ~~
```

//

GVA-501

```
6801 ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~~~~ ~~ACA
6851 GGTTA TGTTC GACTT CGCGT CGGGC TTAGA TCTGA AAGCG CTGAC ATTGC
6901 AAGAA GCTAC TGTGA TACAG GCTAT GCACT CTCGC CTCTT TCGTA CGGAA
6951 GGCGC GAAGG GGGTG TTCAA CGCCC AGTCA TCAGT TGGCG AACAG GCTGT
7001 CGAAA TATAG ATGGA TGACC CATCG TATCT CTCGG GTAGG TCCAC GTATG
```

```
   7051 CTAAG CGTAG GCGCG CTAGG CGCAT GAATG TGTGT AAGTG TGGTG CTATA
   7101 ATGCA CAATA ATAAG GATTG TAAAT CCAGT AGTAT CTCTG GTCAC AAACT
   7151 TGACA GATTA CGGTT CGTGA AAGAG GGAAG AGTAG CCTTA ACAGG CGAGA
   7201 CTCCT GTTTA CCAAA CTTGG ATCGA ATGGG TTCAG GCCGA GTATC ATATA
   7251 TATAT ATTAG AAACC TCAGA CGATG AGGAT TAATC GTCTA TCCTA GCTAA
   7301 GGAGT AAAGT ATAAC TTAAT AATAA AGCAA GCTAG GTTGT CACAC GAAGA
   7351 C
//
```

pSK-GVA-mutagen-3 (containing mutation in CP)

```
      1 CTAAA TTGTA AGCGT TAATA TTTTG TTAAA ATTCG CGTTA AATTT TTGTT
     51 AAATC AGCTC ATTTT TTAAC CAATA GGCCG AAATC GGCAA AATCC CTTAT
    101 AAATC AAAAG AATAG ACCGA GATAG GGTTG AGTGT TGTTC CAGTT TGGAA
    151 CAAGA GTCCA CTATT AAAGA ACGTG GACTC CAACG TCAAA GGGCG AAAAA
    201 CCGTC TATCA GGGCG ATGGC CCACT ACGTG AACCA TCACC CTAAT CAAGT
    251 TTTTT GGGGT CGAGG TGCCG TAAAG CACTA AATCG GAACC CTAAA GGGAG
    301 CCCCC GATTT AGAGC TTGAC GGGGA AAGCC GGCGA ACGTG GCGAG AAAGG
    351 AAGGG AAGAA AGCGA AAGGA GCGGG CGCTA GGGCG CTGGC AAGTG TAGCG
    401 GTCAC GCTGC GCGTA ACCAC CACAC CCGCC GCGCT TAATG CGCCG CTACA
    451 GGGCG CGTCC CATTC GCCAT TCAGG CTGCG CAACT GTTGG GAAGG GCGAT
    501 CGGTG CGGGC CTCTT CGCTA TTACG CCAGC TGGCG AAAGG GGGAT GTGCT
    551 GCAAG GCGAT TAAGT TGGGT AACGC CAGGG TTTTC CCAGT CACGA CGTTG
    601 TAAAA CGACG GCCAG TGAGC GCGCG TAATA CGACT CACTA TAGGG CGAAT
    651 TGGGT ACCGG GCCCC CCCTC GAGGT CGACC TGCAG GCGGC CGCGA ATTCA
    701 CTAGT GATTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTG
    751 TCTTC GTGTG ACAAC CTAGC TTGCT TTATT ATTAA GTTGT ACTTT ACTCC
    801 TTAGC TAGGA TAGAC GATTA ATCCT CATCG TCTGA GGTTT CTAAT ATATA
    851 TATAT GGTAC TCGGC CTGAA CCCAT TCGAT CCAAG TTTGG TAAAC AGGAG
    901 TCTCG CCTGT TAAGG CTACT CTTCC CTCTT TCACG AACCG TAATC TGTCA
    951 AGTTT GTGAC CAGAG ATACT ACTGG ATTTA CAATC CTTAT TATTG TGCAT
   1001 TATAG CACCA CACTT ACACA CATTC ATGCG CCTAG CGCGC CTACG CTTAG
   1051 CATAC GTGGA CCTAC CCGAG AGATA CGATG GGTCA TCCAT CTATA TTTCG
   1101 ACAGC CTGTT CGCCA ACTGA TGACT GAGCG TTGAA TACTC CTTTG CACCT
   1151 TCCGT ACGAA AGAGG CGAGA GTGCA TAGCC TGTAT CACAG TAGCT TCTTG
   1201 TAGTG TCAGT GCTTT CAGAT CTAAG CCCGA CGCGA AGTCG AACAT AACCT
   1251 GTGGC TCTTT AAAGC CTGAC CTAGT CATCT TGGTA GCTAA CTGTG AGTAT
   1301 GCGCC TATTT CAGCC ATCAC TACGA GGAAA TCGTA GGCGT TCTGT GCGAA
   1351 TGGCT CGCAC ATCTG CCTAA GAGTC GCGCC CTTGA CGGGC CCCTC GCTTA
   1401 CTGTC ACACT CAGCG TGCGC ATCCT GCCGA CCAGC TCTGA TATGA CCATC
   1451 TTCCC CCTTG GGTCT ATCCC TGTCT TTTTG CTAGC TTTAA GGCCT ATTAT
   1501 ATCTA CCTCC CCATA ATGAA CCGCC TTTTT GCTTG TCCCG ACCAA GGCGA
   1551 TGTAC CTAAA GAGAG TATTC AGGCA CATAT TTCGG TCGTA GCCGC TCTCA
   1601 GAAGC GTCGT CAGTT GGCTG GGCTT TCGCC AGCAC CAACT CTTCT ATTAT
   1651 CGCCC TTATC TCGAC CCTCT TGGCG TAGTG TGCCA TTTGT CTTCA AACCA
   1701 AGCTT CTTTA TTGTC TCAAA CACCT CTGGC ATAAC TAGGG TATCC GGTAC
   1751 TTTCT GGACC TGCAC GCCCC TTGGG GTTAG TTCGC CATCG CTTTC TGCAT
   1801 TCAGT ATGTT TTGGT GCAGC ACTTT GGAGG CCTCA GTGAT TGCAT ACCTG
   1851 GTCTT CATAT CTTCG CCGCA GAACC GCCCT AAGCT CATCA ATTGG ACGCT
   1901 GAAGG GGTGT ACGCC GTCCA CAAAC TGCAG ATCAG TCGTG GTTAT TGAGA
   1951 GGTGC AATGT GTACC CATTC AGCAG ATCAC TGGTA CTAAT GAAGT ACCCC
   2001 GGGAA ATCAG CACAG GCGCT TAACG GCTTT GACAT ATCCA TGAGG GTTCT
   2051 GGATA TCCTT GACGC CCCGT TTACC AGCCT GGTGT CTAAA ACTTC GACGC
   2101 TCACC TTGCC TTGAA GGTTC TTGCC ACGTG GCATC AGCGC CACCA CTATG
   2151 CACCC TACGT GGTAG TGTGG GAATT CCTCT TCATC TATGT CGTCC AGGCC
   2201 CTCAT CCATG ATGTC CAGAT CGCAG TCCAC GCGGC CGTCC TCGAC AATGA
   2251 GCTCC TTGTG GATGA CGCCC TTCTT TGTTC TTCTT GGGAA GACCT TCTCT
   2301 ATGAG CTCTG TGTCG TACAC GTCAC CCCTG TGCAG AGTCT TGCTG AGAGT
   2351 TTCGA GGTCC TTGGT GCTCC TTTTG ACGTT AAAGA CCTTA ATGTC TTGTA
```

2401 GCTCG GAGGA CGAGG CCTTT GTACC CAGAG AACCT TCTTG CGACA TAGCT
2451 GTAAT CACCT TCCGC TTGCA CCGAT CGCGT GACGA CCTTC TTGTT ATTAA
2501 CGTAA AGGTC TACGC TTAAT CTGGA CGAGT GGAAT GTCAC TAGTA GCTCA
2551 CCTTT GACTG CTCCC AGATC ACTTG GAAGG AAAGC GCTGC CTCTC CCGAG
2601 TCTAG ACAAC AAGTC GTCTA GTATT GCCAG CAGAC TTCTT GGACT GTTGC
2651 ACAAA CTCTT TTTGA CGCCT AGCTG TAAAG CCTCT CCAAT CCCGA TTCCG
2701 TGCCT CCGGA AGATT CGATA GATCT CAAAT CGTTC CGCCT CACCA CCACC
2751 CAGTA TATAT AGTAT ACTAT CTATG CTATG ATATC CTAGG CTCTT TAGAT
2801 CACTC AAAAG TTGGA TGCAC TGTAC ATAAG ACAAA GACTC CAAAG ACCCT
2851 ACACC TAGAC TACGG CTATC ACTAC CACAC CCTAA GAATT CACTC AATCC
2901 TGTAC AATCC TGCGA CGTCA TTGCT CAAGA AGTCA TCATC CCCGT CGCTG
2951 TGAGC TTCGC ACTCG TCCTC GCTGA AAAGA TCAGA TATCT TCTTT GGCAG
3001 TAGGT GCTTC TTGAT CACTA TTTCC CTCAC AAGCT CTTGC TGTGC ATCAA
3051 CATCA ATATT CACAT CATAC AGGTG CTCAC TTAAC CTGTA CCCAA AGCTT
3101 GCCTC TATTG CGTAG TTCAC TAGGC AAAGT GATAG ATTCC CGCTC CTCAG
3151 CGCTA TCTTC CACCT GTCGA GTATG AGGTT TGGCT CCTTG ATTAT ACCGA
3201 ATGGG CTCAT CCTCC ACCCA CAGAA GAGGG GTTCC TCAGA TATGT GGACC
3251 TTCGC CTTGA GTGTC AGCTG GTCGA GAGTG GCTTC GTAAT CCTTC TTTAC
3301 CCTCA GAATG CCTGG CGAAT ACATA TCATC CCCCG CAAAG CACAT CGGGG
3351 TGTTC CGGTC CATGT GGTAA CGCAT ACAGC TAAAC ACCAT GTTGC ACACG
3401 GTATT GAAGA AGAAG GTGCC GAATT CCCCG GAGAA CCTCA TCACC GCCAG
3451 GTCAC CTAAT GCGCA TCCCA TCATA AGCTT GAGCT CCTCG TACTC TCTGA
3501 TTAAA TCTTC AGGCC ACAGG AAGAA GCGCA AGACC TCTAC CTCCA AGTCC
3551 AGCAC TTTTT CGTCT TGTGA TCTGT CGAAT GCTTC ATAGT CGGAG TCCGT
3601 TCCCA TCATG CTTTC GACCC ATGTC TTCGC CCATT TGTCC AAGTC CATGT
3651 AGTTT TTCTG AGAGT ATATC ATTAC GTTCT CAGGT AGGAG CTCCC TTAGT
3701 GCCTT CTCAG TTTGA CGCAG AATAG GCCCG AATCT ACACA GCACC GAATG
3751 AGCAA AACAA GCGAT GGTCT GCCCC GCCTT AGCAT CAACC CCTCT CTTTT
3801 CCATT TTTGT GCAGG TCTGC TGCTT CAGAA ATATC TTGAG GTAGC TACTT
3851 GGCCA GTCAG GATCA CTCCT TCCGG CATGC TTCTC GATAA GCCCG GTACT
3901 CTTCT CTATT CTCTT CTTCA TGAAC TCCAT TTCGC ACCTT TCCAG TTCGG
3951 GCAGT GAGTC AATCT CTTTG AGCTG GTATG TGTCC TTGAA CACCG AGAAC
4001 ATTTG ATGCC CTATA CCGTG GCAAG CCTTG TACTT TCTCC TGTTG GCTTC
4051 ATAGT TCCTG AAGCG CAACC TCTTT TTGAC TGACA TCATG AAGGT AGCTG
4101 TATCA TCAGA TGTGT GCCTC AGGTA CAACG CCTTG TGGGC CATGG GGTCG
4151 CCCAC TTCAC CCTGG TACCC ATCCT CATCT ATCTG TTCGG TTCGC CCTGC
4201 CTCTG TGTGT TGCTC CCGAT GCTCT TTGGC TTTGA GGTCA AATGG TGCAA
4251 GTTCA TTCAT CTTTG TGCTC AGGGC TAGAT GGGTC CTGGT AGGTT CGAGT
4301 ACCAC TGGCT CCTCT ATCTC TATGT CCTCC ATCTC AATCT CGTCG TACAG
4351 GGCCA GAAGG GATTT TAGGG AGGGA TCGCC GGCCA GCCTA GCCTC GTAAT
4401 CCACT GTCTC GGCTA GTCTG CTCTC CTCAA ACAGC ACTGT GGTCT CTGGT
4451 ATGTG CTTTC GGACC ATTGC CACTA TCTCT GACCT TCTGT AGGTC TCGCC
4501 CTTAA GCACT TTCTT CAGCA GTGTG CTTTT AGCAT TGGCG ATCAG ATTCT
4551 TCTTG AGAGC AGGGG TTACA TGTAT GCCTA TCTCA CCTCT GGAGC GAGTC
4601 AAGGC CACCA TCACA TCCTC ATCTT CTTTC TTTGC CCAGT GTTCG TCTAG
4651 GTAAA TGAGG CACGT TTTGA AGCTT AGACC TTGTG TCTCG CTGAC CGTGT
4701 ACCAG CCGGA CCCTT TCTCT TCCTT CATTG CTCTC GTGGC GCATA TGGTA
4751 GGCTC ATCTT TCAGC CTGTT AGAGC GGTAC ACCAC CCTCT TCTCG CTCTC
4801 GGCCC CCATG CATTC TATCT CGAAC AGCTT GCAGT TTCTG GGTAG CCTAT
4851 GTGAA TAGCA TAAGT ATGGT AGTCT CCCTT TCACC CTGTT GAAGA CGTCC
4901 TCTTG ACTAG CCTCC AGCAC CACAT TGTCA CGTTT ACTAT GGTAG GTGCT
4951 TTGAA GAGGG TCTCC CAGGA GCACC AGGGT GCGTG GCTGA TGCGC GGCTA
5001 TCACA AGGTC TATGC ACCCA GGTGG TAGGA GACCA ATCTC GTCTA TCACG
5051 ATGTC TTTGC AGCCG TAGTC ATCCA TGAAT GCCAC TTCGA AGGTC ACCAC
5101 TTTAT GCTTC TTGGA TATTT TTGCT ACCCA ATCAT CTCTG AGCAC TTTCC
5151 TGGGG CTGAT CACAA GCACT TTTTC TACGG AACAG TGTTT GAGTA TGGTT
5201 TGACA CCAGT GACTT TTCCC TGGTC CGGCG AACCC AAAAG CCGTT AGAAC
5251 CTCCT TTGGC CCTTC TATGT CTCCC GTGTG CATCT TCCCT TTGTT GTACT
5301 TGCCG AGTAG CACCC CAGTG TTCCC GTTGT AAAAG CTTTG GGCCA GCCGG

```
5351 TCCGC TTTGC TTTTG TCCAC CCTGA AATGC CCAAG GCCAA CACCA TCACC
5401 AAGTA ACATA GCCTC TCTGA AACCA CCCTT GCTTG TCACC CCTTC GAATG
5451 GCTCC ACATG CCCTG AGGTG GTCTT TAGGT GCAGG CGCCT GTACG ATCCC
5501 GGTGT TTCAG CGTAC CCTCT CTCTC CGTGT ATGGC GATTG GGAAG TCTAG
5551 AGCCT TGGAA AGATT GACGA GTGTG GCTAG ACTGA TCCCG CCATC CTCTA
5601 TCTCC CTCAG AGTTT GAGGG GATGC TCCAT TGACT ATGGA TATCA CGCTT
5651 GGGAT GCTTA ACTGC ATGTG ATCCG CGAAT GCCTT TAGGG AGCAC ATGTT
5701 AGCCT GATTT TTCTG TAGGT ACTCA ATGCC CTCGT CCAGT TCGGC CTCAT
5751 CCTCC TCATA TTCAG AGTCA TCGTC ACCCT TTCTA CTTTC GTAGT CGATC
5801 GTTTT ATTCC GGAGG GTGAT GCTGC TGCGT CCATC AGTAT GGGAG GTGAC
5851 TCTAT GCTTG TGGGT TTGTT GCATC CCCGG CCCCA TCACA TATAC ATCCC
5901 CATCA CTCAT CTCCT TCTTC TCGAT CTTCC CGGAT TGATT CTCTT TTACC
5951 TCAAA GGTGG CTTCA CCGTA TAAAT TGACA GTTAC GACGG AGCCC CCTGG
6001 GAGGT AGCAT GGTTC ATCAT CTGCG TGGAA TCCTA TACTG CCCCC TGCTG
6051 TGTAT CTCTG GATCA GGCAA TGATC GTAGC TATCG TCTAG GCCCA GTTCT
6101 TGTGT GAGCT CATCT AAGGC CTTGT CCCAG CCGAG GGAAC GATGG GAGCC
6151 TCCAT TATAC TTATA CTCCT TGCCG TGCCT GCTAT GGAAC GCCAC CTCGC
6201 GCCCT TTCAA TTTCT CAGAG TAGAT TTCCT CGCGG CACAG GCTCC TCACG
6251 CTCTT GAATT TTTCT ATGTC ACTCT CGCCA CTGGT TCCCG GCATG CTCTC
6301 ATCCT TATGG ATTCC TGTGG TCTCC AGCTT TATGG TGGGT AGTAA GCGCA
6351 GCTCC GCTTC TTCCT CTGTA ATCAC CCCCT CTTCG AGCCC TGATT GTATG
6401 TCCTC AATGT CCTTG ATAGC CCTGC TCTCC ATGAA ACTGA ACCTC CCGAG
6451 CTTGA TGAGG TACTG CCTGA AGTGT ACGAG GGCCC TGGAG GAGCA TACAT
6501 ACCGT GTAAG GTTGA GTTCC ACATC CGTTT GGACC CCTGC AATCA GTTCG
6551 TGGGC CCTGA GTATT CTGGA CAATG GGACC CTCGT CCTGA TTCCT TCTCT
6601 GTGGA TGCAC TCAGT GCCAT AGGGC TCAGA CACTC TATCA TTCCA GCTGA
6651 TGCCG AGGAA GTTCA CCTCG CTGAG AGTGT TAGCA CTGTC GTCTT CCACC
6701 CAATC TATGA CGTTT TGCAC TTTGA ACGTT GCCCC CCTTC TTATG CCCTC
6751 CATAT ATTTC CGCTC CACAT GTATT TCCGT GGTGT CGCAG CTCCG TATGA
6801 ATTTT TCAAG GCACC TGGCA TCGTA CTCCG GTTTG TTGAA CAACC GAGTG
6851 AAGAG GCTTC CGGCT ATGTC GTAGA ATGAT TCCGA CAGTA TGCTC CTTAG
6901 GTTCG GGCTT TCCAT CTTGT CATAC AGAGC CGTCT CCTTG ATCTG CTTGG
6951 CGAGC TGTGC CACAA AGAGG GCCTC ATCCA TGTTT TCCTC TCTAC TGCTA
7001 AGCAT CCTCA ACTTT GCGAC AGCGG AATTT GAGTC GCTTT CCTTC AGAGC
7051 CAGAA GGTAC AGTAG GGTGG GCTTG ATCAT GCGCG CTCTC AAGTA GAGTG
7101 GTTTC CCGCC TCTCC TTCCC TTCAG GATAT CATTA GGGTT TATGA TCGTG
7151 AAGTT ATCGT ACTTG TACGA TTCCT CCGTT AGATA ACTAC CCTTT ATGGT
7201 GCTGA AGAGA TGATG AGCCC CGATT GTTTG GAGCT TTGTG ATCGT CCATG
7251 GGCGT CCCTT CGAAT CCTCT GTCTT TGAGC ATCTC AGCAG CCAAG GGTTC
7301 ACAGG TTGCG TGTAG GCCCC TTCCG CCTTC CCATC TGGGT ACCAC ACTAA
7351 CCTCC CATCC TTCAG GTCGA ACGTG TATGC CTCTG GGAAC AGGCT TTGCG
7401 AATAT CCTGC TTCTA CCTCT GATGG GTACA TGATG CTATA CACTA CCCTA
7451 TCCAC TTTGG ACAAA CTACC TAGGA ACCTT TGCAT TTCAT CTAAA CTCCA
7501 GTATT GGACC TCATC ATGTA TGAAT ATACA TTCGC TCTTG TCTTC CTTGT
7551 TGATC AGATT TGTCA GATTT CCAAC ATCGA GCTCT CTAAA TGCAT TTTCG
7601 TACCT GTGGT GATCC TTTGC ATGCA CCAGT CTGTT GTACT GTGTG AAGTT
7651 CAGAT TCCTG TTGTT TGCAT TTTTC AGACG TAGAC GCTTT ATTTT GCTCT
7701 CTTTA CAGCT AACTA CCAAT GTGTT ACTAG AGAAA TAATT TGGTG CTACT
7751 CTATA TAACA TGTGG TTTTC TATCA TCTTA CTTAC GGGGT GAGCA TGTCC
7801 TTGGA AGGAA TGGAC ACTGA GCTCC ATACC ATAAC CTGTA AGAAT TTCTC
7851 TTAAG GTATC TGGGA TGTAG TAATC AAATA GTCCA TCTAG CCTGG TTTCG
7901 GTTTC CAATA ACCGT TTGCT TTTTA ACTCC TTTAT TGCTT TAACT GACTC
7951 CTCAG ATCCG TTCGT GTACA GGTTG GAGAC TGCGA CACGT TGGGA GGATA
8001 CTGAT ATCGA CATGG TAGTG AATTA GCAGG TGCTT CGCAG AGCGT GGGTG
8051 CTGGC AACGT TTGTT CACTT ATAAT CGATG GGAAT CAAGT TAAAT ATTCT
8101 ATAGT GAGTC GTATT AATCG AATTC CCGCG GTGGA GCTCC AGCTT TTGTT
8151 CCCTT TAGTG AGGGT TAATT GCGCG CTTGG CGTAA TCATG GTCAT AGCTG
8201 TTTCC TGTGT GAAAT TGTTA TCCGC TCACA ATTCC ACACA ACATA CGAGC
8251 CGGAA GCATA AAGTG TAAAG CCTGG GGTGC CTAAT GAGTG AGCTA ACTCA
```

```
 8301 CATTA ATTGC GTTGC GCTCA CTGCC CGCTT TCCAG TCGGG AAACC TGTCG
 8351 TGCCA GCTGC ATTAA TGAAT CGGCC AACGC GCGGG GAGAG GCGGT TTGCG
 8401 TATTG GGCGC TCTTC CGCTT CCTCG CTCAC TGACT CGCTG CGCTC GGTCG
 8451 TTCGG CTGCG GCGAG CGGTA TCAGC TCACT CAAAG GCGGT AATAC GGTTA
 8501 TCCAC AGAAT CAGGG GATAA CGCAG GAAAG AACAT GTGAG CAAAA GGCCA
 8551 GCAAA AGGCC AGGAA CCGTA AAAAG GCCGC GTTGC TGGCG TTTTT CCATA
 8601 GGCTC CGCCC CCCTG ACGAG CATCA CAAAA ATCGA CGCTC AAGTC AGAGG
 8651 TGGCG AAACC CGACA GGACT ATAAA GATAC CAGGC GTTTC CCCCT GGAAG
 8701 CTCCC TCGTG CGCTC TCCTG TTCCG ACCCT GCCGC TTACC GGATA CCTGT
 8751 CCGCC TTTCT CCCTT CGGGA AGCGT GGCGC TTTCT CATAG CTCAC GCTGT
 8801 AGGTA TCTCA GTTCG GTGTA GGTCG TTCGC TCCAA GCTGG GCTGT GTGCA
 8851 CGAAC CCCCC GTTCA GCCCG ACCGC TGCGC CTTAT CCGGT AACTA TCGTC
 8901 TTGAG TCCAA CCCGG TAAGA CACGA CTTAT CGCCA CTGGC AGCAG CCACT
 8951 GGTAA CAGGA TTAGC AGAGC GAGGT ATGTA GGCGG TGCTA CAGAG TTCTT
 9001 GAAGT GGTGG CCTAA CTACG GCTAC ACTAG AAGGA CAGTA TTTGG TATCT
 9051 GCGCT CTGCT GAAGC CAGTT ACCTT CGGAA AAAGA GTTGG TAGCT CTTGA
 9101 TCCGG CAAAC AAACC ACCGC TGGTA GCGGT GGTTT TTTTG TTTGC AAGCA
 9151 GCAGA TTACG CGCAG AAAAA AAGGA TCTCA AGAAG ATCCT TTGAT CTTTT
 9201 CTACG GGGTC TGACG CTCAG TGGAA CGAAA ACTCA CGTTA AGGGA TTTTG
 9251 GTCAT GAGAT TATCA AAAAG GATCT TCACC TAGAT CCTTT TAAAT TAAAA
 9301 ATGAA GTTTT AAATC AATCT AAAGT ATATA TGAGT AAACT TGGTC TGACA
 9351 GTTAC CAATG CTTAA TCAGT GAGGC ACCTA TCTCA GCGAT CTGTC TATTT
 9401 CGTTC ATCCA TAGTT GCCTG ACTCC CCGTC GTGTA GATAA CTACG ATACG
 9451 GGAGG GCTTA CCATC TGGCC CCAGT GCTGC AATGA TACCG CGAGA CCCAC
 9501 GCTCA CCGGC TCCAG ATTTA TCAGC AATAA ACCAG CCAGC CGGAA GGGCC
 9551 GAGCG CAGAA GTGGT CCTGC AACTT TATCC GCCTC CATCC AGTCT ATTAA
 9601 TTGTT GCCGG GAAGC TAGAG TAAGT AGTTC GCCAG TTAAT AGTTT GCGCA
 9651 ACGTT GTTGC CATTG CTACA GGCAT CGTGG TGTCA CGCTC GTCGT TTGGT
 9701 ATGGC TTCAT TCAGC TCCGG TTCCC AACGA TCAAG GCGAG TTACA TGATC
 9751 CCCCA TGTTG TGCAA AAAAG CGGTT AGCTC CTTCG GTCCT CCGAT CGTTG
 9801 TCAGA AGTAA GTTGG CCGCA GTGTT ATCAC TCATG GTTAT GGCAG CACTG
 9851 CATAA TTCTC TTACT GTCAT GCCAT CCGTA AGATG CTTTT CTGTG ACTGG
 9901 TGAGT ACTCA ACCAA GTCAT TCTGA GAATA GTGTA TGCGG CGACC GAGTT
 9951 GCTCT TGCCC GGCGT CAATA CGGGA TAATA CCGCG CCACA TAGCA GAACT
10001 TTAAA AGTGC TCATC ATTGG AAAAC GTTCT TCGGG GCGAA AACTC TCAAG
10051 GATCT TACCG CTGTT GAGAT CCAGT TCGAT GTAAC CCACT CGTGC ACCCA
10101 ACTGA TCTTC AGCAT CTTTT ACTTT CACCA GCGTT TCTGG GTGAG CAAAA
10151 ACAGG AAGGC AAAAT GCCGC AAAAA AGGGA ATAAG GGCGA CACGG AAATG
10201 TTGAA TACTC ATACT CTTCC TTTTT CAATA TTATT GAAGC ATTTA TCAGG
10251 GTTAT TGTCT CATGA GCGGA TACAT ATTTG AATGT ATTTA GAAAA ATAAA
10301 CAAAT AGGGG TTCCG CGCAC ATTTC CCCGA AAAGT GCCAC
```

Appendix 4: ORF alignments

ORFs of pSK-GVA-mutagen-3 were theoretically translated, and aligned to submitted GVA protein sequences on website www.ncbi.nlm.nih.gov/.

Fig. A4.1. ORF1 translation alignment

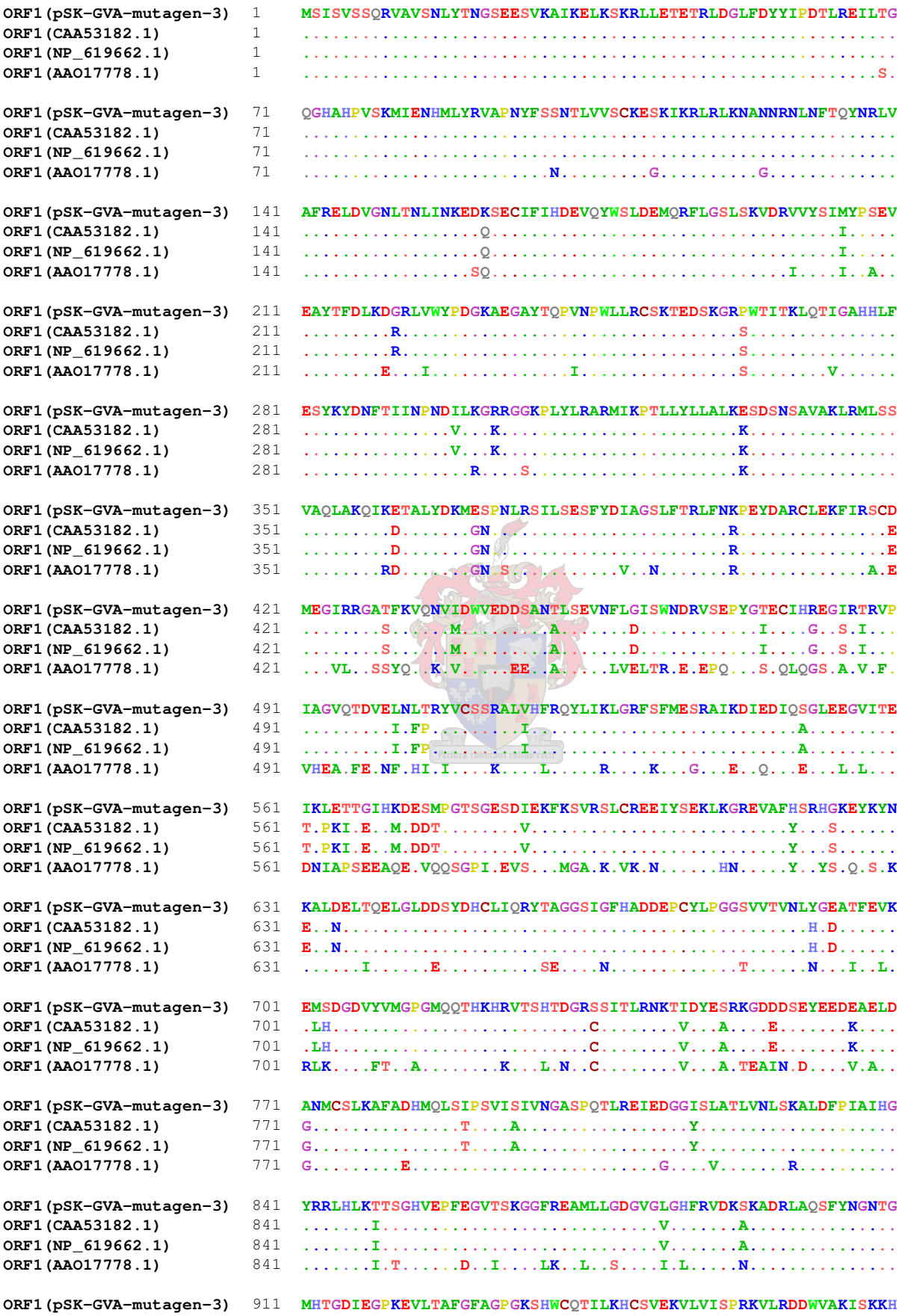


Fig. A4.1. ORF1 translation alignment

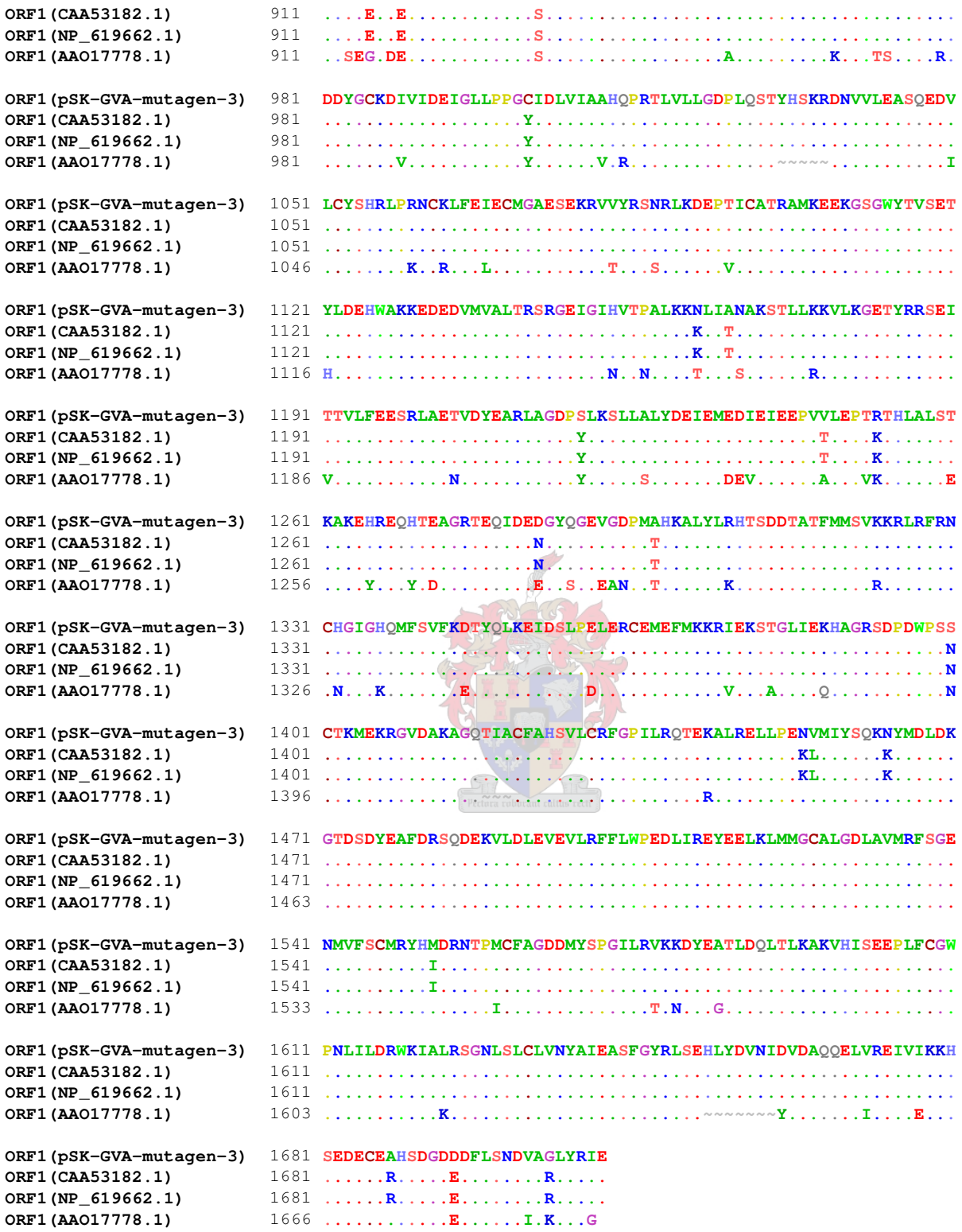


Fig. A4.2. ORF2 translation alignment

```
ORF2(pSK-GVA-mutagen-3)   1   MTSQDCTGLSEFLGCGSDSRSLGVGSLESLSYVQCIQLLSDLKSLGYHSIDSILYILGGGE
ORF2(CAA53183.1)          1   ...R.........H....CDS.S...G.......VN..........Q............
ORF2(AA01779.1)MTSQSCTE   1   ~~~~~~~~~.V...SQ.G.RC..DRT.......T..LY.....RG..........H..ECS.
ORF2(NP_619663.1)         1   ...R.........H....CDS.S...G.......VN..........Q............

ORF2(pSK-GVA-mutagen-3)   71  FRRHGIGIGEALQLGVKKSLCNSPRSLLAILDDLLSRLGRGSAFLPSDLGAVKGELLVTFH
ORF2(CAA53183.1)          71  ..............................................................VA...
ORF2(AA01779.1)MTSQSCTE   63  ...............R....Y....TAE........IS..G..................Q
ORF2(NP_619663.1)         71  ..............................................................VA...

ORF2(pSK-GVA-mutagen-3)   141 VNNKKVVTRSVQAEGDYSYVARRFSGYKGLVLRATRHStop
ORF2(CAA53183.1)          141 ............................I.....Y
ORF2(AA01779.1)MTSQSCTE   133 A.T...C..TC.GK..AG............A..SQGN
ORF2(NP_619663.1)         141 ............................I.....Y
```

Fig. A4.3. ORF3 translation alignment

```
ORF3(pSK-GVA-mutagen-3)   1   MSQEGSLGTKASSSELQDIKVFNVKRSTKDLETLSKTLHRGDVYDTELIEKVFPRRTKKGV
ORF3(AAB62938.1)          1   ..HGD.QA..G.L.DPKE..I.....N.......N.S.............R........C.
ORF3(CAA53184.1)          1   .............F.P......H.....R.....N.S.......N..............C.
ORF3(NP_619664.1)         1   .............F.P......H.....R.....N.S.......N..............C.

ORF3(pSK-GVA-mutagen-3)   71  GRVDCDLDIMDEGLDDIDEEEFPHYHVGCIVVALMPRGKNLQGKVSVEVLDTRLVNGASRI
ORF3(AAB62938.1)          71  .....E..L..G.......T.Y.L.......I....H........V...I....KV.SG..
ORF3(CAA53184.1)          71  ................N.....L............H..................D.....
ORF3(NP_619664.1)         71  ................N.....L............H..................D.....

ORF3(pSK-GVA-mutagen-3)   141 PLSACADFPGYFISTSDLLNGYTLHLSITTTDLQFVDGVHPFSVQLMSLGRFCGEDMKTRY
ORF3(AAB62938.1)          141 ..........................................................NI..L..D......
ORF3(CAA53184.1)          141 ..........................................................I...........
ORF3(NP_619664.1)         141 ..........................................................I...........

ORF3(pSK-GVA-mutagen-3)   211 HQNILNAESDGELTPRGVQVQKVPDTLVMPEVFETIKKLGLKTNGTLRQEGRDKGDNRRVG
ORF3(AAB62938.1)          211 ......SQG...MI.....................Y....R.............G....H.GT.
ORF3(CAA53184.1)          211 ......T.G....I.......................F.....................
ORF3(NP_619664.1)         211 ......T.G....I.......................F.....................
```

Fig. A4.4. ORF4 translation alignment

```
ORF4(pSK-GVA-mutagen-3)   1   MAHYAKRVEIRAIIEELVLAKAQPTDDASESGYDRNMCLNTLFRYIALVGTSKKAVHYGEV
ORF4(NP_619665.1)         1   ....................................Y.....G...........I.....
ORF4(AAB62939.2)          1   ....................T..E...........Y.....G................
ORF4(CAA53185.1)          1   ....................................Y.....G...........I.....
ORF4(BAA92394.1)          1   ....................E.........T.Y.....G................
ORF4(AAQ19964.1)          1   ....................AE..........Y.....G................
ORF4(AAL76173.1)          1   .GQF.R.G....V.......R.ELSE.T.TG....T.Y.....G...........T.....
ORF4(AAL76170.1)          1   ...C..K..M............T..E..T......T.Y.....G...........I.....
ORF4(AAL76167.1)          1   ....................................T.Y.....G................
ORF4(AAM14605.1)          1   ............K.........E...........Y.....G................
ORF4(AAT02513.1)          1   ....................................T.Y.....G................

ORF4(pSK-GVA-mutagen-3)   71  KTGIDPRGKMVISELVGRMRTLSVTVSEGPVKGATLRQMCEPFAQNAYDFLVVMAEIGAYS
ORF4(NP_619665.1)         71  ..........V...........A.............................M.T..
ORF4(AAB62939.2)          71  ..........V...........A.........................L...M.T..
ORF4(CAA53185.1)          71  ..........V...........A.............................M.T..
ORF4(BAA92394.1)          71  ....................A.........................L...M.T..
ORF4(AAQ19964.1)          71  .........L.V..........A..............L.........L...M.T..
ORF4(AAL76173.1)          71  ...L.....INVASM.A.....AA.............................LST.T
ORF4(AAL76170.1)          71  .M.......L.V..........A.........................L...M.T..
ORF4(AAL76167.1)          71  .........L.V..........A.........................L...M.T..
ORF4(AAM14605.1)          71  ..........V..M........A.............................M.T..
ORF4(AAT02513.1)          71  .........I............A.........................L...M.T..

ORF4(pSK-GVA-mutagen-3)   141 GFKEPQVMFDFASGLDLKALTLQEATVIQAMHSRLFRTEGAKEYSTLSHQLANRLSKYRWM
ORF4(NP_619665.1)         141 ..........................................GVFNAQSSIGEQAVEI
ORF4(AAB62939.2)          141 ..........................................GVFNAQSSVGEQAVEI
ORF4(CAA53185.1)          141 ..........................................GVFNAQSSIGEQAVEI
ORF4(BAA92394.1)          141 ..........................................GVFNAQSSVGEQAVEI
ORF4(AAQ19964.1)          141 ..........................................GVFNAQSSVGEQAVEI
ORF4(AAL76173.1)          141 ...........M...T.........S.............GVFNAQSSVGEQAVE
ORF4(AAL76170.1)          141 ..........................................GVFNAQASVGEQAVEL
ORF4(AAL76167.1)          141 ..........................................GVFNAQSSVGEQAVEI
ORF4(AAM14605.1)          141 ...S..................T.........S.............GVFNAQSSVGEQAVEI
ORF4(AAT02513.1)          141 .

ORF4(pSK-GVA-mutagen-3)   211 RMetLSVGALGAStop
ORF4(NP_619665.1)
ORF4(AAB62939.2)
ORF4(CAA53185.1)
ORF4(BAA92394.1)
ORF4(AAQ19964.1)
ORF4(AAL76173.1)
ORF4(AAL76170.1)
ORF4(AAL76167.1)
ORF4(AAM14605.1)
ORF4(AAT02513.1)
```

Fig. A4.5. ORF5 translation alignment

```
ORF5(pSK-GVA-mutagen-3)   1  MDDPSYLSGRSTYAKRRRARRMNVCKCGAIMHNNKDCKSSSISGHKLDRLRFVKEGRVALTG
ORF5(AAL76174.1)          1  .....F.T.....................L......R..T.................E.
ORF5(AAL76171.1)          1  .....F.A................R.............................
ORF5(AAL76168.1)          1  .....F.....................................................
ORF5(AA017780.1)          1  .....F.A....F.....................E.....................T...
ORF5(CAA53186.1)          1  .THHFSRV.PRLLSVG~~.....................S.................
ORF5(NP_619666.1)         1  .THHFSRV.PRLLSVG~~.....................S.................

ORF5(pSK-GVA-mutagen-3)  71  IEWVQAEYHIYILETSDDED
ORF5(AAL76174.1)         71  VK..ET....N........E
ORF5(AAL76171.1)         71  VK..ET.....V.......E
ORF5(AAL76168.1)         71  .....T..............
ORF5(AA017780.1)         71  .R..ET.............K
ORF5(CAA53186.1)         69  VK..ET......V.......
ORF5(NP_619666.1)        69  VK..ET......V.......
```

Table A4.1. Table showing protein homology of constructed clone pSK-GVA-mutagen-3 with submitted protein sequences in Genbank. Percentage identities in ORF 4 do not take the mutation into account. The value given is calculated based on amino acid sequence before the mutation.

Open reading frame of pSK-GVA-mutagen-3	Related proteins	Accession number	E-value	Identities
ORF 1	GVA Putative replicase	gi/1405615	0.0	92%
	GVA RNA-dependent RNA polymerase	gi/27502542	0.0	82%
ORF 2	GVA unknown	gi/1405614	1e-77	84%
ORF 3	GVA ORF in 3' terminal region	gi/1091836	e-144	92%
	GVA movement protein	gi/1405617	e-143	91%
ORF 4 (mutation at 6959)	GVA CP	gi/488390	4e-92	95%
	GVA CP	gi/18642555	7e-91	93%
ORF 5	GVA ORF 5	gi/18642556	3e-43	91%
	GVA ORF 5	gi/18642566	2e-40	82%

Chapter 6: References

Abou-Ghanem, N., Saldarelli, P., Minafra, A., Castellano, M.A., Martelli, G.P. (1997). Properties of grapevine virus D, a novel putative trichovirus. *J. Plant Pathol.* **79**, 1-11.

Ahlquist, P., and Janda, M. (1984). cDNA cloning and the *in vitro* transcription of the complete brome mosaic virus genome. *Mol. Cell. Biol.* **4**, 2876-2882.

Almazan, F., Gonzalez, J. M., Penzes, Z., Izeta, A., Calvo, E., Plana-Duran, J., and Enjuanes, L. (2000). From the cover: Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc. Natl. Acad. Sci. USA* **97**, 5516-5521

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990). Basic Local Allignment Search Tool. *Journal of Molecular Biology,* **215**, 403-410.

Angenent, G.C., Posthumus, E., and Bol, J.F. (1989). Biological activity of transcripts synthesized *in vitro* from full-length and mutated DNA copies of tobacco rattle virus RNA 2. *Virology* **173**, 68-76.

Baulcombe, D.C., Chapman, S., and Santa-Cruz, S. (1995). Jellyfish green fluorescent protein as a reporter for virus infections. *Plant Journal,* **7**, 1045-1053.

Boscia, Minafra and Martelli, in *Filamentous Viruses of Woody Crops*, p. 19, ed. P. L. Monette, Trivandrum: Research Signpost, 1997a.

Boyer, J-C., and Haenni, A-L., 1994. Infectious transcripts and cDNA clones of RNA viruses. *Virology* **198**, 415-426.

Burger, J.G., and Spreeth, N.A. (1993). Occurrence of Shiraz disease in South Africa. In: Gugerli P, ed. Extended Abstracts 11 Meeting ICVG, 1993. Montreux, Switzerland, 56.

Campbell, M. S., and Pletnev, A. G. (2000). Infectious cDNA clones of langat tick-borne flavivirus that differ from their parent in peripheral neurovirulence. *Virology* **269**, 225-237.

Chevalier, S., Greif, C., Clauzel, J.M., Walter, B., Firtsch, C. (1995). Use of an immunocapture-polymerase chain reaction procedure for the detection of GVA in Kober stem grooving-infected grapevines. *J. Phytopathol* **143**, 369-37.

Conti, M., Milne, R.G., Luisoni, E., Boccardo, G. (1980). A closterovirus from a stem pitting-diseased grapevine. *Phytopathology* **70**, 394-399.

Commandeur, U., Jarausch, W., Li, Y., Koenig, R., and Burgermeister, W. (1991). cDNAs of beet necrotic yellow vein virus RNA-3 and RNA-4 are rendered biologically active in a plasmid containing the cauliflower mosaic virus 35S promoter. *Virology* **185,** 493-495.

Davis, N.L., Willis, L.V., Smith, J.F., and Johnston, R.E. (1989). *In vitro* synthesis of infectious Venezuelan equine encephalitis virus RNA from a cDNA clone: Analysis of a viable deletion mutant. *Virology* **171**, 189-204.

Dawson, W.O., Beck, D.L., Knorr, D., and Grantham, G.L. (1986). cDNA cloning of the complete genome of tobacco mosaic virus and production of infectious transcripts. *Proc. Natl. Acad. Sci. USA* **83**, 1832-1836.

Dell'Orco, M., Saldarelli, P., Minafra, A., Boscia, D., and Gallitelli, D. (2002) Epitope mapping of the GVA capsid protein. *Archives of Virology* **147**, 627-634.

Digiaro, Boscia, Simeone and Savino, *Extended Abstracts 12th Meeting of ICVG, Lisbon, 1997*: 169, 1997.

Digiaro, Popovic Bedzrob, D'Onghia, Boscia and Savino (1994), *Phytopathologia Mediterranea* **33**, 187.

Domingo, E., Baranowski, E., Ruiz-Jarabo, C.M., Martin-Hernandez, A.M., Saiz, J.C., and Escarmis, C. (1998). Quasispecies structure and persistence of RNA viruses. *Emerging infectious diseases,* 4.

Dore, J.-M., Erny, C., and Pinck, L. (1990). Biologically active transcripts of alfalfa mosaic virus RNA-3. *FEBS Lett.* **264**, 183-186.

Drake, J.W., Holland, J.J. (1999). Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. USA* **96**, 13910-13913.

Dunn, J.J., and Studier, F.W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the location of T7 genetic elements. *J. Mol. Biol.* **166**, 477-485.

Dzianott, A.M., and Bujarski, J.J. (1989). Derivation of an infectious viral RNA by autocatalytic cleavage of *in vitro* transcribed viral cDNAs. *Proc. Natl. Acad. Sci. USA* **86**, 4823-4827.

Engelbrecht, D.J., Kasdorf G.G.F. (1987). Occurrence and transmission of GVA in South African grapevines. *S. Afr. J. Enol. Vitic*. **8**, 23-29.

Engelbrecht, D.J., Kasdorf G.G.F. (1990). Field spread of corky bark, fleck. leafroll and Shiraz decline diseases and the associated viruses in South African grapevines. *Phytophylactica* **22**, 347-354.

Eggen, R., Verver, J., Wellink, J., De Jong, A., Goldbach, R., and van Kammen, A. (1989). Improvements of the infectivity of *in vitro* transcripts from cloned cowpea mosaic virus cDNA: Impact of terminal nucleotide sequences. *Virology* **173**, 447-455.

Fakhfakh, H., Vilaine, F., Makni, M., and Robaglia, C. (1996). Cell-free cloning and biolistic inoculation of an infectious cDNA of *Potato virus* Y. J. *Gen. Virol*. **77,** 519-523.

Fischer, C.L. and Pei, G.K. (1997). Modification of a PCR-based site-directed mutagenesis method. *Biotechniques* **23**, 570-574.

Forns, X., Purcell, R.H., Bukh, J. (1999). Quasispecies in viral persistence and pathogenesis of hepatitis C virus. *Trend Microbiol.* **7** (10), 402-410.

Freeborough, M-J., (2003). A pathogen-derived resistance strategy for the broad-spectrum control of grapevine leafroll-associated virus infection. Ph.D. thesis.

Martelli, G.P., Minafra, A., Saldarelli, P. (1997). Vitivirus, a new genus of plant viruses [news]. *Arch. Virol*. **142**, 1929-1932.

Furuichi, Y., Lafiandra, A., and Shatkin, A.J. (1977). 5'-Terminal structure and mRNA stability. *Nature* **266**, 235-239.

Galiakparov, N., Tanne, E., Seal, I., Gafny, R., (1999). Infectious RNA transcripts from GVA cDNA clone. *Virus Genes* **19**, 235-242.

Galiakparov, N., Goszczynski, D.E., Che, X., Batuman, O., Bar-Joseph, M., and Mawassi, M. (2003). Two classes of subgenomic RNA of GVA produced by internal controller elements. *Virology* **312**, 434-448.

Galiakparov, N., Tanne, E., Seal, I., Gafny, R. (2003). Functional analysis of the GVA genome. *Virology* **306**, 42-50.

Gal-On, A., Antignus, Y., Rosner, A., and Raccah, B. (1991). Infectious *in vitro* RNA transcripts derived from cloned cDNA of the cucurbit potyvirus, zucchini yellow virus. *J. Gen. Virol.* **72**, 2639-2643.

Gal-On, A., Meiri, E., Huet, H., Hua, W.J., Raccah, B., and Gaba, V. (1995). Particle bombardment drastically increases the infectivity of cloned DNA of zucchini yellow mosaic potyvirus. *J. Gen. Virol.* **76**, 3223-3227.

Garau, R., Prota, V.A., Boscia, D., Fiori, M., Prota, U., 1995. *Pseudococcus affinis* Mask., a new vector of trichovirus A and B. *Vitis* **34**, 67-68.

Geigenmuller, U., Ginzton, N. H., and Matsui, S. M. (1997). Construction of a genome length cDNA clone for human astrovirus serotype 1 and synthesis of infectious RNA transcripts. *J. Virol.* **71**, 1713-1717.

Geiser, M., Cèbe, R., Drewello, D., and Schmitz, R. (2001). Integration of PCR fragments at any specific site within cloning vectors without the use of restriction enzymes and DNA ligase. *BioTechniques* **31**, 88-92.

Goszczynski, D.E., Jooste, A.E.C. (2003). Identification of divergent variants of GVA. *European Journal of Plant Pathology* **109**, 397-403.

Goszczynski, D.E., Jooste, A.E.C. (2002). The application of single-strand conformation polymorphism (SSCP) technique for the analysis of molecular heterogeneity of GVA. *Vitis,* **41**, 77-82.

GVA (2004). Descriptions of plant viruses. (http://www.dpvweb.net/dpv/showdpv.php?dpvno=383, www.dpvweb.net/notes/showgenus.php?genus=Vitivirus)

Gretch, D.R., Polyak, S.J., Wilson, J.J., Carithers, R.L., Perkins, J.D., Jr, Corey, L. (1996). Tracking Hepatitis C virus quasispecies major and minor variants in symptomatic and asymptomatic liver transplant recipients. *J. Virol.* **70** (11), 7622-7631.

Gritsun, T. S., and Gould, E. A. (1998). Development and analysis of a tick-borne encephalitis virus infectious clone using a novel and rapid strategy. *J. Virol. Methods* **76**, 109-120.

Gualano, R. C., Pryor, M. J., Cauchi, M. R., Wright, P. J., and Davidson, A. D. (1998). Identification of a major determinant of mouse neurovirulence of dengue virus type 2 using stably cloned genomic-length cDNA. *J. Gen. Virol.* **79**, 437-446.

Garau, Fiori, Prota, Tolu, Fiori and Prota, *Extended Abstracts 12th Meeting of ICVG, Lisbon, 1997*: 171, 1997

Hamilton, W.D.O., and Baulcombe, D.C. (1989). Infectious RNA produced by *in vitro* transcription of a full-length tobacco rattle virus RNA-1 cDNA. *J. Gen. Virol.* **70**, 963-968.

Hall, T.A. (1999). BioEdit: a user friendly biological sequence alignment editor and analysis program for windows 95/98/NT, *Nucl Acids Symp Ser*, **41**, 95-98.

Hayes, R.J., and Buck, K.W. (1990). Infectious cucumber mosaic virus RNA transcribed *in vitro* from clones obtained from cDNA amplified using the polymerase chain reaction. *J. Gen. Virol.* **71**, 2503-2508.

Heaton, L.A., Carrington, J.C., and Morris, T.J. (1989). Turnip crinkle virus infection from RNA synthesized *in vitro. Virology* **170**, 214-218.

Heiman, M. (1997). Webcutter 2.0, (http://www.firstmarket.com.cutter/cut2.html.).

Herold, J., Thiel, V., and Siddell, S. G. (1998). A strategy for the generation of infectious RNAs and autonomously replicating RNAs based on the HCV 229E genome. *Adv. Exp. Med. Biol.* **440**, 265-268.

Hurrelbrink, R. J., Nestorowicz, A., and McMinn, P. C. (1999). Characterization of infectious Murray Valley encephalitis virus derived from a stably cloned genome-length cDNA. *J. Gen. Virol.* **80**, 3115-3125.

Jakab, G., E. Droz, G. Brigneti, D. Baulcombe, and P. Malnoe. (1997). Infectious in vivo and in vitro transcripts from a full-length cDNA clone of PVY-N605, a Swiss necrotic isolate of potato virus Y. J. *Gen. Viro*. **78,** 3141-3145.

Janda, M., French, R., and Ahlquist, P. (1987). High efficiency T7 polymerase synthesis of infectious RNA from cloned brome mosaic virus cDNA and effects of 5' extensions on transcript infectivity. *Virology* **158**, 259-262.

Keohavong, P., and Thilly, W.G. (1989). Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. USA* **86**, 9253-9257.

Khromykh, A. A., and Westaway, E. G. (1994). Completion of Kunjin virus RNA sequence and recovery of an infectious RNA transcribed from stably cloned full-length cDNA. *J. Virol.* **68**, 4580-4588.

Klump, W.M., Bergmann, I., Müller, B.C., Ameis, D., and Kandolf, R. (1990). Complete nucleotide sequence of infectious coxsakie-virus B3 cDNA: Two initial 5' uridine residues are regained during plus-strand RNA synthesis. *J. Virol.* **64**, 1573-1583.

La Notte, P., Buzkan, N., Choueiri, E., Minafra, A. and Martelli, G.P. (1997) Acquisition and transmission of GVA by the mealybug *Pseudococcus longispinus. Journal of Plant Pathology* **78,** 79-85.

La Notte, P., Minafra, A. and Saldarelli, P. (1997) A spot-PCR technique for the detection of phloem-limited grapevine viruses. *Journal of Virological Methods,* **66,** 103-108.

Lewandowski, D.J., and Dawson, W.O. (1998). Deletion of internal sequences results in tobacco mosaic virus defective RNAs that accumulate to high levels without interfering with replication of the helper virus. *Virology* **251**, 427-437.

Li, H., Li, W. X., and Ding, S.W. (2003). Induction and suppression of RNA silencing by an animal virus. *Science* **296**, 1319-1321.

Lu, M., Funsch, B., Wiese, M., Roggendorf, M. (1995). Analysis of hepatitis C virus quasispecies populations by temperature gradient gel electrophoresis. *J. Gen. Virol.* **76**, 881-887.

Malet, I., Belnard, M., Agut, H., and Cahour, A. (2003). From RNA to quasispecies: a DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *Journal of virological methods* **109**, 161-170.

Martell, M., Esteban, J.T., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J., Gomez, J. (1992). Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution. *J. Virol.* **66** (5), 3225-3229.

Martelli, G.P. (ed) (1993). Graft transmissible diseases of grapevines. Handbook for detection and diagnosis. FAO Publication Division, Rome.

McCormick, A.A., Kumagai, M.H., Hanley, K., Turpen, T.H., Hakim, I., Grill, L.K., Tuse, D., Levy, S., and Levy, R. (1999). Rapid production of specific vaccines for lymphoma by expression of the tumor-derived single-chain Fv epitopes in tobacco plants. *Proc. Natl. Acad. Sci. USA* **96**, 703-708.

McKechnie, V.M., McCruden, E.A., Mills, P.R., Sinclair, K. (2001). Optimisation and sensitivity of singlr-stranded conformation polymorphism for the detection of hepatitis C virus quasispecies. *J. Virol. Methods* **92** (2), 131-139.

Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**, 7035-7056.

Mendez, E., Ruggli, N., Collett, M. S., and Rice, C. M. (1998). Infectious bovine viral diarrhea virus (strain NADL) RNA from stable cDNA clones: A cellular insert determines NS3 production and viral cytopathogenicity. *J. Virol.* **72**, 4737-4745.

Meyers, G., Thiel, H. J., and Rumenapf, T. (1996). Classical swine fever virus: Recovery of infectious viruses from cDNA constructs and generation of recombinant cytopathogenic defective interfering particles. *J. Virol.* **70**, 1588-1595.

Miller, A.M., and Koev, G. (2000). Synthesis of subgenomic RNAs by Positive-strand RNA viruses. *Virology* **273**, 1-8.

Minafra, A., Saldarelli, P., Martelli, G.P. (1997). GVA: nucleotide sequence, genome organization, and relationship in the Trichovirus genus. *Arch. Virol.* **142**, 417-423.

Minafra, A., Saldarelli, P., Grieco, F., Martelli, G.P. (1994). Nucleotide sequence of the 3' terminal region of the RNA of two filamentous grapevine viruses. *Arch. Virol.* **137**, 249-261.

Obreque, J., Minafra, A., Turturo, C., Dell'Orco, M., Saldarelli, P., Diaz, D., Vera, J., Jashes, M., Pena-Cortes, H., and Martelli, G.P. (2003). Preliminary evidence of the presence of defective interfering RNAs in Vitivirus-infected *Nicotiana* plants. *14th ICVG Conference*, Locorotondo, 12-17th September, 2003.

Polo, S., Ketner, G., Levis, R., and Falgout, B. (1997). Infectious RNA transcripts from full-length dengue virus type 2 cDNA clones made in yeast. *J. Virol.* **71**, 5366-5374.

Puurand, U., J. P.T. Valkonen, K. Makinen, F. Rabenstein, and M. Saarma. (1996). Infectious *in vitro* transcripts from cloned cDNA of the potato A potyvirus. *Virus Res*. **40,** 135-140.

Rezaian, M.A. and Krake, L.R. (1987) Nucleic acid extraction and virus detection in grapevine. *Journal of Virological Methods,* **17,** 277-285.

Roossinck, M.J. (2003). Plant RNA virus evolution. *Current Opinion in Microbiology* 6, 406-409.

Rosciglione, B., Castellano, M.A., Martelli, G.P., Savino, V., Cannizzaro, G. (1983). Mealybug transmission of GVA. *Vitis* **22**, 331-347.

Saldarelli, P., Minafra, A., Martelli, G.P. (1996). The nucleotide sequence and genomic organization of grapevine virus B. *Gen. Virol*. **77**, 2645-2652.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning: a laboratory manual, 2nd edition. Cold Spring Harbour Laboratory Press, NY.

Sarnow, P. (1989). Role of 3'-end sequences in infectivity of poliovirus transcripts made *in vitro. J. Virol.* **63**, 467-470.

Satyanarayana, T., Gowda, S., Ayllón, M. A., and Dawson, W.O. (2003). Frameshift mutations in infectious cDNA clones of *Citrus tristeza* virus: a strategy to minimize the toxicity of viral sequences to *Escherichia coli. Virology* **313,** 481-491.

Schneider, W.L., Roossinck, M.J. (2000). Evolutionary related sindbis-like plant viruses maintain different levels of population diversity in a common host. *J. Virol,* **74**, 3130-3134.

Schneider, W.L., Roossinck, M.J. (2001).Genetic diversity in RNA viral quasispecies is controlled by host-virus interactions. *J. Virol.* **7**, 6566-6571

Shih, D.S., Dasgupta, R., and Kaesberg, P. (1976). 7-Methyl-guanosine and efficiency of RNA translation. *J. Virol.* **19**, 637-642.

Simon, A.E., and Bujarski, J.J. (1994). RNA-RNA recombination and the evolution in virus-infected plants. *Annual Review of Phytopathology* **32**, 337-362.

Stamp, J.A. (2004). Syrah Decline. *Wine Business Monthly* **XI** (2), 1 February 2004.

Sumiyoshi, H., Hoke, C. H., and Trent, D. W. (1992). Infectious Japanese encephalitis virus RNA can be synthesized from *in vitro*-ligated cDNA templates. *J. Virol.* **66**, 5425-5431.

Turturo, C., Dell'Orco, M., Saldarelli, P., Minafra, A., and Stavolone, L. (2003). RNA silencing is suppressed by GVA infection. 14th ICVG Conference, Locorotondo, 12th-17th September, 2003.

Ullah, Z., Chai, B., Hammar, S., Raccah, B., Gal-On, A., and Grumet, R., (2003). Effect of the amino termini of coat proteins of distinct potyvirus species on viral infectivity and host specificity. *Physiological and Molecular Plant Pathology* **63**, 129-139.

Van Bakoven, H., Verver, J., Wellinck, J., and van Kammen, A. (1993). Protoplasts transiently expressing the 200K coding sequence of cowpea mosaic virus B-RNA support replication of M-RNA. *J. Gen. Virol.* **74**, 2233-2241.

Verver, J., Goldbach, R., Garcia., J.A., and Vos, P. (1987). *In vitro* expression of a full-length DNA copy of cowpea mosaic virus B RNA: Identification of the B RNA encoded 24-kd protein as a viral protease. *EMBO J* **6**, 549-554.

Voinnet, O. (2001). RNA silencing as a plant immune system against viruses. *Trends Genetics* **17**, 449-459.

Voinnet, O., Lederer, C., and Baulcombe, D.C. (2000). A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana. Cell* **103**, 157-167.

Voinnet, O., Pinto, Y.M., and Baulcombe, D.C. (1999). Suppression of gene silencing: a general strategy used by diverse DNA and RNA viruses of plants. *Proc. Natl. Acad. Sci. USA* **96**, 14147-14152.

Viry, M., Serghini, M.A., Hans, F., Ritzenthaler, C., Pinck, M., and Pinck, L. (1993). Biologically active transcripts from cloned cDNA of genomic grapevine fanleaf nepovirus RNAs. *J. Gen. Virol.* **74**, 169-174.

Wait, D., (2003). South African Wine Industry Statistics (SAWIS). (http://www.sawis.co.za)

White, K.A., and Morris, T.J. (1999). Defective and defective interfering RNAs of monopartite plus-strand RNA plant viruses. *Curr Top Microbiol. Immunol.* **239**, 1-17.

World Development Report (2003). *Systainable Development in a Dynamic World.* World Band and Oxford University Press.

Worldwide statistics (2004). Wines of South Africa. (http://www.wosa.co.za/statistics.asp)

Yamamiya, A., and Shirako, Y. (2000). Construction of full-length cDNA clones to *Soil-Borne Wheat Mosaic Virus* RNA1 and RNA2, from which infectious RNAs are transcribed *in vitro*: virion formation and systemic infection without expression of the N-terminal and C-terminal extensions to the capsid protein. *Virology* **277**, 66-75.