Construction of a cDNA library for the vine mealybug, Planococcus ficus (Signoret)

by

Kora Holm

Thesis presented in partial fulfilment of the requirements for the degree of

Master of Science at the Department of Genetics, Stellenbosch University

Study leader: Prof J.T. Burger

December 2008

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 8 December 2008

Abstract

Construction of a cDNA library for the vine mealybug, *Planococcus ficus* (Signoret)

K. Holm

Department of Genetics
University of Stellenbosch
Private Bag X1, 7602 Matieland, South Africa

Thesis:

December 2008

The vine mealybug, *Planococcus ficus* (Signoret), is a severe pest of grapevine in many grape and wine producing countries around the world. It is renowned not only for the considerable damage it infers to grapevine of its own accord, but in particular for its role in transmitting deleterious viral diseases such as grapevine leafroll disease, Kober stem grooving, Shiraz disease and corky bark. Incidentally, it is an exceptionally tenacious antagonist of grapevine, being resistant to both chemical and biological control mechanisms. As a result, finding an effective strategy for *P. ficus* control has become a main priority of viticultural industries worldwide.

Possible implementation of biotechnological approaches to pest management has resulted in a need for *P. ficus* genetic data - of which there are currently very little available. The transcribed genes of an organism can be captured in a cDNA library, and the sequences of the various transcripts can then be characterized.

In this study altogether five cDNA libraries were constructed from the transcribed sequences of *Planococcus ficus* (Signoret). Instrumental to their construction was the identification of an RNA extraction protocol that provided large quantities of high quality RNA from mealybugs. The five cDNA libraries were the result of a set of modifications to the Creator™ SMART™ cDNA Library Construction Kit (used for Primary Library construction), and differed mainly with regards to range of insert sizes they contain. Whereas an abundance of short fragments were found in the Primary Library (42% of screened inserts $\leqslant$0.5 kb, and 20% $\geqslant$1 kb), the Fractionated Libraries contained inserts of specific size ranges that were more-or-less equally represented. The broadest size range was found in Fractionated Library 4, for which a uniform distribution over the range $\sim$0.25 kb - 4 kb was observed. Average insert sizes of Fractionated Libraries 1 to 4 were estimated at 0.25 kb, 0.5 kb, 1 kb and 2 kb respectively. These results demonstrated the importance of using a protocol designed to circumvent the bias towards incorporation of shorter transcripts in cDNA libraries.

Although the libraries were not exhaustively analyzed, the outcome of a pilot investigation indicated that 41% of the submitted sequences had matches in the non-redundant database of the National Center for Biotechnology Information (NCBI, E-value $\leqslant 10^{-5}$), and that approximately 82% of these were of insect origin. Moreover, two potential targets for an RNAi-mediated approach to *P. ficus* pest control were identified. With one exception, these sequences seemed to be unique to arthropods. Future research needs to investigate the efficiency by which these sequences are able to constrain *P. ficus* proliferation, and their suitability for grapevine transformation.

Uittreksel

Construction of a cDNA library for the vine mealybug, *Planococcus ficus* (Signoret)

K. Holm

Departement Genetika
Universiteit van Stellenbosch
Privaatsak X1, 7602 Matieland, Suid-Afrika

Tesis:

Desember 2008

Die wingerdwitluis, *Planococcus ficus* (Signoret), is 'n ernstige plaag in menige wyn- en druifproduserende lande reg oor die wêreld. Die plaag beskik oor die vermoë om grootskaalse wingerdskade aan te rig as 'n direkte gevolg van oorvoeding en 'n indirekte gevolg van die afskeiding van heuningdou, maar is veral gedug vanweë sy rol as 'n vektor van ongeneesbare virussiektes in wingerd soos Shiraz siekte, gleufstamsiekte, skurfbassiekte en wingerd rolblaar siekte. *P. ficus* is boonop weerstandbiedend teen beide chemiese en biologiese beheermaatreëls, en die soeke na alternatiewe plaagbeheermaatreëls het gevolglik 'n hoë prioriteit vir die wingerd- en wynbou bedryf geraak.

Die moontlike gebruik van biotegnologiese plaagbeheermetodes teen *P. ficus* plaagbeheer het 'n behoefte laat onstaan om die huidige tekort aan genetiese informasie oor *P. ficus* aan te vul. Komplementêre deoksieribonukleïensuur (kDNS)-biblioteke is vir hierdie doel geskik, aangesien hulle fisiese argiewe van die uitgedrukte gene van 'n organisme daarstel en as sulks grondige fondamente bied vir die ontginning van voorheen onbekende geen-transkripte.

Gedurende die verloop van hierdie studie is daar altesame vyf kDNS-biblioteke van wingerdwitluis geen-transkripte gemaak. Onderliggend hieraan was die identifisering van 'n ribonukleïensuur (RNS)-isoleringsprotokol wat geskik was vir gebruik op wasbedekte insekte soos die wingerdwitluis. Die vyf biblioteke was die eindresultaat van 'n reeks veranderinge aan die Creator™ SMART™ cDNA Library Construction Kit protokol aangebring, en hulle verskil hoofsaaklik ten opsigte van die grootte-verspreiding en mees algemene kDNS-insetselgroottes in elk. Waar die Primêre Biblioteek gekenmerk word aan 'n oormaat kort insetsels (42% van alle ondersoekte fragmente $\leqslant$0.5 kb, en 20% $\geqslant$1 kb), bestaan die Fraksionele Biblioteke uit reëlmatige verspreidings van insetsels van spesifieke grootte-ordes. Die breedste insetselgrootteverspreiding was die van die vierde Fraksionele Biblioteek, wat 'n uniforme verspreiding oor die grootte-reeks ∼0.25 kb - 4 kb beslaan. Die gemiddelde kDNS-fragmentgroottes in die vier Fraksionele Biblioteke was ∼0.25 kb, 0.5 kb, 1 kb en ∼2 kb onderskeidelik. Hierdie resultate wys op die belang van eksperimentele ontwerpe wat gerig is om partydigheid vir inkorporering van korter insetsels in kDNS-biblioteke uit die weg te ruim.

Hoewel die vyf biblioteke nie in meer detail ondersoek was nie, het 'n steekproefondersoek aangedui dat ongeveer 41% van die ondersoekde kDNS-volgordes wat ondersoek was 'n ooreenkomstige volgorde in die *National Center for Biotechnology Information* het (uitsnywaarde van E $\sim 10^{-5}$) en dat ongeveer 82% van die volgordes van insekte afkomstig is. Daarbenewens is twee potensiële teikens vir *P. ficus* plaagbeheer deur middel van geenonderdrukking uitgewys. Behalwe vir een geval, blyk die teiken kDNS-volgordes uniek te wees vir die filum Arthropoda. Toekomstige navorsing sou kon aandui of hierdie teiken volgordes *P. ficus*-aanwas kan inperk, en of hulle geskik sou wees vir wingerdstok transformering. Indien nie, sou die biblioteke moontlik gebruik kan word vir die ontginning van meer geskikte teiken volgordes.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations who have contributed to making this work possible:

- Prof Burger for always whenever and wherever being willing to encourage a generalist who, somehow erroneously, ended up in a highly specialized environment such as Lab 228.
- The Vitis lab: Annerie, Antman, Dirk, Hano, Jacques, Johan, Liza, Mandi, Marguerite, Mark, Michael-John for making it all so much fun - especially when it wasn't funny anymore.
- Gloudi - for her hardships and advice with the mealybug sequences.
- Shimoda Biotech (Pty) Ltd., KWV-SA and the Harry Crossley Foundation for research and personal funding during the course of my studies.
- My little team of involuntary editors: Jacky, Riaan & Vater (none of which should be held responsible for the errors that I introduced after their painstaking efforts. Or for the good advice ignored.)
- My wonderful friends.
- My delightfully bizarre, fantastic, fanatic, crazy, loop-headed and (presumably) misplaced family - for creating alternative universes everywhere they go.

Dedications

Vir my wetenskapper gewetenskepper,
Vater.

Contents

Declaration ii

Abstract iii

Uittreksel v

Acknowledgements vii

Dedications viii

Contents ix

List of Figures xii

List of Tables xiv

Abbreviations xv

1 Introduction 1
- 1.1 General Introduction . . . 1
- 1.2 Project Proposal . . . 3

2 Literature Review 5
- 2.1 The vine mealybug, *Planococcus ficus* (Signoret) . . . 5
  - 2.1.1 Identification in South African Vineyards . . . 6
  - 2.1.2 Biology . . . 6
  - 2.1.3 Detriments of Infestation . . . 8
  - 2.1.4 Mealybug Control . . . 10
- 2.2 The need for Novel Approaches to Insect Pest Control . . . 11
- 2.3 Current Status of the Genetic Information available for *Planococcus ficus* (Signoret) . . . 13
- 2.4 Technologies by which Expressible Genetic Sequence Information can be obtained . . . 14
- 2.5 Isolating Ribonucleic Acids . . . 15
  - 2.5.1 Membrane Solubilization . . . 16

2.5.2 Inhibition of Ribonucleases . . . 16
2.5.3 RNA Separation from Cellular Lysates . . . 16
2.5.4 Storage of Purified RNA . . . 17
2.5.5 Quantification and Quality Assessment of RNA . . . 18
2.6 Constructing cDNA Libraries . . . 18
2.6.1 First-strand Synthesis . . . 19
2.6.2 Second-strand Synthesis . . . 20
2.6.3 cDNA Cloning . . . 21
2.6.4 Library Screening . . . 21
2.7 Creator™ SMART™ cDNA library Construction Technology . . 22

3 Experimental Procedures 25
3.1 Introduction . . . 25
3.2 Mealybug Rearing . . . 26
3.3 RNA Isolation . . . 26
3.4 Preparation and Ligation of Primary Library cDNA . . . 27
3.5 Preparation and Ligation of Fractionated Library cDNAs . . . 28
3.5.1 Reverse Transcription . . . 28
3.5.2 Second-strand Synthesis . . . 29
3.5.3 Size Fractionation . . . 29
3.5.4 cDNA Amplification . . . 29
3.5.5 cDNA Purification . . . 30
3.5.6 Enzymatic Digestion . . . 30
3.5.7 Vector Preparation . . . 31
3.5.8 Ligations . . . 33
3.6 Transformations . . . 33
3.7 Library Titering . . . 34
3.8 Screening for Inserts . . . 34
3.9 Library Amplification . . . 35
3.10 Insert Sequencing & Data Analysis . . . 36
3.10.1 Sequencing . . . 36
3.10.2 Sequence Editing . . . 36
3.10.3 Blast Analysis . . . 36
3.10.4 Identification of Potential Target Sequences . . . 37
3.10.5 Sequence Submission . . . 38
3.11 Frequently used Techniques . . . 38
3.11.1 Agarose Gel Electrophoresis . . . 38
3.11.2 Culture Media . . . 38
3.11.3 Overnight Cultures . . . 39

4 Results and Discussion 40
4.1 RNA Isolation from Mealybugs . . . 40
4.2 Construction of the Primary Library . . . 42
4.2.1 Evaluating the Size-distribution of Clone Inserts . . . 45

4.2.2 Troubleshooting 47
4.2.3 Formulating a New Methodology 48
4.3 Construction of the Fractionated Libraries 48
4.3.1 Recovering the pDNR-LIB Vector Backbone 49
4.3.2 Preparing the cDNA Size-fractions 50
4.3.3 Cloning the Four Size-fractions 52
4.3.4 Evaluating the Size-distribution of Clone Inserts 53
4.4 Comparative Analysis of the Sequencing- and Blast Results of the Primary- and Fractionated Libraries 55
4.5 Identification of Potential Targets for RNAi- mediated *P.ficus* Control 57

5 General Discussion **59**

6 Conclusion **62**

Appendices **63**

A Precautions when working with RNA **63**
A.1 Creating an RNase-free working environment 63

B RNA Extraction Protocols less suitable for use on Mealybugs **65**

C Protocol for preparation of Chemically Competent Cells **66**

D Calculations **67**
D.1 Primer addition to second-strand reaction 67
D.2 pDNR-LIB: Calculation of picomole ends 69
D.3 Ligations: Volume of insert needed to establish the desired vector:insert ratio 70
D.4 Library titering 71
D.5 Library amplification: Volume of transformation reaction needed 72

E GeneRuler™ 1kb DNA Ladder **73**

F Python Scripts **74**

G Blast Results **82**

Bibliography **100**

List of Figures

2.1 The female and male forms of the vine mealybug, *Planococcus ficus* (Signoret) . . . 7
(a) Mature female . . . 7
(b) Mature male . . . 7
2.2 Illustration of the two *Sfi*I restriction sites . . . 23
2.3 Flow chart of the Creator™ SMART™ cDNA Library Construction Kit protocols . . . 24

4.1 Electrophoresis profile of mealybug total RNA . . . 41
4.2 Amplified cDNA of the Primary Library . . . 43
4.3 Size-fractionation of Primary Library cDNA . . . 43
4.4 Control for the ligation reactions of the Primary Library . . . 44
4.5 Transformation reactions of the Primary Library after plating on selective medium . . . 45
(a) Ligation A . . . 45
(b) Ligation B . . . 45
(c) Ligation C . . . 45
(d) Negative Control . . . 45
(e) No selection . . . 45
4.6 Colony PCRs of the Primary Library . . . 46
4.7 Size-fractionation of GeneRuler™1kb DNA Ladder . . . 47
4.8 Agarose gel displaying the most important stages of vector purification . . . 49
4.9 Second-strand cDNA & size-fractionation for the creation of the Fractionated Libraries . . . 50
(a) Second-strand reaction. . . . 50
(b) Incisions for size-fractionation. . . . 50
4.10 Trial amplification productss of Fractionated Library cDNA after size-fractionation . . . 51
4.11 Amplification products of Fractionated Library Fractions 1-4 . . . 51
4.12 Colony PCRs of the first three Fractionated Libraries . . . 54
(a) Amplification products of Fractionated Library 1 . . . 54
(b) Amplification products of Fractionated Library 2 . . . 54

(c) Amplification products of Fractionated Library 3 54
4.13 Colony PCRs of the fourth Fractionated Library 55
4.14 Quantitative overview of sequencing- and blast results obtained for the Primary- and Fractionated Libraries 56

List of Tables

2.1 Increased global distribution of *P. ficus* from 1994 to 2007. 9
2.2 GenBank sequence data currently available for *P. ficus* (Signoret). 13

4.1 Concentrations, purity parameters and quantities of total RNA and mRNA used for Primary- and Fractionated Library construction. . 42
4.2 Recovery of the purified size-fractions before and after *Sfi*I digestion. 52
4.3 Calculated titers of the four Fractionated Libraries. 53
4.4 Clones containing potential target sequences for an RNAi-mediated approach to *P. ficus* pest control. 57

A.1 RNase-inactivating compounds suitable for use in molecular biology. 64

B.1 RNA extraction protocols found to be less suitable for use on mealybugs 65

G.1 Sequences with matches in the non-redundant databases of the NCBI, but not in RefSeq (E-value $\leqslant 10^{-5}$). 83
G.2 Sequences for which a match was found in a query against all GenBank non-mouse and non-human EST entries (E-value $\leqslant 10^{-5}$). . . 84
G.3 Sequences for which matches were obtained from the RefSeq databases of the NCBI (E-value $\leqslant 10^{-5}$) 85

Abbreviations

A	Adenine
$A_{260/230}$	Ratio of absorption at 260 nm and 230 nm wavelength
$A_{260/280}$	Ratio of absorptionat 260 nm and 280 nm wavelength
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment and Search Tool
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
cds	Coding sequence
cDNA	Complementary DNA
cfu	Colony forming units
CsCl	Cesium chloride
ddH_2O	Double distilled H_2O
DEPC	Diethylpyrocarbonate
dH_2O	Distilled H_2O
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra-acetic acid di-sodium salt
EST	Expressed sequence tag
EtOH	Ethanol
G	Guanine
GLD	Grapevine leafroll disease
GLRaV-1	Grapevine leafroll-associated virus 1
GLRaV-3	Grapevine leafroll-associated virus 3
GVA	Grapevine virus A
GVB	Grapevine virus B
HCl	Hydrochloric acid
IPM	Integrated pest management
kb	kilobase pairs
LB	Luria-Berthani
min	Minutes
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information

nt	Nucleotides
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PTGS	Post-transcriptional gene silencing
PVPP	Polyvinylpolypyrrolidone
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	revolutions per minute
rRNA	Ribosomal RNA
RT	Reverse transcription
SAP	Shrimp alkaline phosphatase
SDS	Sodium dodecyl sulphate
seq	sequence
SIT	Sterile Insect Technique
SMART	Switching Mechanism At 5' end of RNA Transcript
T	Thymine
TAE	Tris/acetic acid/EDTA
Tris	Tris(hydroxymethyl)aminomethane
U	Units
UV	Ultraviolet
V	Volt
v/v	volume per volume
w/v	weight per volume
YT	Yeast tryptone

Chapter 1

Introduction

1.1 General Introduction

The grape and wine industry is an important source of economic growth and development in South Africa. According to the South African Wine Industry Information and Systems (SAWIS), the sector contributed roughly R22.5 billion to South African gross domestic product (GDP) in 2004, and 3.5% to national agricultural production in 2005. Grapes comprised 33% of the country's deciduous fruit harvest in 2005, and contributed 5.3% to its total earnings from agricultural exports. In 2006, 38.3% of all wines produced in the country were exported, ranking South Africa fourth in new world wine sales and ninth in global wine production[1]. Adding to this is the sector's contribution to full time employment (amounting to 257 000 people in 2004) and its importance to the tourism industry, on which South Africa relies heavily as a source of capital influx. Furthermore, nearly 19% of the household income generated annually is destined for lower income groups and, according to a recent survey, the sector is imperative to the optimal functioning of nearly 65% of business activities in the Western Cape (http://www.sawis.co.za). Taken together, these statistics illustrate the importance of the stability and well-being of the wine industry to the economic well-being of South Africa.

Like any agricultural industry, the grape and wine industry is vulnerable to the adverse effects of pathogens and pests on the crops they cultivate. According to Martelli, grapevine are susceptible to more potential pathogens than any other woody perennial (at least 47 viruses, 5 viroids and 8 prokaryotes have been recorded)[2]. Added to this is their susceptibility to external pests such as nematodes and insects. Of all of these, the most devastating are the viruses. Viral diseases pose a special problem to the grape and wine industry, as infected plants are currently impossible to cure and no resistant cultivars have been discovered to date[3].

A grapevine viral disease of special interest is grapevine leafroll disease (GLD)[†]. This disease is currently regarded as one of the principal antagonists of grapevine worldwide, and is the most destructive of diseases in South Africa[6,7]. GLD can spread at an alarming rate, as was illustrated by its increased incidence from 1 820 to 23 425 infected vines in a monitoring block of 60 vineyards in the Western Cape (2001 to 2005, Pietersen pers. comm.).

Many viruses are implicated in the etiology of GLD, but the primary causative agent has been identified as grapevine leafroll-associated virus 3 (GLRaV-3)[8]. Infection with GLRaV-3 stems mainly from two sources: *(i)* graft transmission from infected to uninfected vines, and *(ii)* in-field transmission by insect vectors. Although the use of virus-free propagation material has greatly reduced the incidence of GLD in South African vineyards, phytosanitation does not provide protection against reinfection of materials in the field. Consequently vector control has become an area of great importance.

The vine mealybug, *Planococcus ficus* (Signoret), is the main vector for GLRaV-3 in South Africa[7]. It is also a vector of viruses implicated in Shiraz disease, corky bark disease and Kober stem grooving[2]. Apart from spreading viral diseases, *P. ficus* creates favorable conditions for fungal growth, and is capable of inducing grave damage to grapevine of its own accord. In addition, grape bunches that have become infested by the pest are unmarketable. In view of these adversities, *P. ficus* has become a renowned pest of grapevine not only in South Africa, but also in other wine and grape producing countries around the world.

[†]GLD is a complex disease of which the most important adverse affects are a progressive reduction in berry yield (losses of up to 50% per year), delayed fruit maturity, loss of grape pigmentation, diminished quality of must, and reduced vine vigor[4,5].

Effective control of *P. ficus* is currently a main priority of the viticulture industry, both locally and globally. However, like GLD, it is an exceptionally tenacious antagonist of grapevine. The adverse effects it has on grapevine, along with its resistance to both chemical and biological control measures has earned it the top rank on the American Vineyard Foundation's list of major pests of grapevine[9]. Despite implementation of all available control strategies, the pest has spread from four geographical areas spanning twenty countries in 1994, to five geographical areas spanning thirty-nine countries in 2007 (Table 2.1, Chapter 2).

Biotechnology is currently a valuable resource of novel and sustainable insect pest control strategies. Promising fields in this regard are the production of sterile males, mass releases of males carrying female-specific lethal alleles and the engineering of insect-resistant crops (detailed discussion in Chapter 2, Section 2.2). These strategies are based on the manipulation of genetic sequences, however, and their implementation is dependent on the availability of appropriate target sequence information. Currently there is very little sequence information available for *P. ficus*.

1.2 Project Proposal

The Masters study forms an integral part of a larger research project, the goal of which is to effectively control the spread of grapevine leafroll disease (GLD) in grapevine. The current approach is to eliminate the main viral vector, *Planococcus ficus* (Signoret), by means of an RNAi-mediated approach to pest control.

The aim of this project was to create a resource of *P. ficus* expressible coding sequences from which a transcript suitable for RNAi-mediated pest control could be obtained.

Since cDNA libraries provide reliable long-term resources of genetic information, the construction of such a library was proposed[†]. Establishing such a library is an advantage to a laboratory, as it provides a very good platform

[†]Although high-throughput sequencing could have delivered many sequences in a short amount of time, suitable technology (i.e. LCM-454 technology) only became available in South Africa by the end of the study.

from which to characterize gene structures, develop nucleic acid probes and express proteins of interest.

A commercial kit was chosen for cDNA library construction, following the recommendations of two of the current authoritative molecular biology compendiums[10,11]. The Creator™ SMART™ cDNA Library Construction Kit (Clontech) was selected, as it uses technology which greatly enhances the representation of full-length cDNA clones in the library and is suitable for use on limited quantities of starting material. It also involves little enzymatic manipulation of mRNA and consequently has a lower risk of template degradation prior to cDNA synthesis. In addition, the technology is designed to streamline downstream manipulation and characterization of inserts[12].

Our approach was to construct a whole organism cDNA library from female mealybugs primarily, as this is the only sex capable of transmitting grapevine viral diseases. Inclusion of mealybug egg sacks and younger instars was proposed as a method to ensure that mealybug developmental genes would also be represented in the library.

Specific tasks of this study were:

- To find and optimize a suitable protocol for RNA extraction from mealybugs.
- Synthesize cDNA from the purified mRNA.
- Ligate the cDNA into suitable cloning vectors.
- Transform host cells with these vectors.
- Titer the library and screen its clones to verify cDNA incorporation.
- Amplify the library to secure the stability of all transformed cells.
- Carry out EST sequencing on a selection of clones to elucidate the content of the library.
- Identify EST's that could be instrumental to the design of a RNAi-based pest control strategy.

Chapter 2

Literature Review

2.1 The vine mealybug, *Planococcus ficus* (Signoret)

As is the case for most insect pest species, *P. ficus* has been described, renamed and reclassified repeatedly. The first recorded sighting was in 1869, when it was erroneously identified as *Coccus vitis* Linnaeous[13]. The error was corrected the following year, when it was classified as a new pest species, *Dactylopius vitis* Lichtenstein. It was renamed at least seven times more, before arriving at the currently accepted *Planococcus ficus* (Signoret) (Ezzat & McConnel, 1956 *teste*[14]). According to the latest taxonomic classification, *P. ficus* is of the Order Hemiptera, Superorder Homoptera, Superfamily Coccoidea, and Family Pseudococcidae[13].

In parallel to its large array of scientific synonyms, *P. ficus* is also referred to by a variety of colloquial names. The most frequently used are: Mediterranean vine mealybug, subterranean vine mealybug, vine mealybug and grapevine mealybug. To confuse matters even more, the similarly named grape mealybug, refers to a closely related species, *Pseudococcus maritimus* (Ehrhorn)[15].

2.1.1 Identification in South African Vineyards

The first recorded sighting of *P. ficus* in a South African vineyard was in 1914, when the insect was identified by its then current synonym, *Pseudococcus vitis*[14]. It was declared a pest of grapevine in 1943 under the misnomer *Planococcus citri* (Risso) (Joubert, 1943 *teste*[15]). Subsequent studies of its biology[16], predators and parasites[17] were performed under the same name. The insect was finally correctly identified as *Planococcus ficus* (Signoret) in 1975[14]. Its status as a key pest in the South African grape and wine industry was affirmed after a recent survey of the mealybug species in the major grape producing provinces of the country[18].

2.1.2 Biology

In his masters dissertation, *'n Bydrae tot die kennis van Planococcus citri (Risso) (Homoptera: Pseudococcidae)**, Kriegler (1954) provided a detailed description of the biology and economic importance of *Planococcus ficus* in South African vineyards. Unless otherwise stated, the following information was derived from this dissertation.

2.1.2.1 Morphometrics and Life Cycle

Like most coccids, the vine mealybug is sexually dimorphic. Female mealybugs are ovate, flesh- to slate grey-colored and covered in a white powdery wax that protrudes into thick uniform filaments around the edge of the body (Figure 2.1). Before reaching maturity they undergo incomplete metamorphosis, and pass through three distinct nymphal stages. The fully matured female is clearly segmented and approximately 4 mm long, 1.5 mm thick and 2 mm wide. In contrast, males undergo a more complete metamorphosis, with the penultimate stage a pseudopupa. They are extremely small (∼1 mm body length), have no mouth parts, a single pair of wings on the metothorax, and two long filamentous anal setae †.

*English translation: *A contribution to the knowledge of Planococcus citri (Risso) (Homoptera: Pseudococcidae)*

†Due to the high mobility and extremely small size of males, mealybug identifications are usually done on females[19]. Species can be distinguished by various characteristics such as body shape, length, color, and the amount and appearance of the waxy filaments that

(a) Mature female **(b)** Mature male

Figure 2.1: The female and male forms of the vine mealybug, *Planococcus ficus* (Signoret) - courtesy of Pia Addison, Conservation Ecology & Entomology, SU.

The vine mealybug, in contrast to some other mealybug species, does not diapause. All life stages can be found in all seasons, and populations can achieve up to six generations per year. Optimal progress through the developmental stages is achieved around 25-27°C, with the pre-oviposition period as the determining factor (3.44 days at 25°C versus 15.79 days at 27°C, or 36.11 days at 20°C)[20].

2.1.2.2 Seasonal Population Dynamics

The distribution and dynamics of mealybug populations in natural environments are closely correlated with ambient temperature, presence or absence of natural enemies, and the availability of nutrient rich food.

Populations usually peak in spring time when the ambient temperature is optimal for reproduction and natural enemies are still scarce. In a pesticide-free environment, the population will typically decrease to its annual minimum in late summer as a result of peak predation and parasitism. This is usually followed by a slight increase in autumn, and a relatively stable population dynamic throughout the winter[17].

surround the female body. However, *P. ficus* and its closest relative, *P. citri*, are extremely difficult to distinguish in this manner and misidentifications are often made.

Vine mealybug populations spend the winter in colonies under the bark of the lower, and underground regions of infected vines. In South Africa, upward migration of the younger stages (first and second female instars) typically commence in mid September when colonies are formed at the nutrient rich bases of shoots and young buds. Populations peak in mid November, and by December are found mainly on the foliage. Mealybugs start infesting the grape bunches late in January, where they feed on the abundance of plant sap and nutrients available. After harvest the majority of these populations move back onto the leaves and, when these start to dry, start moving back to the lower stem and rootstock.

2.1.2.3 Dispersal

Technically mealybugs have limited power of dispersal as females are wingless and their movement is restricted to neighbouring vines. Despite this, *P. ficus* can be found in most of the wine producing countries around the world, and its distribution has more than doubled over the past 13 years (Table 2.1, data from[13] and unpublished data (Yair Ben-Dov, 2008)). Long distance dispersal is usually the result of unsanitary pruning and harvesting practices, as well as the distribution of infested rootstocks and grafting material. Shorter distances are covered when strong winds (and possibly birds) carry younger instars and contaminated leaves to neighboring blocks.

2.1.3 Detriments of Infestation

Mealybugs cause damage to grapevines by *(i)* weakening the vines , *(ii)* creating favourable conditions for fungal growth , and *(iii)* spreading viral diseases. They are of economic importance not only due to the damage they inflict on vines however, but also because of the reduced marketability of grape bunches they have infested.

(i) Weakening of Vines: Large scale infestations of mealybug can cause progressive weakening of grapevine as a result of early defoliation and a loss of nutrients due to excessive feeding. Vines that suffer prolonged exposure to high levels of infestation can lose their vigour and die[21].

Table 2.1: Increased global distribution of *P. ficus* from 1994 to 2007.

Distribution in 1994[13]	
Oriental Region	India; Pakistan
Ethiopian Region	South Africa
Nearctic Region	None
Neotropical Region	Argentina (Tucuman); Brazil
Palearctic Region	Azerbaijan; Canary Islands; Cyprus; Egypt; France; Greece; Iran; Iraq; Israel; Italy; Lebanon; Libya; Saudi Arabia; Spain; Tunisia
Distribution in 2007 (unpublished data by Y. Ben-Dov, 2008)	
Oriental Region	India; Pakistan
Afrotropical Region	South Africa; Mauritius
Nearctic Region	United States of America (Alabama, California, Florida, Georgia, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Texas)
Neotropical Region	Argentina (Catamarca, La Rioja, Mendoza, San Juan, Tucuman); Brazil (Rio Grande do Sul); Chile; Dominican Republic; Trinidad and Tobago (Trinidad); Uruguay
Palearctic Region	Afghanistan; Azerbaijan; Azores; Canary Islands; Corsica; Crete; Cyprus; Egypt; France; Greece; Iran; Iraq; Israel; Italy; Jordan; Lebanon; Libya; Majorca; Portugal; Sardinia; Saudi Arabia; Sicily; Spain; Syria; Tunisia; Turkey; Turkmenistan

(ii) Favouring Fungal Growth: Like most sap-sucking insects, mealybugs secrete large quantities of honeydew. Apart from the cosmetic problem it poses for the table grape industry, honeydew also serves as a substrate for a saprophytic fungus called sooty mold. Although not damaging to the plant in itself, the mold can greatly reduce plant vigour when covering the foliage to such an extent that it interferes with photosynthesis. Grape bunches covered with sooty mould and mealybugs are also virtually unmarketable - especially in the table grape industry where bunch presentation is of utmost importance.

(iii) Spread of Viral Diseases: Perhaps the most important adverse effect of mealybugs on grapevine lies in their capacity to distribute phloem-bound viral diseases[22,23,24]. The vine mealybug has been shown to transmit grapevine leafroll-associated virus type 1 (GLRaV-1), grapevine leafroll-associated virus type 3 (GLRaV-3), grapevine virus A (GVA), and grapevine virus B (GVB). These viruses are implicated in grapevine diseases such as Kober stem grooving and shiraz disease (GVA), corky bark disease (GVB), and grapevine leafroll disease (GLRaV-1 and -3)[2].

2.1.4 Mealybug Control

The vine mealybug is an exceptionally obstinate pest of grapevine as part of its population remains below ground and in bark crevices, where it is sheltered from pesticides and natural enemies. With a broader awareness of the adverse effects of pesticides on the environment and the inevitable development of resistance in insects, the control of mealybugs has moved to a more integrated approach. This entails the careful coordination of a variety of cultural, biological and chemical measures to complement each other in a year round pest management program[25,26,15,27].

2.1.4.1 Chemical Control

Chemical control of mealybugs used to rely heavily on delayed dormant organophosphates (e.g. chlorpyrifos and tokuthion), short-residual organophosphates (e.g. mevinphos and diazinon) and carbamates (methomyl)[28,29]. Due to their unselectiveness, these compounds were not only detrimental to the proliferation of mealybug populations, but also to that of their natural enemies[30,19]. Reduced-risk insecticides such as the neonicotinoids (imidacloprid) and insect growth regulators (buprofesin and pyriproxyfen) are the present choice. Although not necessarily more beneficial to the broader environment, these compounds are more accommodating towards the predators and parasitoids of mealybugs and are therefore more compliable with integrated pest management (IPM) programs[31].

2.1.4.2 Biological Control

(i) Predators and parasitoids are among the most effective biological control agents of mealybugs. Predatory Coccinellid beetles (most of which are from the genus *Nephus*), and Hymenopteran parasites like *Coccidoxenoides perminutus* (Timberlake), *Anagyrus pseudococci* (Girault) and *Leptomastix dactylopii* (Howard) are the most prevalent enemies of *P. ficus* in South African vineyards[32]. Augmentation of natural populations is achieved through scheduled releases of mass reared populations early in the breeding season. However, parasitoids can only efficiently target mealybugs when they are found on ex-

posed locations and are unable to reach them when they are hidden in the deep crevices of the lower and underground rootstocks of grapevine.

(ii) The mealybug sex pheromone was recently identified. A synthetic version, racemic lavandulyl senecioate, is currently used in sticky pheromone traps which are used to monitor the development of *P. ficus* infestations in vineyards[33]. These traps are not intended as a pest control strategy *per se*, but provide information about the the most appropriate times at which to implement actual pest control strategies. Recent attempts at mating disruption through the application of a microencapsulated formulation of the sex pheromone were ineffective, but increased parasitism by *Anagyrus* was observed with possible implications for enhanced parasitoid augmentation[28,34].

(iii) Cultural control of mealybugs is designed to hinder the spread of existing infestations through infested-, and to uninfested vineyards[19,35,16]. Sterilisation of pruning and harvesting equipment and the use of heat-treated nursery stocks are the most important precautions taken. Nursery material intended for vine propagation is treated by sequential immersions in water baths of 30°C (pre-warming), 52.8°C (heat treatment) and 23°C (cooling). According to a recent study, this method of sanitation is 99.8-100% effective, but is not recommended as a cure for evidently infested materials[36]. Rather, it is regarded as a precaution to elliminate inconspicuous mealybug infestations in propagation material that is destined for dispersal.

2.2 The need for Novel Approaches to Insect Pest Control

Although IPM strategies can be very efficient when correctly administrated, the sad reality is that most of the time, this is not the case. The injudicous implementation of a single strategy can thwart the success of a whole IPM program. Adding to this are the limitations inherent to the pest control strategies themselves. Over the past 20 years, biotechnology has become a valuable resource of novel and sustainable pest control strategies. Three of the most promising fields are the production of sterile insects, mass releases of males carrying female-specific lethal alleles and the engineering of insect-resistant crops.

(i) Sterile Insect Technique (SIT): Successful reproduction of harmful insect species can be greatly reduced through mass releases of sterile males early in the mating season. Males can be sterilized either by radiation or, more recently, through genetic engineering[3].

(ii) Release of Insects carrying a Dominant Lethal (RIDL): A strategy which closely resembles SIT, but which involves mass releases of males homozygous for a female-specific dominant lethal allele. This strategy has the advantage that a single release will not only affect the first generation of its target population, but also penetrate the subsequent generations[37]. Computational analysis of this strategy indicated that such releases would not only provide good control of pest populations, but could also significantly reduce the percentage of refuge crops necessary for insect resistance management in transgenic crops[37].

(iii) Insect Resistance through Genetic Engineering: Transformation of commercially important crops with insecticidal genes has revolutionized pest management programs in agriculture[38]. Conventionally, plants are transformed with entomopathic genes that are normally expressed in other organisms[39,40]. A more advanced approach is to make use of post-transcriptional gene silencing (PTGS)[41]. An advantage of this approach is that it can be tailored to be highly specific for the target organism. Attaining resistance through PTGS is still a very young contribution to the field and no plants have been released for commercial production[41]. However, successful implementation of the technology was recently reported for in-lab experiments[42,43,44].

Apart from male sterilization through radiation (which is incidentally not applicable to mealybugs[‡]), all of the approaches described above are dependent on genetic engineering. More specifically, they are focussed on the expressible components of target organism genomes. If any of these methods are to be used for mealybug pest control, it is imperative that knowledge about the expressible genes of the insect is available.

[‡]In contrast to other insects on which this technique has been used, mealybugs have been shown to survive high doses of ionizing radiation and still reproduce[45].

2.3 Current Status of the Genetic Information available for *Planococcus ficus* (Signoret)

Very little information is currently available on the genes expressed by *P. ficus*. A search of the available sequence data on GenBank (http://www.ncbi.nlm.nih.gov) provided only eleven entries, four of which were derived from mitochondrial DNA (including one concatenated sequence), three from ribosomal RNA, two from the primary- and secondary endosymbionts of *P. ficus* and two from its own elongation factor 1α (Table 2.2). As similar sequences are also expressed in plants and other organisms, these sequences are not suitable for pest control through PTGS. They are also not suitable targets for the induction of male sterility, or female specific lethality in mealybugs. More suitable expressible genetic information is therefore needed.

Table 2.2: GenBank sequence data currently available for *P. ficus* (Signoret).

Accession number	**Gene product description**	**Sequence description**	**Source origin**
EU250515	elongation factor 1α (EF-1α)	partial cds	genomic DNA
EU250573	cytochrome c oxidase subunit 1 (COI)	partial cds	mitochondrial
DQ238220	cytochrome c oxidase subunit 1 (COI)	partial cds	mitochondrial
DQ238218	cytochrome c oxidase subunit 1 (COI)	partial cds	mitochondrial
AY691420	cytochrome b (cytb) NADH dehydrogenase subunit 1 (nd1) large subunit ribosomal RNA small subunit ribosomal RNA	partial seq complete seq complete seq partial seq	mitochondrial mitochondrial mitochondrial mitochondrial
AY427341	28S large subunit ribosomal RNA	partial seq	genomic DNA
AY426055	18S small subunit ribosomal RNA	partial seq	genomic DNA
AY427405	28S large subunit ribosomal RNA	partial seq	genomic DNA
AY427233	elongation factor 1α (EF-1α)	partial cds	genomic DNA
AF476092	16S and 23S ribosomal RNA	partial seq	Candidatus Tremblaya princeps
AF476108	16S and 23S ribosomal RNA	partial seq	Secondary endosymbiont

2.4 Technologies by which Expressible Genetic Sequence Information can be obtained

High-throughput transcript sequencing and cDNA library construction are two of the most frequently cited current technologies by which novel expressible genetic sequence information can be obtained. Both of these approaches have their merits and shortcomings. Whereas gene discovery through high-throughput sequencing follows a "from-information-to-clone" approach, cDNA libraries represent a "from-clone-to-information" approach. Depending on the research question at hand, a different approach might be more appropriate.

(i) High-throughput Sequencing : High-throughput sequencing of cDNA-ends is a very efficient way in which to obtain large quantities of sequence information in a very short time. Short reads of up to 300 bases can be obtained for each end of a cDNA molecule within a matter of hours[46]. Unfortunately the technology is not advanced enough to create full length sequences yet and the middle portions of larger transcripts are not represented in the data obtained. Another disadvantage is that, after interesting transcripts have been identified by data analysis, their corresponding hard copies still need to be isolated from the source material. An advantage of the technology is that low abundance transcripts are better represented in these data than what they are in conventional cDNA libraries[47].

(ii) Complementary DNA Libraries: cDNA libraries are physical archives of the genes that are transcribed in living organisms[10]. Although not able to compete with the high rate of information generation which is the trademark of high-throughput sequencing, obtaining expressible sequence information by means of cDNA libraries has its own merits. The most important advantage is that, when a transcript of interest is discovered in a cDNA library, the cloned sequence is physically available and analysis and/or manipulation thereof can directly commence. Another advantage is that nucleic acid probes can be used to quickly identify cDNA clones which contain inserts of interest if genetic sequence information is already available for a similar transcript in a related organism. Furthermore, full-length cDNA sequences can be obtained from cDNA libraries and if a library was constructed using expression vectors, the proteins for which they encode can be directly expressed and isolated. A disadvantage of the method is that low abundance transcripts are often lost

during the many steps of the library construction protocol. Library construction is also a more labour intensive approach than high-throughput cDNA-end sequencing.

A NOTE ON THE FEASIBILITY OF USING GENOMIC LIBRARIES FOR GENE DISCOVERY: Genomic libraries are constructed from chromosomal DNA and contain large stretches of non-coding sequence[11]. Although genes can be discovered in these libraries, their coding regions are mostly interrupted by one or more introns, and therefore the length of a gene-coding sequence can be quite long. Because of the diffuse prevalence of coding regions in genomic libraries, they are not an optimal source for gene discovery[48]. Nonetheless, they are a valuable asset when characterizing genes, as the promotor region and intron-exon boundaries of a transcript of interest can be identified in these libraries[48].

Both high-throughput sequencing and cDNA libraries make use of ribonucleic acids (RNAs) as the initial material from which sequence information is retrieved. To obtain the best results from either of these strategies, it is imperative that good quality RNA be obtained from the organism under investigation.

2.5 Isolating Ribonucleic Acids

The ostensible indestructibility and ubiquity of RNA degrading enzymes (RNases) can make RNA extractions - to say the very least - a taxing experience. This does not need to be so. When proper equipment and technique is used, RNA extractions can be performed without any setbacks. Depending on the type of tissue and the research question at hand, different approaches to RNA extractions are assumed. Factors that influence the choice of a protocol are: The type and amount of RNA required, the required purity of the sample and the necessity of integrity. These factors are addressed in the five components of an RNA extraction protocol. When the RNA has been purified, it is important to assess its quality before continuing with downstream applications.

2.5.1 Membrane Solubilization

The first step to RNA purification is to disrupt the cell walls and membranes that enclose the cytoplasm and organelles of cells. Cell walls can be sheared mechanically by grinding samples in a mortar and pestle. Alternatively, non-ionic, hypotonic buffers can be used to lyse membranes with minimal disruption of sub-cellular organelles (i.e. nuclei, mitochondria and chloroplasts). These organelles can then be separated from the cytoplasmic lysate by differential centrifugation, and RNA extractions can be continued either on the cytoplasmic fraction, or on the isolated organelles. If there is no need for compartmental RNA isolation, membranes can be disrupted by harsher reagents like sodium dodecyl sulphate (SDS) or guanidinium thiocyanate.

2.5.2 Inhibition of Ribonucleases

The key to working with RNA is to minimize RNase activity. This can be achieved through the use of RNase-inactivating compounds, and by avoiding reintroduction of RNAses after purification. Since all RNA extraction buffers contain one of more RNase-inactivating compounds, RNA is protected from endogenous RNases while in the extraction buffer, but becomes vulnerable again as soon as it is purified from the buffer. Consequently all equipment used for RNA work should be treated to make them RNase-free and RNase-inactivating enzymes such as RNasin should be incorporated in protocols whenever necessary (see Appendix A, p. 63 for a detailed description of RNase-inactivation).

2.5.3 RNA Separation from Cellular Lysates

RNA can be isolated from cellular lysates by aqueous-organic phase separation, or by density gradient (isopycnic) centrifugation. Most of the time both of these methods are followed-up with purification and recovery by precipitation.

When extracting RNA by means of *aqueous-organic phase separation*, it is important to realize that nucleic acid partitioning is highly dependent on temperature and pH. Whereas only RNA will partition into the aqueous phase under acidic conditions, both RNA and DNA will remain in the aqueous phase in alkaline conditions[10]. For this reason phenol should be saturated to a low pH

(∼4.5) before RNA extractions are performed. Also, in pure phenolic preparations mRNA partitions in the aqueous phase at low temperatures, in the interface at room temperature, and in the aqueous phase at high (60°C) temperature. Addition of chloroform facilitates RNA partitioning in the aqueous phase[49].

RNA isolation by isopycnic centrifugation is usually performed in a density gradient of either cesium chloride (CsCl), cesium trifluoroacetate (CsTFA) or cesium sulfate (Cs_2SO_4). The technique relies on discrepancies in the boyant densities of macromolecules. Molecules in the cellular lysate migrate through the density gradient until they reach a density similar to their own, at which point they accumulate. Since both proteins and DNA have lower boyant densities than RNA, they segregate at an earlier stage in the gradient. The maximum density of the gradient is usually designed to be slightly less than that of RNA, and so the molecules accumulate as a pellet in the bottom of the ultracentrifuge tube.

Isolating messenger RNA from total RNA preparations: Both cDNA libraries and the cDNA-ends used for high-throughput sequencing are conventionally created only from the protein encoding messenger RNA (mRNA) population of the transcriptome[10]. mRNA constitutes a very small proportion (approximately 1-5%) of the total cellular RNA in eukaryotes[50]. It can be isolated from total RNA preparations by oligo(dT) selection, because poly(A)$^+$ mRNAs bind with oligo(dT) sequences under high-salt conditions[10]. A variety of commercial kits make use of this characteristic by covalently binding oligo(dT)-sequences to stationary materials such as magnetic beads, cellulose membranes and other inert polymers. Total RNA is usually washed over these materials at high-salt conditions, after which the unbound nucleic acids are removed. mRNA is then released by lowering the ionic strength[10].

2.5.4 Storage of Purified RNA

The correct best way to store RNA is often a source of major debate. Most researchers agree that long term storage should be done at -80°C, while short term storage can be done at -20°C. Storage in highly purified 100% formamide is a reliable option for long term storage of RNA samples[51]. Good protection is also provided if RNA is stored at -80°C as an ethanol precipitate[52]. Pellets

dissolved in RNase-free water or TE buffer (10 mM Tris, 0.1 mM EDTA, pH 7.5) are more prone to degradation, but samples are ready for use in most downstream applications[10].

2.5.5 Quantification and Quality Assessment of RNA

Both the purity and the quantity of RNA can be assessed by spectrophotometry. RNA purity is assessed by calculating the ratio of sample absorbance at wavelengths of 260 nm and 280 nm. An absorption ratio of $1.8\pm 0.1(A_{260/A280})$ is generally accepted as a sign of good purity. Secondary confirmation is provided by the $A_{260/A230}$ ratio, which should ideally be 2.0 ± 0.1[53]. The quantity of RNA in a sample is calculated from its absorbance at 260 nm. In a pure sample, an optical density reading of 1 corresponds to a concentration of approximately 40 µg/ml RNA.

RNA integrity can be assessed either by agarose gel-electrophoresis or by analysis on an Agilent® BioAnalyzer[54]. Gel-electrophoresis should be performed under denaturing conditions, as RNA secondary structures can alter the electrophoresis profile. Normally samples are run on denaturing agarose gels[54], but non-denaturing gels can also be used if the sample is denatured in a suitable loading dye prior to electrophoresis[10]. The objective of all of these procedures is to determine the intensity, and the distinction, of the fluorescence signals obtained from the 18S and 28S ribosomal RNA (rRNA) fragments. A electrophoresis profile with clear, distinct bands exhibiting a fluorescence ratio of 2 to 1 (28S rRNA to 18S rRNA) is generally accepted as an indication of good RNA integrity. Only RNA that has good integrity is suitable for cDNA library construction and/or high-throughput transcript sequencing.

2.6 Constructing cDNA Libraries

There are a vast number of ways in which to construct cDNA libraries. All methods are conducted in a stepwise fashion, however, and three essential procedures are performed in each protocol: First-strand synthesis, second-strand syntheses and cDNA cloning[11,10,55]. Other procedures such as cDNA amplification, size fractionation, normalization and/or subtractive hybridization can

be incorporated in the protocol in order to establish different representation of fragments[56,57,11]. After a library has been constructed, information can be retrieved by any of a number of screening methodologies[†].

2.6.1 First-strand Synthesis

First-strand synthesis (also referred to as "reverse transcription") is the first critical step of any cDNA synthesis reaction. It entails the use of a RNA-dependent DNA polymerase enzyme to create a single stranded DNA molecule that is complementary to the RNA template it was made from[55]. The two most frequently used polymerases are AMV (derived form the avian myeloblastosis virus) and MMLV (Moloney strain of murine leukemia virus)[58,10]. These enzymes differ with regards to their fidelity, processivity, RNase H activity and the temperatures at which they optimally function[11]. Modified enzymes with improved functionality are most frequently used, but in general AMV reverse transcriptase is active at 50°C and higher, while MMLV reverse transcriptase functions only between 37 and 40°C. AMV reverse transcriptase is also more processive than MMLV reverse transcriptase, but has higher levels of RNase H activity, making it a more risky choice when full length transcripts are needed[10]. The primers used for reverse transcription are mostly either homopolymers of thymine nucleotides (oligo(dT) primers)[‡], or randomly assembled combinations of six nucleotides (random hexamers)[59,10]. Whereas oligo(dT) primers anneal to the poly(A) tails of mRNA transcripts, random hexamers initiate strand synthesis at various locations along the template[55]. Since poly(A) tails can be very long, oligo(dT) primers are frequently anchored to the 5'-most end of the poly(A) tail by incorporating a single A, C or G nucleotide to the 3'-end of the primer[10]. The 5'-ends of these primers are also frequently linked to adapter sequences which contain restriction enzyme recognition sequences that are useful for site-directed fragment cloning[10].

[†]It is not within the scope of this dissertation to provide a detailed description of all the available cDNA library technologies. Hence, a brief description of the most essential steps are provided, and the technology that was used in this study is outlined at the end of the chapter.

[‡]The one-letter abbreviations for the deoxyribibonucleotides are: A (adenine), C (cytosine), G (guanine) and T (thymine)

2.6.2 Second-strand Synthesis

Second-strand synthesis entails the replacement of the RNA template with a DNA strand that is complementary to the first DNA strand synthesized[10]. This is acchieved through the use of a DNA-dependent DNA polymerase, of which numerous formulations are available[11]. At the inception of cDNA library construction technology, second-strand synthesis was primed by a transient hairpin loop that forms at the 3'-end of a first-strand molecule[11]. Removal of hairpins after second-strand synthesis resulted in a loss of sequence information, and was often accompanied by fragment degradation[10,60]. A subsequent approach was to include RNase H in second-strand synthesis reactions. Since this enzyme nicks the RNA strand in a DNA:RNA hybrid molecule, free 3'-OH groups are made available and serve as primers for DNA synthesis[60]. A second-strand that is primed in this way is fragmented and shouldering DNA fragments must be ligated to obtain a single covalently bound DNA strand[11]. Also, cDNA fragments created in this manner do not include the bases corresponding to the 5'-end of the original RNA transcript, as priming of the extreme 5'-end of a cDNA fragment is a very rare occurence when using this approach[61].

Modern methodologies of cDNA library construction are designed to obtain the full length transcript sequence[55]. While 3'-ends are easily obtainable by oligo(dT)-priming, 5'-ends are often lost. This can be ascribed in part to the limited capabilities of reverse transcriptases (which stall at secondary structures and often terminate strand synthesis before reaching the 5'-end), and in part to suboptimal second-strand priming[61]. A number of methods have been developed to improve 5'-end representation in cDNA libraries - most of these are modifications of the original protocol for Rapid Amplification of cDNA Ends (RACE)[62]. In principle 5'-ends are incorporated in cDNA fragments by flanking the 5'-end of the original template sequence with a sequence to which the primers of the second-strand reaction could anneal. Flanking can be accomplished by ligating an adapter fragment to the 5'-end of the mRNA transcript, or to the 3'-end of the first strand[61]. Another approach is to make use of terminal deoxynucleotidyl transferase activity and the template switching capabilities of reverse transcriptases[63,64]. The number of full length transcripts from which cDNA synthesis commences can be enhanced prior to first-strand synthesis by 5'cap affinity selection technologies[65].

2.6.3 cDNA Cloning

There are two types of vector into which cDNA fragments can be ligated, through which host cells can be transformed. Viral vectors (or phage vectors) can be packaged into bacteriophages, which are used to transform bacterial cells[66,11]. Plasmids are used to directly transform bacterial cells through electroporation[‡]. The choice of a cloning vector depends on the type of research to be done. Phage libraries have much higher transformation efficiencies and far more plaques can be screened per single plating[11]. Plasmid libraries are less complex to construct and are better suited for *de novo* gene identification, as it is easier to isolate a single cDNA fragment (the template for a sequencing reaction) from a plasmid library than from a phage library[68].

2.6.4 Library Screening

cDNA libraries are conventionally screened either by molecular hybridization techniques, or by sequencing the inserts of selected clones[48]. Whereas sequencing is used to obtain novel coding sequence information, hybridization techniques are used to identify clones that contain sequences which have formerly been determined to be of interest. Nucleic acid probes can be used to identify clones that contain complementary and/or closely related sequences[11,48]. Alternatively, if part of the sequence of a cDNA fragment of interest is already available, other approaches such as PCR screening of arrayed cDNA library pools[69], large-scale concatenated cDNA sequencing (CCS)[70], or one of the inverse PCR strategies such as MACH-1, MACH-2 and SLIP can be used to obtain the rest of the relevant sequence[71,72]. cDNA libraries that were transformed into expression vectors can also be screened with antibodies to identify clones that contain the coding sequence for a particular protein[48].

[‡]Bacterial cells can also be transformed chemically, but as this technique exhibits much lower transformation efficiencies electroporation is the method of choice when constructing libraries[67].

2.7 Creator™ SMART™ cDNA library Construction Technology

The following is a brief description of the main distinguishing features of cDNA synthesis by means of the SMART™ (Switching Mechanism At the 5'-end of RNA Transcript) protocol. All information was obtained from the Creator™ SMART™ cDNA Library Construction Kit User Manual, and the information provided on the Clontech website (www.clontech.com).

Creator™ SMART™ cDNA library construction technology makes use of the combined terminal deoxynucleotidyl transferase activity and template switching capabilities of PowerScript™ reverse transcriptase in order to eliminate the need for adapter ligation to obtain full length cDNA fragments. First-strand synthesis is primed by a lock-docking oligo(dT) primer with an adaptor sequence covalently bound to its 5'-end. The adaptor sequence contains a *Sfi*I restriction site which is asymmetric to the *Sfi*I restriction site incorporated into the adaptor sequence of the second-strand primer (Figure 2.2). This facilitates site-directed cloning after enzymatic cleavage with only one restriction enzyme. The second *Sfi*I restriction site is linked to the first-strand cDNA fragment when PowerScript™ reverse transcriptase starts to transcribe the second-strand primer (template switching) after it annealed to the short string of deoxycytidine nucleotides by which PowerScript™ extended the 5'-end of the first-strand fragment (terminal deoxynucleotidyl transferase activity)[§].

The first-strand cDNA fragment is thus flanked by an adapter sequence on each end, only by performing a reverse transcription reaction. It can then be used either for second-strand synthesis by primer extension, or complemented and amplified by Long-Distance PCR (LD-PCR) [†]. The amplified cDNA fragments are then cleaved by the *Sfi*I endonuclease, which results in each cDNA fragment having two non-compatible overhanging cDNA ends. These are used to clone the insert into the pDNR-LIB cloning vector in a site-directed orientation.

[§]PowerScript™ reverse transcriptase preferentially extends DNA fragments by adding deoxycytidine nucleotides.

[†]The choice of the procedure depends on the amount of starting material used for reverse transcription. LD-PCR is suggested for cDNA synthesis from a minimum of 25 ng poly(A)$^+$ mRNA, while primer extension can be performed on 0.5 to 2.0 µg poly(A)$^+$ mRNA.

Figure 2.2: Illustration of the two *Sfi*I restriction sites.

The two fortes of SMART™ cDNA library construction technology lie in the following events:

1. PowerScript™ reverse transcriptase extends the first-strand cDNA fragment which it created only after it has reached the 5'-end of the RNA template. Therefore, most truncated cDNA fragments will not have the poly(C)-tail which is necessary for second-strand primer annealing, and consequently the second-strand primer sequence will not become part of these strands becuase PowerScript™ will not be able to switch templates and perform first-strand elongation. Therefore, libraries constructed with SMART™ technology have enhanced representation of full-length cDNA transcripts.

2. *Sfi*I-enzyme is a rare cutting enzyme, which is able to cut through any four base pairs that span the middle of its recognition sequence. Hence, site directed cloning can be accomplished after a singe digestion reaction, using the primer sites that were designed to create incompatible sticky-ends.

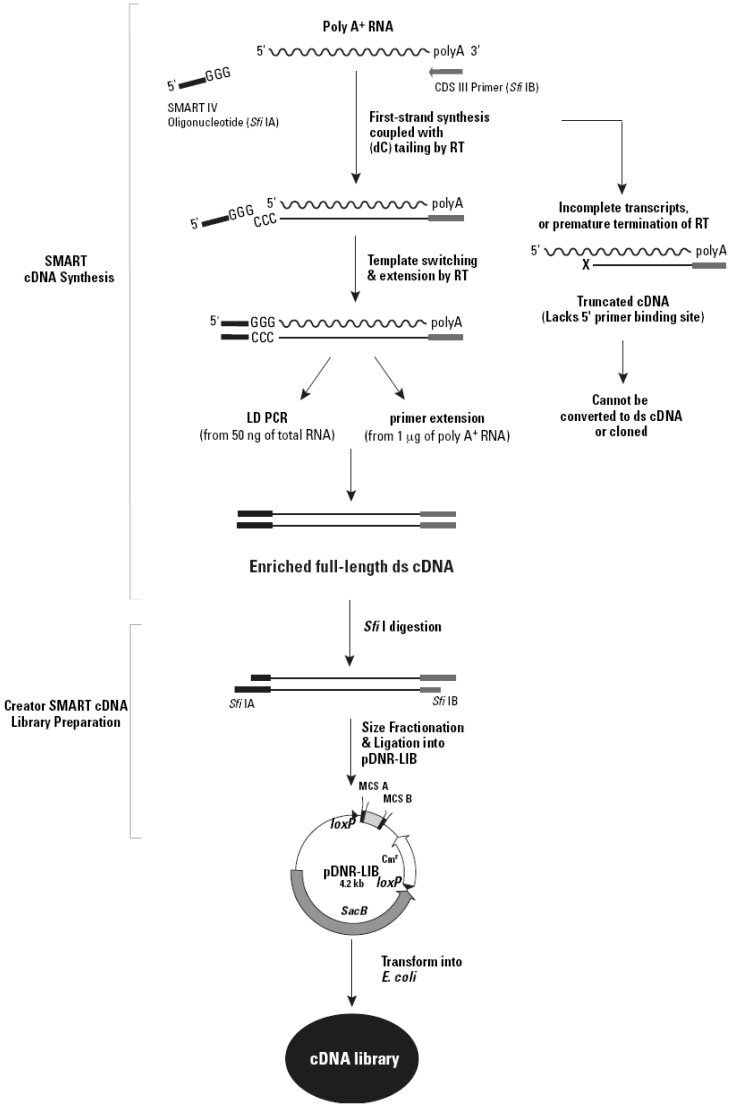


Figure 2.3: Flow chart of the Creator™ SMART™ cDNA Library Construction Kit protocols - graphic obtained from the Creator™ SMART™ cDNA Library Construction Kit User Manual.

Chapter 3

Experimental Procedures

3.1 Introduction

This chapter contains a description of the procedures followed for cDNA library construction, using SMART™ Library Construction technology. Since two approaches with both overlapping and divergent procedures were utilized, an attempt was made to minimize repetition by describing shared procedures simultaneously. Since many of the procedures (mealybug rearing, RNA extraction, vector transformation, library titering, screening for inserts, library amplification and sequencing) were essentially the same for all libraries, this seemed like the most appropriate approach. Differences in library construction were mainly during cDNA synthesis, size fractionation and vector ligation. In short, the Primary Library was constructed essentially according to the instructions provided in the Creator™ SMART™ Library Construction kit (CLONTECH), while the Fractionated Libraries were made with numerous modifications to the instructions provided. Since a detailed description of the SMART™ protocol is freely available (http://www.clontech.com) the construction of the Primary Library is described only succinctly, while that of the Fractionated Libraries is described in detail. General lab techniques that were frequently used are described under a single heading at the end of the chapter.

3.2 Mealybug Rearing

Mealybugs were obtained from the insectarium at ARC Infruitec-Nietvoorbij, Stellenbosch. A small breeding colony was established on butternut pumpkins, *Cucurbita moschata* L., and kept in containers (300 mm × 300 mm × 500 mm) inside a closed incubator. The incubator was set to maintain an optimal breeding temperature of 25°C, at 55% relative humidity with an equalized photoperiod (12 hours light:12 hours darkness).

3.3 RNA Isolation

RNA was extracted with a modified version of the method described by Chirgwin *et al.*[73]. A cesium chloride density gradient was used to separate the RNA from other cellular components and nucleic acids*.

Mealybugs were harvested from butternuts with sterilized forceps. They were collected in 1.5 ml microfuge tubes, flash frozen in liquid nitrogen and ground to a fine powder using a chilled microfuge grinder. Care was taken to keep samples frozen throughout the grinding procedure by continual exposure to liquid nitrogen. The ground material was added to an extraction buffer (5 M guanidium thiocyanate, 30 mM sodium citrate (pH 7.0), 0.5% sarkosyl, 0.7% v/v β-mercapto ethanol) at a w/v ratio of 15 mg/ml, and immediately homogenized using an Ultra-Turrax® T8 S8N-8G disperser (IKA® Labortechnik). The homogenate was split into two 50 ml Oak Ridge tubes (BLD Science) and centrifuged for 40 minutes at 10 000 × g and 4°C (Beckman Allegra™ X-22R centrifuge with F0850 rotor). In order to ascertain the exclusion of all larger particulate matter from the sample, the supernatants were decanted into pre-chilled sterile 50 ml polypropylene tubes through a fine gauze sieve. It was then carefully layered onto cesium chloride cushions (5.7 M in 0.1 M EDTA (pH 8.0, 4 cm depth) in ultracentrifuge tubes. Samples were ultracentrifuged for 19 hours at 27 000 rpm and 20°C (Beckman L-70 Ultracentrifuge with SW28 swing-out rotor). The wax plaque that formed on the surface of each sample was removed with a spatula, and the supernatant was carefully

*Although several protocols for RNA extraction were tested, a detailed description is provided only for the protocol that was used for the actual library construction. For a brief description of methods found to be less suitable, please refer to Table B.1, Appendix B.

drawn off with a pasteur pipet. A small volume of the cesium chloride cushion was left to cover the pellet, in order to avoid any accidental contact with the extraction buffer or the pipet tip. It was expelled with a quick flick of the wrist, and the tubes were kept upside down in order to prevent any remaining liquid to flow onto the purified RNA. The inside wall of each tube was dried with a piece of RNase-free tissue paper while taking great care not to disturb the pellet. The tubes were then turned over again, and the pellets were left to dry for ten minutes before re-suspending them in 100 µl RNase-free water. The dissolved RNA was transferred to RNase-free microfuge tubes (Quality Scientific Plastics). The bottoms of ultracentrifuge tubes were rinsed with an additional 40 µl water and added to the 100 µl samples. RNA quality was assessed (procedures described in the next paragraph) and mRNA was isolated from the complete sample of total RNA (347 ng) according to the instructions in the Oligotex mRNA Spin-Column kit for isolation of poly A^+ mRNA from total RNA (QIAGEN). All RNA samples were stored at -80°C.

Quality Assessment: RNA samples were quantified, and their purity assessed, by means of a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific). The integrity of the total RNA sample was assessed by electrophoresis on a 1.5% w/v non-denaturing agarose gel run in RNase-free 1× TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8.0). RNA was loaded in a 3 volume dilution of NorthernMax® Formaldehyde Load Dye (Ambion), after it had been incubated at 65°C for 15 min, briefly spun down, and placed on ice. All gel casting and running apparates were treated for RNase-contamination by soaking it in a 2% v/v AbSolve™-solution (PerkinElmer) for one hour, and rinsing it with diethyl pyrocarbonate (DEPC)-treated water.

3.4 Preparation and Ligation of Primary Library cDNA

cDNA synthesis was performed by means of the Creator™ SMART™ Library Construction Kit (CLONTECH), essentially according to the manufacturer's instructions. The following is a brief overview of the procedures followed:

Approximately 60 ng of mRNA was used for reverse transcription in a 10 µl reaction volume, and 2 µl thereof was used for cDNA synthesis. cDNA

was amplified for 14 cycles in a GeneAmp 9700 thermal cycler (Perkin-Elmer) that was set to run in 9600-mode[†]. After amplification, the polymerase enzyme was inactivated with proteinase K and the nucleic acid was purified by ethanol precipitation. The resuspended cDNA was digested with *Sfi*I-enzyme to create sticky ends for vector ligation. Short fragments were discarded when the cDNA was size fractionated on a CHROMA SPIN-400 column that had previously been calibrated to a flow rate of one drop every 40 seconds. Sixteen fractions were collected, and 3 µl of each was loaded onto an agarose gel for electrophoresis (see section 3.11, p. 38). The first three fractions producing a detectable signal were pooled, and precipitated with ethanol. The pellet was resuspended in 7 µl deionized water. cDNA was ligated to the provided vector backbone using the suggested volumetric vector to insert ratios (1:0.5, 1:1, and 1:1.5), and incubated at 16°C overnight. Ligation reactions were precipitated for 48 hours with 95% ethanol at -70°C and pellets were resuspended in 5 µl deionized water.

3.5 Preparation and Ligation of Fractionated Library cDNAs

3.5.1 Reverse Transcription

6 µl (550 ng) of mRNA was mixed with 3 µl CDS III/3' PCR Primer (5'-ATTC TAGAGGCCGAGGCGGCCGACATG-d(T)30N–1N–3') and 3 µl SMART IV™ Oligonucleotide (5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGC CGGG-3') and was incubated for 3 minutes at 72°C in a GeneAmp 9700 thermal cycler (Perkin-Elmer). The tube was cooled on ice for 2 minutes before adding a cocktail of 3 µl RiboLock™ Ribonuclease Inhibitor (Fermentas), 6 µl 5× First-Strand Buffer, 6 µl dithiothreitol (20 mM), 3 µl dNTP (10 mM) and 3 µl PowerScript™ Reverse Transcriptase. The content of the tube was mixed by finger tapping, and incubated at 42°C for 3 hours.

[†]The kit provided optimal cycling conditions for the GeneAmp 9600 cycler only. Since the GeneAmp 9700 cycler has a maximum capacity of 50 µl per reaction when run in 9600-mode, the sample was split into two 50 µl reactions after mixing it on a vortex, and both of these were placed in the cycler.

3.5.2 Second-strand Synthesis

10 µl of the reverse transcription reaction was mixed with 0.5 µl 5' PCR Primer (5'-AAGCAGTGGTATCAACGCAGAGT-3'), and was denatured at 94°C for 5 minutes. The reaction was cooled to 60°C, and second-strand reaction mix of the same temperature was added, to a final concentration of 0.5 mM dNTP Mix, 1× *Ex Taq*™ Buffer‡ and 0.1 unit/µl *Ex Taq*™ DNA Polymerase (TaKaRa) in a 30 µl reaction volume. The sample was incubated for three cycles of 1 min/60°C and 15 min/68 °C in a GeneAmp 9700 thermal cycler (Perkin-Elmer) running in 9600 mode.

3.5.3 Size Fractionation

The second-strand reaction was size fractionated on an agarose gel containing no ethidium bromide. A 5 µl aliquot was loaded in the first lane, and served as an indicator lane for the size distribution of the cDNA. The GeneRuler™ 1 kb DNA Ladder (Fermentas) was loaded adjacent to the indicator lane, and the remaining cDNA was loaded one well separated. After electrophoresis, the gel was cut along the length of the separating well, and the piece containing the marker and indicator lane was stained for 10 minutes in 1× TAE buffer containing 1 µg/ml ethidium bromide. The two gel pieces were then placed side-by-side on a UV-transilluminator with the indicator lane and ladder exposed to the light source, and the unstained piece shielded. Four size fractions with ranges 0.25 kb - 0.5 kb (Fraction 1), 0.5 kb - 1 kb (Fraction 2), 1 kb - 2 kb (Fraction 3) and $\geqslant$2 kb (Fraction 4) were excised from the unstained piece, using the transilluminated indicator and marker lanes as guidelines. Each fraction was excised with a new scalpel blade, and the cDNA was recovered by means of the Zymoclean Gel DNA Recovery Kit™ (Zymo Research), following the manufacturer's instructions.

3.5.4 cDNA Amplification

The number of cycles needed for the amplification of each fraction was ascertained by preparing three 10 µl trial reactions for each, and subjecting them

‡Composition proprietory

to 14, 16 and 18 cycles of amplification respectively. The reaction mix in each tube consisted of 1 µl gel-purified cDNA template, 0.5 mM dNTP Mix (TaKaRa), 0.5 mM 5' PCR Primer (SMART™ Library Construction Kit), 0.5 mM CDS III/3' PCR Primer (SMART™ Library Construction Kit), 0.1 unit Ex Taq™ DNA Polymerase (TaKaRa) and 1× *Ex Taq*™ Buffer. Samples were placed in a GeneAmp 9700 thermal cycler (Perkin-Elmer) running the following program in 9600 mode: 5 min/95°C and 14, 16, or 18 repeats of 10 sec/95°C, 10 min/68°C. 5 µl of each reaction was then loaded onto an agarose gel for electrophoresis.

The actual cDNA amplification reactions were prepared in exactly the same way as the trial reactions, with the only difference being a 10-fold increase in volume of each reagent. The resultant 100 µl reactions were also divided into two 50 µl reactions each, in order to suit the specifications of the GeneAmp 9700 thermal cycler (Perkin-Elmer) when run in 9600 mode. A negative control reaction was also prepared, with 10 µl ddH_2O substituting the cDNA template in the reaction. Each fraction was amplified for 17 cycles, and 5 µl of each reaction was electrophoresed on an agarose gel to verify the success of amplification.

3.5.5 cDNA Purification

Amplified cDNA fractions were purified with the MSB® Spin PCRapace kit (Invitek), following the manufacturer's instructions. Yield was assessed on the NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific).

3.5.6 Enzymatic Digestion

*Sfi*I-digestion reactions were prepared with nucleic acid concentrations approximating the specifications of the manufacturer. cDNA content and reaction volumes for the four cDNA fractions were as follows:

Fraction 1: 14 µl (1.715 µg), 50 µl reaction volume
Fraction 2: 17.5 µl (4.023 µg), 100 µl reaction volume
Fraction 3: 16 µl (3.106 µg), 100 µl reaction volume
Fraction 4: 15 µl (1.347 µg), 50 µl reaction volume

All digestions were performed with 1 µl *Sfi*I enzyme (Fermentas), and had a final buffer concentration of 1× Buffer G (10 mM Tris-HCl (pH 7.5 at 37°C), 10 mM $MgCl_2$, 50 mM NaCl, 0.1 mg/ml BSA)(Fermentas). Reactions were incubated in a heated-lid thermocycler at 50°C. After 24 hours the tubes were taken out, briefly spun down, and another 1 µl of *Sfi*I enzyme was added to each. Reactions were mixed by finger tapping and placed back in the thermocycler for an additional 24 hour incubation period. Digestions were purified with the MSB Spin PCRapace kit (Invitek), following the manufacturer's instructions.

3.5.7 Vector Preparation

An additional stock of pDNR LIB vector was prepared as follows:

(i) Enzymatic digestion: A clone with a insert of approximately 2 kb in length was isolated from the Primary Library, and an overnight culture was prepared from it. The plasmid was purified using the GeneJET™ Plasmid Miniprep Kit (Fermentas), and the yield was quantified on a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific). Four 50 µl *Sfi*I-enzyme digestion reactions were prepared, each one containing 6 µg plasmid DNA, 10 units *Sfi*I-enzyme (Fermentas), and 1× Buffer G (10 mM Tris-HCl (pH 7.5 at 37°C), 10 mM $MgCl_2$, 50 mM NaCl and 0.1 mg/ml BSA) (Fermentas). The reactions were incubated at 50°C for 72 hours, with 10 units of fresh *Sfi*I-enzyme being added twice at 24 hour intervals, and an inactivation step of 20 minutes at 80°C preceding the second addition.

(ii) Purification: The vector backbone and insert were separated by agarose gel-electrophoresis. The gel was run for 60 minutes in freshly prepared 1× TAE buffer. A scalpel blade was used to excise the backbone from the gel, using the position of the xylene cyanol electrophoresis front as an indication of its position, as UV-irradiation can cause damage to DNA. The gels were stained by ethidium bromide nonetheless, as correct excision had to be ascertained afterwards by irradiating the remaining gel pieces with UV-light. The insert was also excised, and both fragments were purified by means of the Wizard® SV Gel and PCR Clean-Up

System (Promega). Yield was quantified on a NanoDrop® ND-1000 spectrophotometer (Thermo Fisher Scientific).

(iii) Dephosphorylation: The vector was dephosphorylated with shrimp alkaline phosphatase (SAP) (Fermentas), using the suggested ratio of enzyme to pmol ends (calculations provided in Appendix D, p. 69). A 50 µl reaction was prepared with 1.9 µg purified pDNR-LIB and 0.2 units of SAP in 1× Reaction Buffer (0.1 M Tris-HCl (pH 7.5 at 37°C), 0.1 M $MgCl_2$ and 0.1 mg/ml BSA) (Fermentas). The mixture was incubated at 37°C for 15 minutes, whereafter the enzyme was inactivated at 65°C for 15 minutes. The SAP was removed, and the vector concentrated by purification through a Zymoclean Gel DNA Recovery Kit™ (Zymo Research).

(iv) Trial Ligations: Ligation reactions were prepared both for the SAP-treated- and the untreated vector. They were prepared in 10 µl volumes, and consisted of 190 ng vector (5 µl of the SAP-treated-, and 1 µl of the untreated vector respectively), 1.25 units of T4 DNA Ligase (Fermentas), 1× Ligation Buffer (40 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT and 0.5 mM ATP (pH 7.8 at 25°C)) and 5% w/v polyethylene glycol (PEG) 4000. The effectiveness of the ligation reaction itself was monitored by a control reaction consisting of 5 µl GeneRuler™ 1 kb DNA Ladder (Fermentas) and the same concentrations of T4 DNA Ligase, Ligation Buffer and PEG 4000 as was used for the other ligation reactions. Samples were incubated overnight in a pre-cooled thermocycler at 16°C. Effectiveness of ligation was assessed by gel-electrophoresis of the control reaction.

(v) Trial Transformations: 5 µl of ligation reaction was added to 100 µl chemically competent *E. coli* DH5α cells (preparation described in Appendix C, p. 66), and was mixed by gentle finger tapping. The reactions were incubated on ice for 15 minutes, heat shocked for 45 seconds at 42°C, and placed on ice for another 1 minute. 900 µl LB broth was added to each tube, and they were incubated for 60 minutes at 37°C and 150 rpm. The reactions were spun down at 8 000 × g for 3 minutes, and the cells were resuspended in 100 µl LB broth. The complete samples were then plated onto pre-warmed YT-plates with chloramphenicol selection (see section 3.11, p. 38). Colonies were counted after 16 hours incubation at

37°C[§].

3.5.8 Ligations

Ligation reactions were set up for each of the four cDNA size fractions, using an insert to vector ratio of 3 to 1 in a 10 µl reaction volume. Reactions were prepared with 1 µl SAP-treated pDNR-LIB vector and an appropriate volume of *Sfi*I-digested insert (an example calculation and the actual insert volumes that were used are provided in Appendix D, p. 70). Reaction conditions were the same as for the trial ligations (see 3.5.7, section *(iv)*), only with the appropriate volume of insert included in the ligation mix. A control reaction was also included, and assessed by gel-electrophoresis.

3.6 Transformations

Vector constructs of both the Primary and Fractionated Libraries were transformed into *E. cloni*® Electrocompetent Cells (Lucigen® Corporation). For each transformation, 1 µl of purified ligation reaction was used along with 25 µl competent cells. Electroporations were performed in pre-chilled 1.0 mm gap cuvettes (Bio-Rad), using a Bio-Rad Gene Pulser set to 200 Ω, 25 µF, and 1500 V. The transformed cells were resuspended in 970 µl Recovery Medium immediately after electroporation, and incubated at 37°C and 150 rpm for one hour. 5 µl of each reaction was mixed with 45 µl recovery medium and plated onto a 2× YT agar plate with chloramphenicol selection (30 µg/ml). The inoculum was allowed to soak into the agar for 10 min before plates were turned over and incubated at 37°C for 16 hours.

[§]NOTE: After the vector had been tested and found suitable for library construction, another set of ligations were prepared to ascertain the optimal ratio of vector to insert for each of the cDNA fractions created previously. These procedures and results are not described in detail. Briefly four *Sfi*I-digested fragments of 0.35 kb, 0.75 kb, 1.5 kb and 2 kb were ligated to 1µl purified vector at molar ratios of 1:1, 1:3 and 3:1 (insert:vector) respectively. Insert volumes were calculated using the same formula as described in Appendix D, p. 70 (calculations and results not given), and ligations were prepared with the same reaction conditions as described for other trial ligations and transformed like the rest of the ligations.

3.7 Library Titering

Titering of both the Primary and Fractionated libraries were performed according to the instructions provided in the SMART manual. Serial dilutions were made from all transformation reactions in order to determine their individual titers. Dilutions were made in the following way: A 5 µl aliquot was taken from the transformation reaction and thoroughly mixed with 995 µl sterile LB broth, using a vortex. Three aliquots (1 µl, 5 µl and 50 µl) were then plated onto YT agar plates with chloramphenicol selection (30 µg/ml)[¶]. A second dilution was made by mixing 5 µl of the first dilution with 995 µl sterile LB broth. Three aliquots (10 µl, 50 µl and 100 µl) of this dilution were then plated onto YT agar plates with chloramphenicol selection (30 µg/ml). After allowing the inoculum to soak into the agar for 10 minutes, the plates were turned over and incubated at 37°C overnight. Colonies were counted by hand, and the titer for each transformation reaction was calculated (for an example of such a calculation, please refer to Appendix D, p. 71).

3.8 Screening for Inserts

The percentage of recombinant clones in each library was determined by means of the polymerase chain reaction (PCR). Isolated colonies were picked from the titer plates and inoculated into 25 µl aliquots of PCR cocktail (see next paragraphs). The reactions were placed in a GeneAmp 9700 thermal cycler (Perkin-Elmer) and amplified with the following program: 1 cycle of 30 sec/94°C; 20 cycles of 30 sec/94°C and 2 min/68°C; and 1 cycle of 8 min/68°C. 5 µl of each sample was loaded into a gel for electrophoresis.

Primary Library PCR cocktail: 1× Advantage™ 2 PCR buffer, 0.2 mM dNTPs, 0.4 µM M13 forward primer (5'-GTAAAACGACGGCCAGT-3'), 0.4 µM M13 reverse primer (5'-AACAGCTATGACCATG-3') and 0.5 µl of 50× BD Advantage™ 2 Polymerase Mix[‖].

[¶]Whenever less than 50 µl of a dilution was plated out, the volume was adjusted to 50 µl by adding an appropriate amount of sterile LB medium to the sample.

[‖]50× BD Advantage™ 2 Polymerase Mix is a cocktail of BD TITANIUM *Taq* DNA Polymerase, BD TaqStart Antibody, and an unknown proofreading polymerase. The exact quantities of each of these enzymes are not provided.

Fractionated Libraries PCR cocktail: 1× KAPATaq High Yield Buffer (75 mM Tris-HCl (pH 8.8), 0.01% (v/v) Tween 20, 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$), 0.2 mM dNTPs, 0.4 µM M13 forward and -reverse primer each (Integrated DNA Technologies), 1× cresol loading dye and 0.05 U/µl KAPATaq DNA Polymerase (Kapa Biosystems).

3.9 Library Amplification

Amplification of the Primary and Fractionated Libraries were done similarly, with the exception that their individual titers dictated the use of different plating volumes (example calculation provided in Appendix D, p. 72):

Primary Library:	6 µl/plate
Fractionated Library 1:	100 µl/plate
Fractionated Library 2:	100 µl/plate
Fractionated Library 3:	20 µl/plate
Fractionated Library 4:	20 µl/plate

Libraries were amplified on 145 mm diameter 2× YT agar plates with chloramphenicol selection (30 µg/ml). A suitable aliquot was mixed with 150 µl neat LB medium and spread onto each plate. After allowing the inoculum to soak into the agar, plates were turned over and incubated at 37°C overnight. The amplified library was then harvested by adding 5 ml LB medium to each plate and scraping the colonies from the surface with a sterile glas plate. The resuspended cell mixture was poured into sterile 50 ml polypropylene tubes containing 12.5 ml sterilized glycerol each and homogenized on a vortex. Four 1 ml aliquots were taken from the 50 ml tubes and placed in 1.5 ml microfuge tubes for convenience of later platings. All tubes were placed in a -70°C freezer for long term storage**.

**NOTE: Since the Primary Library was created from three 1 ml transformation reactions, it consisted of too large a volume to completely amplify. After plating out approximately 1 ml (165 plates) thereof, 0.5 ml sterilized glycerol was added to the remaining 2 ml, and it was mixed thoroughly on a vortex. The tube was stored along with the amplified fractions at -70°C

3.10 Insert Sequencing & Data Analysis

3.10.1 Sequencing

Isolated colonies were picked from the titration plates and screened for the presence of multiple inserts by PCR. Overnight cultures were prepared from the selected clones, and their plasmids were isolated with the GeneJET™ Plasmid Miniprep Kit (Fermentas), according to the manufacturer's instructions.

Sequencing was performed on a ABI 3130*xl* Genetic Analyzer (Applied Biosystems) at the Core DNA Sequencing Facility, Stellenbosch University. Single reads were prepared from the 5'-end only, using the M13 forward primer.

3.10.2 Sequence Editing

Raw trace files were edited by means of the BioEdit sequence alignment editor (Version 7.0.9, Tom Hall, www.mbio.ncsu.edu/BioEdit/bioedit.html) and the Sequence Scanner Software v1.0 package (www.appliedbiosys tems.com). The 5'-ends of trace files were vector trimmed right after the *SfiI*-recognition sequence (5'-GGCCATTACGGCC-3'), while their 3'-ends were trimmed as soon as a poly-A motive was found. Sequences that did not contain a poly-A motive were trimmed at the first ambiguous peak that had a low confidence value (<20) assigned by the Sequence Scanner Software v1.0 package (www.appliedbiosys tems.com). Whenever the Sequence Scanner displayed high confidence values at very poor peak resolutions, the sequence was trimmed at the first instance where base calling was uncertain, using personal discretion.

3.10.3 Blast Analysis

The edited sequences were concatenated into five fasta files, each representative of a cDNA library. Redundancy within each library was assessed by multiple sequence alignment, making use of Clustal W (http://www.ebi.ac.uk/ Tools/-clustalw/)[74].

The files were submitted for blast analysis via the interface of the National Center for Biotechnology Information (NCBI, http://blast.ncbi.nlm.nih.gov) using three different search algorithms: tblastx (against the RefSeq-

and non-redundant nucleotide database), blastx (against the RefSeq-, and non-redundant protein database), and megablast (against all GenBank non-mouse and non-human EST entries). Gene-specific html files were extracted from the automatically generated directories and saved locally.

Hits with an expectancy value of 10^{-5} or smaller were parsed from the html files using a script written in Python (version 2.4, Python Software Foundation, http://www.python.org/psf/ - program code supplied in Appendix F, Listing F.1). The top scoring hit of insect origin was extracted from each file, and its unique identifier was used to retrieve a description of its molecular- and biological functions from Entrez Gene (http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene) or Flybase (http://flybase.org/reports/FBgn0019789.html)[†].

3.10.4 Identification of Potential Target Sequences

Potential targets for RNAi-mediated suppression of *P. ficus* genes were identified by:

1. Matching the blast results to a set of genes that had already been shown to be essential for insect viability or development[43] (Python code supplied in Appendix F, Listing F.2).
2. Manually selecting potential target genes from the refined blast results, while taking into consideration only:
   - Hits with an expectancy value of $\leq 10^{-30}$[‡].
   - Hits for which an explicit functional description was available.
   - Genes that were deemed essential for life or normal development.

The nucleotide sequences of the identified gene targets were re-submit-ted for blast analysis, using the megablast- and discontiguous megablast alignment algorithms. Potential targets with high sequence similarity to plant orthologs were discarded.

[†]Matches of insect origin were given priority over better scoring hits of other origin in order to ensure that the functional annotation of the encoded protein would be appropriate for insect metabolism specifically. Whenever no insect match was available, the best scoring alignment itself was used.

[‡]A high expectancy cutt-off value was used in order to discard with falsely identified genes. Only RefSeq blastx and tblastx results used.

3.10.5 Sequence Submission

All non-mitochondrial and non-ribosomal EST sequences were submitted to GenBank and the database of Expressed Sequence Tags (dbEST) according to the data format and procedure specified for streamlined submission (Python code supplied in Appendix F, Listing F.3).

3.11 Frequently used Techniques

3.11.1 Agarose Gel Electrophoresis

Unless stated otherwise, nucleic acids were separated on 1.1% w/v agarose D1 LE (Hispanagar) gels that were run for 45 minutes at 4 V/cm in a 1× TAE (40 mM Tris-acetate, 1 mM EDTA pH 8.0) buffer. Gels contained 0.5 µg/ml ethidium bromide for visualization with ultra-violet (UV) transilllumination and image capture by GeneSnap image acquisition software (Syngene). Except for colony PCRs, all samples were loaded with a 6× Loading Dye Solution (10 mM Tris-HCl (pH 7.6), 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol and 60 mM EDTA)(Fermentas), and the GeneRuler™ 1 kb DNA Ladder (Fermentas) was used as a size marker. Since colony PCRs were prepared with a loading dye (1 M sucrose, 0.02% cresol) as part of the master mix, they were loaded onto gels directly.

3.11.2 Culture Media

YT medium was prepared with 5 g Yeast Extract, 8 g Tryptone, 5 g NaCl and 15 g Agar bacteriological (all products from MERCK), brought to a total volume of 1 ℓ with distilled water (dH_2O). It was autoclaved and allowed to cool to approximately 55°C before adding 1ml of a 30 mg/ml chloramphenicol stock solution to it. Plates were cast in a laminar flow cabinet, and stored in a dark container at 4°C for a maximum period of two weeks. Before use, plates were pre-warmed for two hours at 37°C.

3.11.3 Overnight Cultures

Well isolated colonies were picked with sterilized toothpicks and inoculated into 5 ml autoclaved aliquots of Luria-Bertani broth (MERCK) containing no selection. They were incubated at 37°C, while shaking at 225 rpm overnight.

Chapter 4

Results and Discussion

In this chapter the results obtained for RNA isolation, the various steps of cDNA library construction, and the preliminary screening of libraries is presented and discussed. With the exception of the BLAST results (comparatively discussed in Section 4.4), all results are discussed in chronology of their acquisition in the laboratory. Sequences that had been identified as potential targets for RNAi-mediated *P. ficus* pest control are discussed in Section 4.5.

4.1 RNA Isolation from Mealybugs

Finding a suitable protocol for RNA extraction was a high priority of this project, as RNA integrity is of paramount importance for the construction of full-length cDNA libraries.

A unique problem was posed by the high amount of wax (dorsal covering of mealybug spp.) that inevitably accompanied mealybug RNA extractions. Although the wax is probably inert *per se*, it interfered with phase separation in organic extraction protocols and complicated sampling of the aqueous phase by clogging pipet tips. The wax also clogged the membranes of RNA purification columns, thereby obstructing RNA binding and the flow of homogenized samples. Consequently all organic- and column-based extraction protocols were rejected (tested methods listed in Table B.1, Appendix B along with comments on their additional shortcomings).

The method of Chirgwin *et al.*[73] provided solutions for all of the difficulties encountered with other extraction protocols. Large quantities of high quality RNA could be extracted, and samples did not need to be treated with DNase before downstream implementation[†].

An unusual fluorescence profile was observed when the RNA samples were electrophoresed on a non-denaturing agarose gel. The distinct three-banded profile (Figure 4.1) was in stark contrast to the conventional two-banded profile which is characteristic of most mammalian total RNA preparations. The profile also did not conform to the conventional 2 to 1 signal ratio of the 28S and 18S rRNA bands[‡]. Nonetheless, distinct rRNA banding was observed, and the faint fluorescence signal that could be discerned above, between and below the three rRNA bands, was an indication of excellent RNA integrity[10,77].

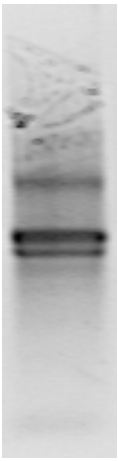


Figure 4.1: Electrophoresis profile of a mealybug total RNA sample. (100ng aliquot loaded in NorthernMax® Formaldehyde Load Dye (Ambion) and electrophoresed on a 1.5% w/v non-denaturing agarose gel run in RNase-free 1× TAE at 4V/cm for 45 min).

[†]See Chapter 2, Section 2.5.3, p. 16

[‡]The unusual fluorescence profile of the mealybug RNA samples could be explained by the presence of a hidden-break in the 28S rRNA transcript - as is characteristic for many insect species[75,76]. If this was the case, the lowest band in the profile would represent a fragment of the 28S rRNA transcript (approximately 1.8 kb), while the middle band would contain the 18S rRNA transcript and the remaining fragment of the 28S rRNA transcript. Co-migration of the latter fragments would explain the high fluorescence signal obtained from the middle band, while the weak signal obtained from the top-most band would represent a diminished population of unfragmented 28S rRNA transcripts migrating at their usual position.

The yield and purity of all total RNA preparations was within the acceptable range (Table 4.1)[§]. Although no explanation could be found for the elevated ratios of the purified mRNA samples, a survey of the literature indicated that this is a common phenomenon, and that the samples could be regarded as sufficiently purified for cDNA library construction[10].

Table 4.1: Concentrations, purity parameters and quantities of total RNA and mRNA used for Primary- and Fractionated Library construction.

Sample	**Concentration**	$\mathbf{A}_{260/280}$	$\mathbf{A}_{260/230}$	Quantity used
total RNA_{prim}	518.27 ng/µl	1.99	2.31	0.26 µg
$mRNA_{prim}$	20 ng/µl	2.40	3.73	60 ng
total RNA_{frac}	3091.7 ng/µl	2.11	2.20	1.6 µg
$mRNA_{frac}$	92 ng/µl	2.25	2.24	550 ng

4.2 Construction of the Primary Library

Construction of the Primary Library was accomplished without difficulties. The only alteration to the Creator™ SMART™ Library Construction Kit protocol was in the number of cycles used for cDNA amplification. These had to be optimized, as the proposed number of cycle repeats resulted in a cDNA smear that continued up into the high molecular weight region of the gel. Since this is usually an indication of overcycling[78], the number of cycles was lowered until good results were obtained (Figure 4.2).

The range of the cDNA smear (∼0.25 kb to ∼4 kb) correlated well with the size ranges reported for other insect cDNA libraries (0.18 - 3.3 kb for *Solenopsis invicta*[79] and 0.113 - 1.87 kb for *Glossina moritans*[80]). Three distinct bands (electrophoresing at ∼0.7 kb, 1 kb and 1.5 kb) represented highly expressed sequences in the transcriptome of the mealybug. Since the results indicated that reverse transcription and cDNA amplification had been successful, the amplified product was digested with *Sfi*I-enzyme and subjected to size-fractionation.

[§]Parameters discussed in Chapter 2, Section 2.5.5, p.18.

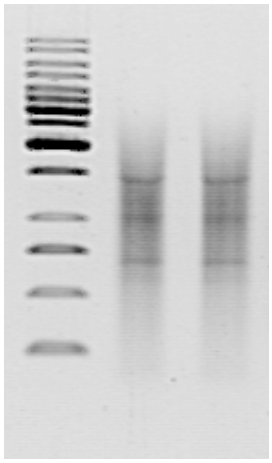


Figure 4.2: Amplified cDNA profile after fourteen PCR cycles (5µl aliquots electrophoresed on a 1.1% agarose gel - specifics of the DNA size marker provided in Appendix E, p. 73).

The distribution of cDNA fragments in the different size-fractions was visualized by agarose gel electrophoresis. A fluorescence signal could be detected from fraction 7 onwards, with the most intense signals observed in fractions 9 and 10 (Figure 4.3). It was not possible to determine the size ranges of the cDNA fragments in the different fractions, as the cDNA was too dilute to detect by the time the size marker had adequately separated. As a conservative measure to exclude smaller fragments, only the first three fractions (7, 8 and 9) were pooled for use in the ensuing ligation reactions.

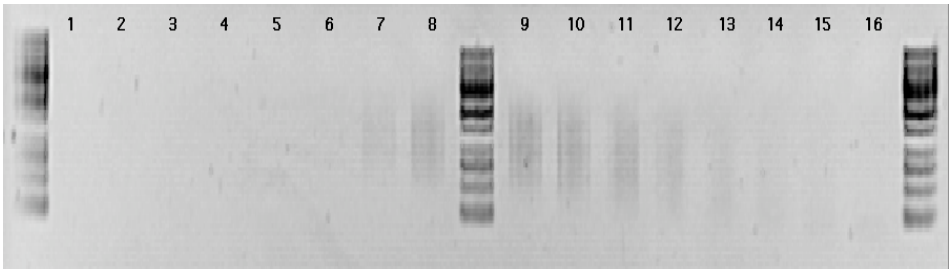


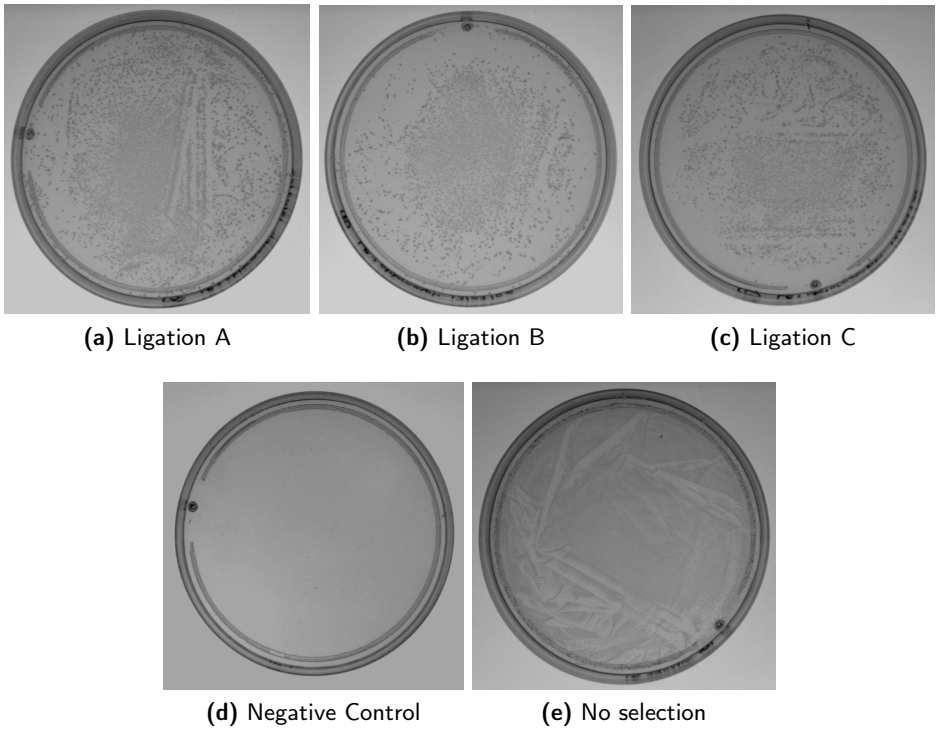
Figure 4.3: Visualization of the cDNA content of each of the size-fractions obtained from the CHROMA SPIN-400 column (3 µl of each fraction loaded onto a 1.1% agarose gel and electrophoresed for 20 min. in 1× TAE).

Ligation reaction conditions were monitored by a control reaction containing GeneRuler™ 1 kb DNA Ladder (Fermentas). The dense smear above the normal electrophoresis range of the ladder represented the ligated ladder fragments, and was an indication of very good ligation conditions (Figure 4.4).

Figure 4.4: Ligation control reaction. Left lane: Unligated ladder. Right lane: The same ladder after overnight incubation at 16°C with 0.25 µl T4 DNA Ligase and 1× Ligation Buffer.

All transformation reactions were successful, with nearly confluent plates obtained from all but the negative control reaction (Figure 4.5). Lack of growth on the negative control plate indicated that there had been no contamination of the ligation reaction (with insert), or the transformation reaction (with uncut vector). It also suggested that the antibiotic selection was effective, and that growth on the other plates was a result of antibiotic resistance acquired through plasmid transformation. This was substantiated by a control plate which contained no chloramphenicol, and onto which an aliquot of the negative control reaction was plated. Confluent growth on this plate demonstrated the efficiency of antibiotic selection on all other plates.

Similar titer values were obtained for all of the transformation reactions (results not shown), and hence the three transformation reactions were pooled. The pooled library had an estimated titer of 3.57×10^6 cfu/ml (calculations provided in Appendix D, section D.4). This compared well with the titers of other cDNA libraries constructed from insects[80,81].



(a) Ligation A **(b)** Ligation B **(c)** Ligation C

(d) Negative Control **(e)** No selection

Figure 4.5: *E. coli* growth on 30 µg/ml chloramphenicol selective medium after platings with various electro-transformed cell suspensions. Plates (a), (b) and (c) represent ligation reactions with volumetric vector:insert ratios of 0.5:1, 1:1 and 1.5:1. Controls were provided by platings of the negative control reaction onto selective (d), and non-selective medium (e).

4.2.1 Evaluating the Size-distribution of Clone Inserts

The size-distribution of the library was estimated by PCR amplification of the inserts of 134 randomly picked colonies. Clones that produced multiple products (identified by multiple banding in the electrophoresis profile), were discarded from the survey, as these results were ambiguous. Nearly all of the remaining amplification products migrated in the 2 kb to 0.5 kb range, with most of them migrating below the front of the 1 kb ladder rung (Figure 4.6). After subtracting the nearly 270 bp of co-amplified vector backbone from the length of each insert, 56 of the screened inserts were estimated to be shorter than 0.5 kb, while 27 were 1 kb or longer. The largest insert was estimated to be 2.25 kb (clone I1, Figure 4.6).

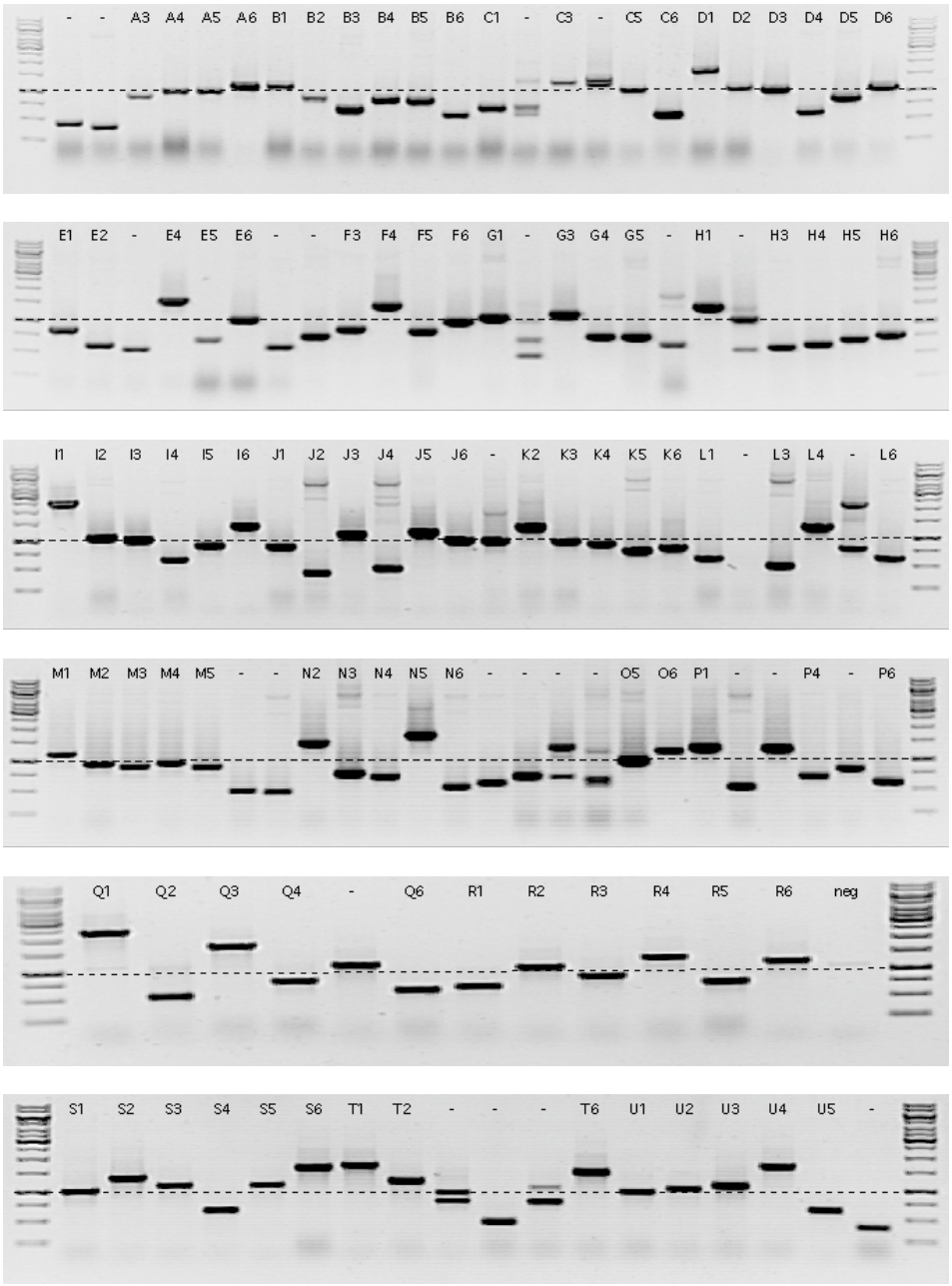


Figure 4.6: Colony PCRs of randomly picked colonies from the titration plates of the Primary Library. The electrophoresis front of the 1 kb rung of the GeneRuler™ 1 kb DNA Ladder (Fermentas) is indicated by the dotted line. Only the inserts of clones corresponding to the numbered lanes were sequenced.

Although the insert size range was still comparable to the size ranges reported for other insect libraries[79,80], the high prevalence of smaller inserts was disconcerning. The average insert length also did not compare well with the 2 kb average reported for a *D. melanogaster* cDNA library[82].

4.2.2 Troubleshooting

Since most smaller cDNA fragments were supposed to be discarded during the size fractionation step, the resolution of the CHROMA SPIN-400 column was investigated. An aliquot of the GeneRuler™ 1 kb DNA Ladder (Fermentas) was loaded onto the column and the profile of its size fractionation was inspected. Interestingly, all detectable fractions but the first contained virtually all ladder fragment sizes (Figure 4.7). Two further repeats of the procedure confirmed the first result. Since the ladder spanned fragments of 0.25 - 10 kb, the column was deemed unsuitable for high resolution fractionation of the particular size-distribution.

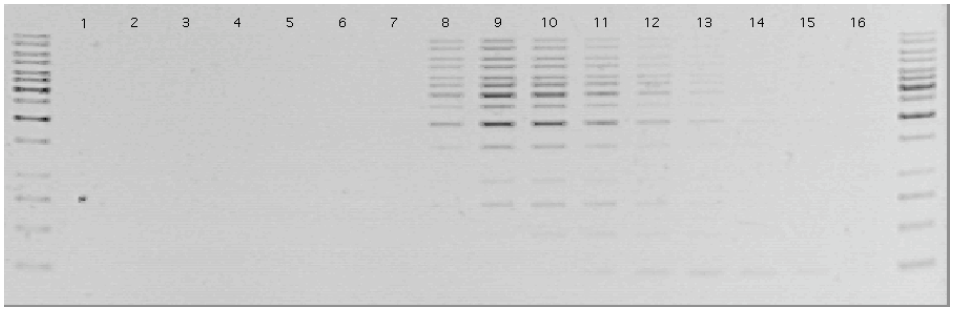


Figure 4.7: Electrophoresis profile of the size-fractions obtained from a 20µl aliquot of GeneRuler™ 1kb DNA Ladder, after fractionation through a CHROMA SPIN-400 column (5 µl of each fraction was loaded onto a 1.1% agarose gel and electrophoresed for 45 min. in 1× TAE.).

The results suggested that an equally poor resolution had been obtained during the fractionation of Primary Library cDNA, and provided a plausible explanation for the high abundance of short inserts observed.

4.2.3 Formulating a New Methodology

In order to obtain longer clone inserts, the Creator™ SMART™ Library Construction protocol had to be adapted. Modifications were introduced according to a method proposed by Wellenreuther *et al.*[83]:

1. Instead of using the CHROMA SPIN-400 column, size-fractionation was performed on an agarose gel by electrophoresis. A set of different size-fractions was excised from the gel and purified.
2. Size-fractionation was no longer performed on the amplified cDNA, but rather on the second-strand product. The various size-fractions were then amplified individually.
3. Instead of pooling the amplified products, ligation reactions were performed for each of the size-fractions individually.
4. Transformation reactions were likewise performed individually.

Motivation: The above procedure addressed the various stages of the original protocol at which a bias towards the incorporation of smaller cDNA fragments was promoted. Research has shown that shorter PCR templates have a competitive advantage over longer ones if they are amplified in a conjugated reaction[84]. Moreover, shorter nucleic acids are more readily cloned into vectors than longer ones[85], and larger vectors transform less efficiently than smaller ones[86]. By keeping the different size-fractions separate during these procedures, the competitive advantage that smaller fragments have for incorporation into a library can be attenuated.

4.3 Construction of the Fractionated Libraries

Besides the main alterations to the protocol, the pDNR-LIB vector backbone had to be recovered from a clone of the Primary Library.

4.3.1 Recovering the pDNR-LIB Vector Backbone

Preparing the vector backbone was more problematic than originally anticipated. A special problem was posed by the *Sfi*I restriction enzyme by which the insert had to be excised from the Primary Library plasmid.

*Sfi*I differs from most endonucleases in that it is a tetrameric enzyme which has to engage with two binding sites before cleaving them in a single concerted process[87,88]. The efficiency of the enzyme is a factor not only of its chances to encounter two of its binding sites at the same time, but also of the frequency at which the second binding site is blocked as a result of its occupation by another enzyme. The lower the concentration of available binding sites, the higher the incidence at which secondary binding sites will be blocked.

As a result, complete digestion of plasmids by *Sfi*I digestion is a very difficult goal to attain. Optimization of the reaction conditions resulted in an extensive incubation-&-recovery protocol, the result of which was a vector backbone with a calculated transformation background of 0.5% (Figure 4.8). This was regarded as an acceptable purity, and the vector was used in the subsequent library construction protocol.

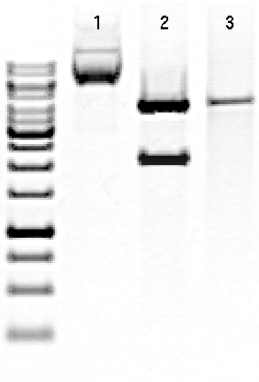
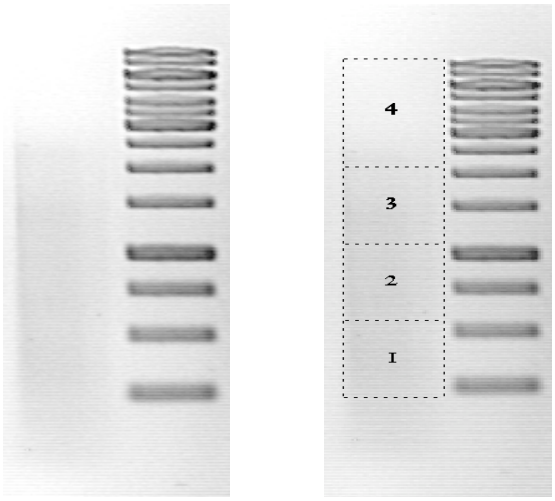


Figure 4.8: Some of the most important stages of vector purification: The purified plasmid (1), the *Sfi*I-digested miniprep (2), and the backbone after gel-purification and SAP-treatment (3).

4.3.2 Preparing the cDNA Size-fractions

The RNA used for construction of the Fractionated Libraries was of a higher concentration than that which was used for Primary Library construction (Table 4.1, p. 42). The fluorescence profile of the second-strand product had a similar distribution to the amplified cDNA of the Primary Library, but the cDNA was too dilute to discern the three distinct bands that were observed before (Figure 4.9). Four size fragments (indicated in Figure 4.9(b)) were excised from the gel and purified.



(a) Second-strand reaction.

(b) Incisions for size-fractionation.

Figure 4.9: Size-fractionation of second-strand cDNA by gel-electrophoresis: Four size-fractions, ranging from 0.25 kb - 0.5 kb (fraction 1), 0.5 kb - 1 kb (fraction 2), 1 kb - 2 kb (fraction 3) and $\geqslant$ 2 kb (fraction 4), were obtained after excision and purification of the gel slices indicated by the dotted lines in (b).

Small aliquots of the purified cDNA fractions were used in a set of trial amplification reactions by which the optimal number of cycles needed for fraction amplification was determined. A comparative gel indicated that 18 cycle repeats would suffice for amplification of fractions 1 to 3, and that a few more cycles could be used for amplification of fraction 4 (Figure 4.10). The distinct stepwise fashion by which the size-distribution of the amplified cDNA fractions increased was a clear indication that the size-fractionation procedure

had been effective. Also, the three bands that were observed in the Primary Library cDNA were now detectable in the profiles of the first, second and third fractions.

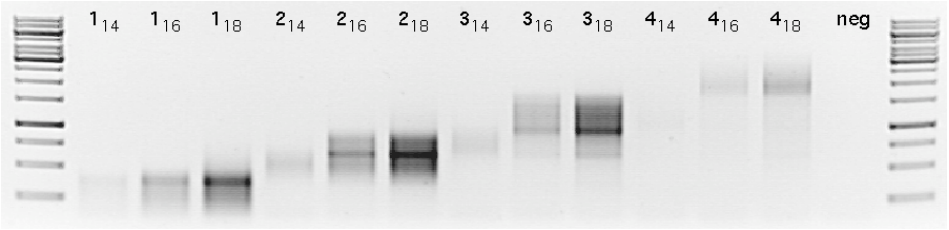


Figure 4.10: Results of the trial amplifications of the four cDNA size-fractions (each lane is labeled by its fraction number, and is subscripted with the number of cycle repeats used for amplification - i.e. lane "2_{18}" represents the cDNA of fraction 2 after 18 PCR cycle repeats).

As a precaution against the introduction of too much template or enzyme due to pipetting errors, the upscaled reactions were amplified for 17 cycles only. Electrophoresis of the amplified products indicated that 17 cycle repeats had indeed been sufficient for the amplification of all four cDNA fractions. Interestingly, the fluorescence signal of the fourth fraction was now much stronger than what it had been in the optimization reactions. This could be attributed, as was explained before, to an increased template concentration in the upscaled amplification reaction.

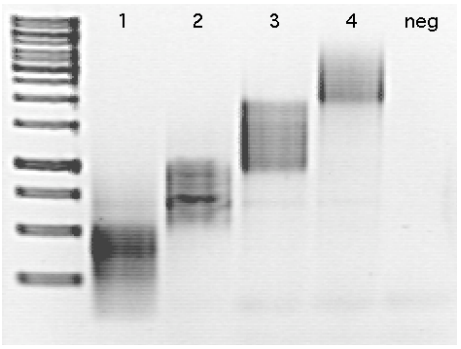


Figure 4.11: Fractionated Library cDNA after 17 rounds of amplification. GeneRuler™ 1 kb DNA Ladder followed by Fraction 1 (lane 1), Fraction 2 (lane 2), Fraction 3 (lane 3), Fraction (lane 4) and the negative control (lane 5).

cDNA recovery after PCR-cleanup was high. Unfortunately, almost half of the material was lost during subsequent *Sfi*I-digestion and purification (Table 4.2). The loss could be ascribed - to some extend - to the desired loss of short terminal adapter fragments, but was most likely also due to the loss of some of the desired inserts. Nonetheless, enough material had been purified to proceed with cDNA cloning.

Table 4.2: Recovery of the purified size-fractions before and after *Sfi*I digestion.

Sample	**Recovery before *Sfi*I digestion**	**Recovery after *Sfi*I digestion**
Fraction 1	122.5 ng/µl	64.91 ng/µl
Fraction 2	229.9 ng/µl	141.42 ng/µl
Fraction 3	194.1 ng/µl	75.75 ng/µl
Fraction 4	89.8 ng/µl	54.17 ng/µl

4.3.3 Cloning the Four Size-fractions

The four cDNA fractions were ligated to the purified pDNR-LIB vector backbone, and the electrophoresis profile of the control reaction was the same as for the Primary Library (Figure 4.4, p. 44)[†]. All cDNA fractions transformed successfully and produced confluent plates (results not shown). Five colonies were detected on the negative control plate, but these were identified as artifacts of the earlier vector purification procedure.

The calculated titers of Fractionated Libraries 3 and 4 were similar to the titer of the Primary Library, but Fractionated Libraries 1 and 2 had titers that were one order of magnitude lower (Table 4.3). The difference in titer values could have been caused by a number of factors, including: Differences in cDNA recovery after purification of the ligation reactions; concatenation of shorter

[†]A number of optimization experiments were performed to determine the optimal insert to vector ratio needed for cDNA cloning. To spare the limited amount of fractionated cDNA, these reactions were performed on cDNA fragments isolated from the Primary Library, which were representative of the mean insert sizes of the four cDNA fractions. Best results were obtained from an insert to vector ratio of 3 to 1 for all clones.

inserts in the ligation reactions; and/or differences in the electroporation conditions due to unequivocal success of ligation purification. Nonetheless, the results were satisfactory and the size ranges of library inserts were investigated.

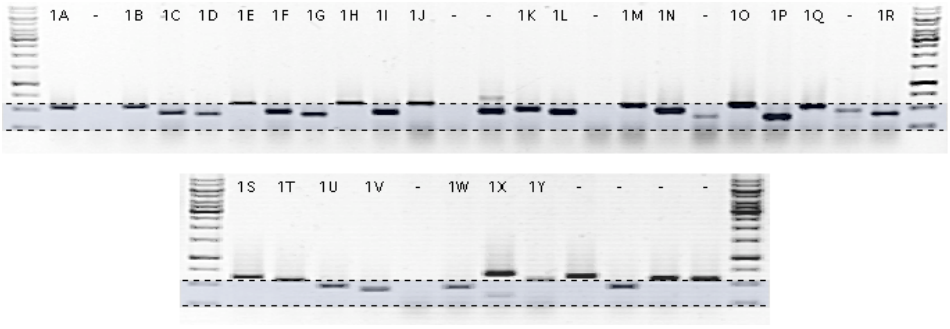
Table 4.3: Calculated titers of the four Fractionated Libraries.

Library	**Titer (*cfu*/ml)**
Fractionated Library 1	2.20×10^5
Fractionated Library 2	2.40×10^5
Fractionated Library 3	1.25×10^6
Fractionated Library 4	1.19×10^6

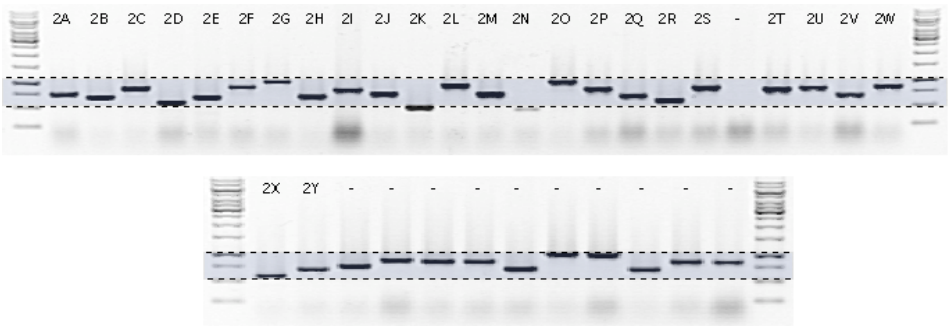
4.3.4 Evaluating the Size-distribution of Clone Inserts

The size-distributions of the inserts of the first three libraries correlated well with the positions of the gel slices from which their cDNA templates were obtained (Figure 4.12). The inserts of the fourth library, however, were uniformly distributed over the range ∼0.25 kb - 4 kb. This was most likely the result of a small quantity of shorter fragments that had been purified along with the larger fragments, and had competed with them for successful amplification, cloning and transformation. Nonetheless, the library had an encouraging amount of large inserts and contained the longest insert of all clones screened (∼4 kb - clone 4D, Figure 4.13).

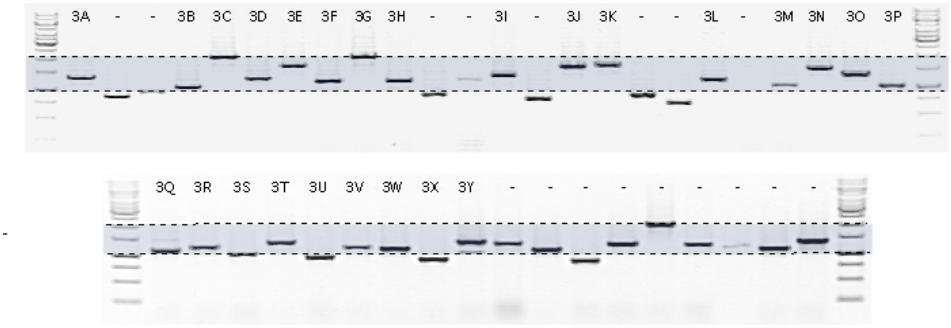
Average insert sizes of the four Fractionated Libraries were estimated at ∼0.25 kb, 0.5 kb, 1 kb and 2 kb respectively. Fractionated Libraries 3 and 4 were thus on par with other insect cDNA libraries with regards to their average insert sizes[79,80,82].



(a) Amplification products of Fractionated Library 1



(b) Amplification products of Fractionated Library 2



(c) Amplification products of Fractionated Library 3

Figure 4.12: Amplification products of randomly selected clones of the first (a), second (b) and third (c) Fractionated Libraries (highlighted sections represent the position of the gel slice from which the second-strand cDNA was recovered. Numbered lanes indicate the clones of which the inserts were sequenced.).

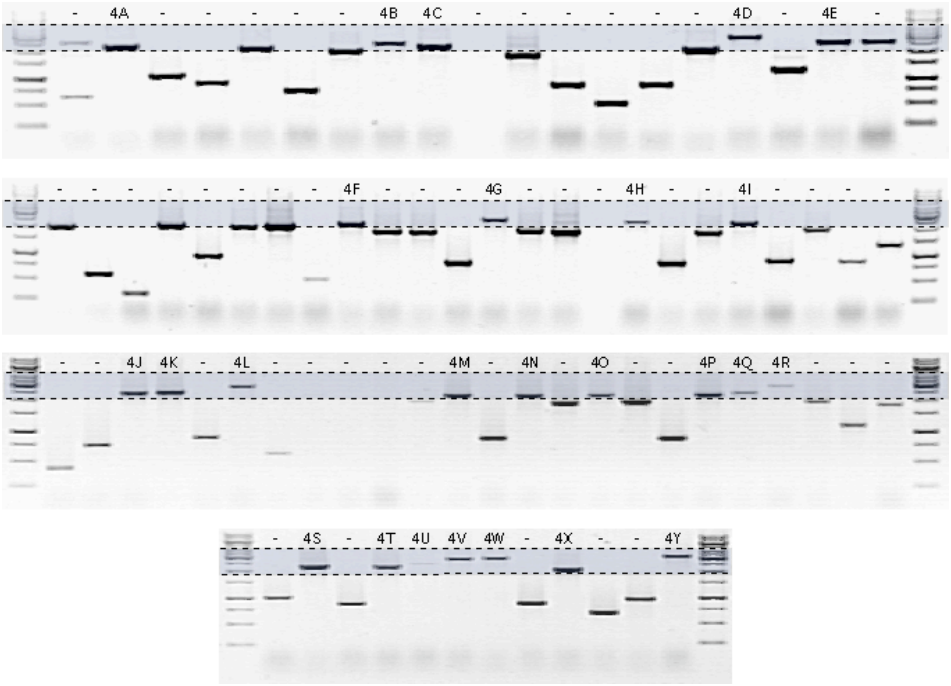


Figure 4.13: Amplification products of randomly selected clones of the fourth Fractionated Library (highlighted sections represent the position of the gel slice from which the second-strand cDNA was recovered. Numbered lanes indicate the clones of which the inserts were sequenced.).

4.4 Comparative Analysis of the Sequencing- and Blast Results of the Primary- and Fractionated Libraries

A pilot investigation was launched to provide a birds eye-view of the content of the five cDNA libraries. The investigation involved a hundred clones selected from the Primary Library, and twenty-five clones selected from each of the Fractionated Libraries[‡].

High quality sequence data could be obtained for 95 of the clones selected from the Primary Library, and 82 of the clones selected from the Fractionated Libraries. Sequence redundancy was highest for Fractionated Libraries 2 and

[‡]Although a larger-scale sequencing project would certainly have provided a more accurate view of the character of the individual libraries, the costs of single-clone sequencing was prohibitively expensive and the laboratory was not yet geared for a more streamlined sequencing procedure.

3, while no redundant clones were detected in the sampling of Fractionated Library 4. Analysis of the blast results indicated that Fractionated Library 4 returned the largest quantity of RefSeq hits when expressed as a percentage of the number of sequences submitted (Figure 4.14). The second most informative library was Fractionated Library 3, and the least so was the Fractionated Library 2.

It was interesting to see that approximately 82% the sequence matches returned from the non-redundant database were from organisms of the class Insecta. Moreover, two of the eleven available *P. ficus* sequences were matched in this query (Table G.1, Appendix G), and many of the remainder were from a close relative of *P. ficus*, the pink hibiscus mealybug (*Maconellicoccus hirsutus*). This was also true for the EST results, ∼50% of which were obtained from *M. hirsutus* (Table G.2, Appendix G). In contrast, most RefSeq results were from *Drosophila melanogaster* (Table G.3, Appendix G).

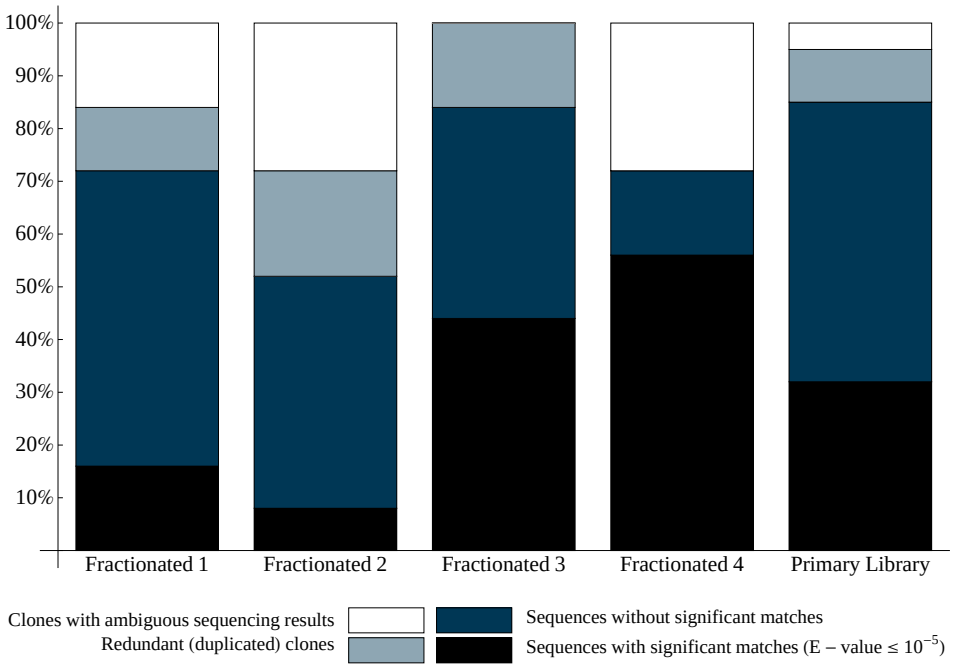


Figure 4.14: Quantitative overview of sequencing- and blast results obtained for the Primary- and Fractionated Libraries. Results were normalized over the number of clones analyzed from each library (i.e. 25 clones from each of the Fractionated Libraries, and 100 clones from the Primary Library).

4.5 Identification of Potential Targets for RNAi-mediated *P.ficus* Control

For RNAi-mediated pest control to be effective, the target sequence should be: *(i)* Essential for the pest's development or vital functions, and *(ii)* dissimilar to any of the genes transcribed in the organism hosting the silencing-construct. Moreover, the strategy can only be implemented if the silencing mechanism is harmless to all non-target organisms consuming the agricultural product. By this reasoning, two potential target genes were identified from the clones analyzed in the study (Table 4.4).

Table 4.4: Clones containing potential target sequences for an RNAi-mediated approach to *P. ficus* pest control.

Clone ID	**Gene Description**	**Confidence of Protein Prediction**	**Nucleotide Sequence Matches with:** *Non-arthropod Seq.*
3S	Lethal(2)essential for life	9e-27	-
H2	Chickadee	3e-53	yes

The small heat shock protein encoded by the insert of clone 3S, and the profilin encoding transcript of H2 were both essential for insect homeostasis[†], and - with one exception[§] - all available nucleotide sequences aligning with these inserts were of arthropod origin.

Another transcript worth mentioning is the actin encoding transcript of clone 4S. This transcript has previously been identified as a potential target for RNAi-mediated pest control by Baum *et al.*[43]. In their study, expression of the gene was suppressed in the western corn rootworm (WCR) by means of an orally administered silencing construct. Significant stunting of larval development and high mortality was observed, but high concentrations of dsRNAs had to be used ($\geq$520 ng/cm^3 feeding medium). Since such high concentrations of dsRNA are unlikely to be expressible in grapevine, effective control of *P.*

[†]Descriptions of the molecular- and biological functions of the identified transcripts are provided in Table G.3, Appendix G.

[§]*Picea sitchesis* entry (gb|EF083566.1|), spanning 55% of the *P. ficus* chickadee sequence with 68% sequence identity, including 3 gaps.

ficus through suppression of this gene transcript does not seem very feasible. Another criticism is that the target sequence has significant similarity to the nucleotide sequence of an actin transcript of human origin (dbj|AK313294.1|). Since humans are consumers of grapevine products, grapevine transformation with a silencing construct directed against this gene is unlikely to be accepted.

Chapter 5

General Discussion

The purpose of this study was to create a resource of *P. ficus* coding sequences from which a suitable target for RNAi-mediated pest control could be obtained. To this end, the construction of whole organism cDNA library was proposed.

In accordance with the specific tasks formulated at the start of this study, an RNA isolation protocol suitable for use on mealybugs was identified and optimized, and altogether five cDNA libraries were created, titered, screened, and amplified. The study demonstrated the importance of using a protocol designed to circumvent the bias towards incorporation of shorter transcripts in cDNA libraries. It also contributed some 162 *P. ficus* transcripts to the knowledge database of the scientific community, 63 of which were unrelated to any of the sequences previously available. Although a larger-scale sequencing project would certainly have provided a more accurate view of the character of the individual libraries, the laboratory was not geared for such a project by the time the libraries had been constructed. Nonetheless, the pilot investigation increased the hitherto available amount of data on *P. ficus* coding sequences to a large extent.

During the library screening procedure, two potential targets for an RNAi-mediated approach to *P. ficus* pest control were identified. With one exception, the sequences seemed to be unique to arthropods. This was a very encouraging result, as it suggested that *P. ficus* expression of the transcripts could be suppressed via a silencing construct hosted in grapevine, without the dsRNA posing any risks to either grapevine- or human metabolism.

A great deal of research needs to be done to confirm the suitability of these transcripts for grapevine transformation and - most importantly - the efficiency by which they are able to constrain *P. ficus* proliferation. Amongst others, the inserts of the identified clones should be characterized to identify the areas useful for their suppression in *P. ficus*, and a set of trial experiments should be performed where dsRNA is fed to *P. ficus* instars and the effects are monitored for an appropriate period of time. It would also be important to determine whether the transcript sequences are complementary to any of the transcripts commonly expressed in grapevine. The combined results of these experiments would determine whether it would be sensible to proceed with the construction of a silencing construct suitable for grapevine transformation.

Should the proposed target sequences be found unsuitable for RNAi-mediated *P. ficus* pest control, the cDNA libraries created in this study could be instrumental to the discovery of more suitable target sequences. Potential targets could be obtained by means of directed screening techniques such as array PCR or complementary hybridization or, alternatively, identified by means of a streamlined procession of clone selection and miniprep sequencing. The latter procedure holds the potential for identification of novel transcripts suitable for insect pest control, but is heuristic and would require a great deal of bioinformatic exercises to process the large amount of data thus generated. The more focussed procedure would be to directionally screen the libraries for the presence of transcripts that have formerly been identified as potential targets for insect pest control[†].

To date, only two studies have reported successful introduction of RNAi-mediated resistance to agricultural pests. The target transcripts in these cases were the coding region of a putative V-ATPase A of the western corn rootworm[43], and a region of the P450 monooxygenase gene of the cotton bollworm[42]. Identification of the orthologs of these sequences from the *P. ficus* cDNA libraries might provide a good point from which to proceed.

Fundamental research on plant-pathogen and plant-insect interactions is also a rich source of information from which potential target genes can be

[†]Although the initial objective of the altered cDNA library construction protocol was simply to obtain a better representation of larger cDNA inserts, it has come to mind that the four Fractionated Libraries lend themselves to size-directed cDNA library screening. This could be useful, for instance, if a library has to be screened for a transcript of which the approximate length is known.

identified. In the course of their evolution, plants have developed an array of strategies to outwit their many pathogenic symbionts: Insect attack induces the production of toxins, digestibility reducers (proteinase inhibitors), insect repellents and volatile organic compounds that lure predators and parasitoids[89,90]. In reaction, herbivorous insects have developed an assortment of counterstrategies: Various species are able to neutralize plant toxins through the use of cytochrome P450 monooxydases and glutathione *S*-transferases[91,90]. Suppression of plant responses to wounding has also been demonstrated for the corn earworm, which secretes salivary glucose oxidase while feeding[92]. Gene expression profiles of plants and their antagonists could provide valuable insight into the dynamics of their interactions, and potentially elucidate suitable targets for gene suppression.

An elegant alternative would be to identify target transcripts through investigation of the dynamics of the three-way plant::virus::vector interaction. With *P. ficus* being a phloem feeding insect, grapevine responses to mealybug infestation are likely to be closely related to the plant's responses to phloem virus infections[93,94]. To investigate grapevine responses to viral infection, an expression profile of GLRaV-3 infected *Vitis vinifera* red wine cultivars was recently established[95]. Moreover, a circulatory mechanism of viral transmission was recently proposed for *Planococcus citri*[96] - the closest known relative to *P. ficus*. The presence of GLRaV-3 particles in the salivary glands of *P. citri* suggests that transportation of the virus occurs through the gut epithelium into the haemolymph, and again over the basal lamina and plasmalemma of these structures. Transportation of viral particles in the haemolymph of hemipteran insects is most often possible only because of the protection provided by chaperonin GroEL homologues[97,98]. Although viral protection by GroEL homologs in *P.citri* haemolymph has been questioned, the presence of a similar protective compound is very likely. Molecular interactions between *P.ficus* and its associated grapevine viruses are therefore most likely. The outcome of a three-way interactive investagation might elucidate gene targets suitable not only for biological control of the pest, but also its capacity to transmit grapevine viral diseases.

Chapter 6

Conclusion

In this study, altogether five cDNA libraries were created from the transcribed sequences of the vine mealybug, *Planococcus ficus* (Signoret). Instrumental to their construction was the identification of an RNA isolation protocol that was suitable for use on samples containing large amounts of wax. The cDNA libraries differed with regards to the prevalence and average size distributions of cDNA fragments they contained. Although they were not exhaustively analyzed, a pilot investigation of 200 clones indicated two potential targets for an RNAi-mediated approach to *P. ficus* pest control. The investigation also proved that the libraries have a good representation of insect-related coding sequences and that, as such, they represented a reliable resource of coding sequences from which new genes could be discovered and characterized. Therefore, should the initial target sequences provide unsuitable for implementation, the libraries could be instrumental to the identification of more suitable target sequences in the future.

Appendix A

Precautions when working with RNA

A.1 Creating an RNase-free working environment

Most laboratory suppliers can provide certified RNase-free laboratory equipment and reagents. These are a welcome and trustworthy alternative to in-house cleansing procedures. To prevent re-contamination of these materials while handling them, gloves should be worn and frequently changed. Reusable laboratory equipment should be treated prior to use. Glassware can be treated with a 0.1% solution of DEPC or 0.1 N sodium hydroxide (NaOH), or could be baked at 250°C overnight. Plastic equipment and surface areas should be treated with a solution of 0.5 N sodium hydroxide (NaOH) and 0.5% SDS, or a 3% solution of hydrogen peroxide (H_2O_2). Alternatively commercial solutions such as AbSolve™ and RNase AWAY® can also be used. Buffers and solutions should be made from reagents reserved for RNA work. They could be treated with DEPC or, if they contain DEPC-inactivating compounds such as Tris, filter sterilized.

Table A.1: RNase-inactivating compounds suitable for use in molecular biology.

Chaotropic agents such as urea, guanidine hydrochloride and guanidinium thiocyanate denature proteins by disrupting their secondary and tertiary structures. Urea achieves this through the disruption of hydrogen bonds and solvation of the hydrophobic core[99,100], while guanidine destabilizes the electrostatic interactions needed to stabilize the protein in its functional conformation[101].
Anionic surfactants like sodium dodecyl sulphate (SDS)[a] can induce compaction of proteins at low concentrations, but are paradoxically also very potent denaturants of proteins at intermediate concentrations. This is due to the dual hydrophobic and electrostatic interactions of the molecule with the various protein moieties[100]. Another frequently used ionic detergent is *N*-Laurylsarcosine[b]. In contrast to SDS, sarkosyl is less prone to precipitation and can be used along with chaotropic agents[10].
Diethylpyrocarbonate(DEPC) inactivates enzymes permanently by modifying the histidine residues at their active sites[102]. It is not suitable for use in Tris-containing buffers, as Tris incapacitates DEPC by binding to it. DEPC must be inactivated before using treated solutions, as it is carcinogenic and can chemically alter RNA molecules and render them unsuitable for downstream applications.
Reducing agents such as dithiothreitol (DTT), β-mercaptoethanol (β-ME)and Tris(carboxyethyl) phosphine (TCEP) denature proteins by disruption of the disulphide bonds needed to maintain their secondary and tertiary structures[103,104,105].
RNasin inacitvates RNases by noncovalent competitive binding to the enzyme. It is effective against RNase A, B and C, but not against RNase T_1, S1 nuclease or RNase from *Aspergillis*, but newer modified versions have broader effective ranges. RNasin is most useful in protocols where protein denaturation is not desirable[106]. In contrast to vanadyl ribonucleoside complexes (VRCs), it does not interfere with downstream applications such as reverse transcription[107]. It is, in fact, frequently implemented as a precaution against RNA degradation during reverse transcription reactions[10].
Organic solvents such as phenol and chloroform both denature and precipitate proteins. They also facilitate RNA isolation by separating extractions into an aqueous (nucleic acid containing) and organic (other cellular components) phase. Phenol can be used as a standalone, but is not a very stable compound[10]. Most frequently it used in combination with chloroform and isoamyl alcohol in a 25:24:1 ratio. Chloroform aids in phase separation and stabilizes phenol, while isoamyl alcohol reduces foaming of proteins[10].
Proteinase K is a proteolytic enzyme with broad-range stability both for temperature (25°C - 65°C) and acidity (pH 4.0 - 12.5)[108]. It is usually removed from reactions by phenol:chloroform extraction after incubation at optimal conditions.

[a] Also known as lauryl sulfate

[b] Also known as sarkosyl

Appendix B

RNA Extraction Protocols less suitable for use on Mealybugs

Table B.1: RNA extraction protocols found to be less suitable for use on mealybugs

Buffer Type	Buffer Content	Remarks
GES-buffer[a]	2 M guanidine thiocyanate, 10 mM EDTA, 0.24% sarkosyl, 10 mM citric acid, 90 mM acetate, 60 mM β-mercaptoethanol	Low yield; high DNA content
TENS-buffer[b]	0.1 M Tris-HCl (pH 7.5), 1 mM EDTA, 0.1 M sodium chloride, 1% SDS, 20 mM β-mercaptoethanol	Low yield; poor integrity; DNA co-purified
$NaClO_{4\text{-buffer}}$[c]	10 M $NaClO_4$, 1 M Tris-HCl (pH 8.3), 20% SDS, 8.5% w/v PVPP, 2% w/v PEG 6000, 1% v/v β-mercaptoethanol	Moderate yield; DNA co-purified
RNeasy Plant mini kit[d]	Proprietary - contains guanidine thiocyanate	Column membranes block up with wax; high background when samples are run on gel

[a] Chomczynski, P. and Sacchi, N. (1987)
[b] Bugos *et al.* (1995)
[c] Davies, C. and Robinson, S. P. (1996)
[d] QIAGEN

Appendix C

Protocol for preparation of Chemically Competent Cells

An isolated colony of *E. coli* DH5α was inoculated into 5 ml LB medium and placed in a rotary incubator at 37°C, 225 rpm. After 16 hours the complete culture was inoculated into 500 ml LB medium (2 ℓ Erlenmeyer flask), and placed back in the rotary incubator. A 1 ml aliquot was taken after 2 hours incubation and every 30 min thereafter until an optical density of 0.4 - 0.6 was obtained at 600 nm wavelength (OD_{600}: 0.4 - 0.6). As soon as the desired optical density was obtained, the culture was spun down for 10 min at 10 000 rpm, 4°C. The supernatant was discarded and the cells were gently resuspended in 100 ml ice-cold 100 mM $MgCl_2$. After 25 min incubation on ice the cells were spun down for another 10 min at 4 000 rpm, 4°C. Pellets were gently resuspended in 10 ml ice-cold resuspension buffer (100 mM $CaCl_2$, 15% glycerol), and 100 µl aliquots were pipetted into pre-chilled 1 ml eppendorf tubes. The tubes were flash frozen in liquid nitrogen and the aliquots were stored at -80°C.

Protocol adapted from Hanahan[109].

Appendix D

Calculations

D.1 Primer addition to second-strand reaction

The concentration of the 5' PCR primer in the second-strand reaction needed to be corrected, as it was diluted after addition of the first-strand reaction to the second-strand reaction mix. The volume of 5' PCR Primer needed to supplement the SMART IV™ Oligonucleotide already in the reaction was calculated as follows:

$$\begin{aligned}\text{Primer concentration in RT reaction} &= \frac{\text{primer volume} \cdot \text{primer concentration}}{\text{reaction volume}} \\ &= \frac{3\mu l \cdot 10\mu M}{30\mu l} \\ &= 1\mu M\end{aligned}$$

Calculating the initial 5' primer concentration in the second-strand reaction:

$$\begin{aligned}&= \frac{\text{volume RT added} \cdot \text{primer concentration in RT}}{\text{reaction volume}} \\ &= \frac{10\mu l \cdot 1\mu M}{30\mu l} \\ &= 0.33\mu M\end{aligned}$$

Calculating the volume of 5' primer needed to bring concentration to 0.5 µM:

$$\begin{aligned}
\text{5' primer}_{\textit{moles available}} &= \text{primer conc}_{\textit{2ndStrand reaction}} \cdot \text{volume}_{\textit{2ndStrand reaction}} \\
&= 0.33\mu M \cdot 0.00003l \\
&= 1 \times 10^{-5}\mu mol
\end{aligned}$$

$$\begin{aligned}
\text{5' primer}_{\textit{moles wanted}} &= \text{primer conc}_{\textit{wanted}} \cdot \text{volume}_{\textit{2ndStrand reaction}} \\
&= 0.5\mu M \cdot 0.00003l \\
&= 1.5 \times 10^{-5}\mu mol
\end{aligned}$$

$$\begin{aligned}
\text{5' primer}_{\textit{moles needed}} &= \text{5' primer}_{\textit{moles wanted}} - \text{5' primer}_{\textit{moles available}} \\
&= 1.5 \times 10^{-5}\mu mol - 1 \times 10^{-5}\mu mol \\
&= 0.5 \times 10^{-5}\mu mol
\end{aligned}$$

$$\begin{aligned}
\text{5' primer}_{\textit{volume needed}} &= \frac{\text{5' primer}_{\textit{moles needed}}}{\text{5' primer}_{\textit{stock concentration}}} \\
&= \frac{0.5 \times 10^{-5}\mu mol}{10\mu mol/l} \\
&= 0.5 \times 10^{-6}l \\
&= 0.5\mu l
\end{aligned}$$

D.2 pDNR-LIB: Calculation of picomole ends

According to the specifications for the SAP-enzyme (Fermentas), a minimum of 0.05 units of enzyme should be used to dephosphorylate one picomole of 5'-protruding termini, while a minimum of 0.1 units should be used per picomole of 5'-recessed termini. Since *Sfi*I-digestion of pDNR-LIB creates both 5'-protruding and 5'-recessed termini, the larger amount of enzyme (0.1 units per picomole DNA termini) was used in the dephosphorylation reaction. Calculations were done taking onto consideration that there are two sticky ends per vector fragment, and that the enzyme is provided in a stock concentration of 1 unit/µl.

$$\begin{aligned}\text{Amount of DNA in 10 µl of the purified vector} &= 10\mu l \cdot 190ng/\mu l \\ &= 1900ng \\ &= 1.9{\times}10^6 pg\end{aligned}$$

$$\begin{aligned}\text{Quantity picomole ends for a fragment of 3 943 bp} &= 2 \text{ ends} \cdot \frac{pg \text{ dsDNA}}{\text{MW in } Da} \\ &= 2 \text{ ends} \cdot \frac{1.9{\times}10^6 pg}{3943bp \cdot 660Da} \\ &= 1.46 \text{ pmol ends}\end{aligned}$$

$$\begin{aligned}\text{Amount of SAP needed for dephosphorylation} &= \frac{1.46 \text{ pmol ends} \cdot 0.1\text{unit/pmol end}}{1\text{unit}/\mu l} \\ &= 0.146\mu l \text{ SAP} \\ &\approx 0.2\mu l \text{ SAP}\end{aligned}$$

D.3 Ligations: Volume of insert needed to establish the desired vector:insert ratio

All ligation reactions were prepared with 1 µl purified vector, and calculations were done with its length and concentration rounded to 4 kb and 190 ng/µl respectively. All other parameters used for calculations are presented in the table below:

Insert	**Length (kb)**[a]	**Concentration (ng/µl)**
Trial insert	2	210
Fraction 1	0.35	64.91
Fraction 2	0.75	141.42
Fraction 3	1.50	75.75
Fraction 4	3.00	54.17

[a]Estimated average fragment lengths were used for calculations of the fragmented libraries

The following formula was used:

$$\text{Insert}_{volume} = \text{ratio} \cdot \frac{\text{Vector}_{concentration} \cdot \text{Insert}_{length} \cdot \text{Vector}_{volume}}{\text{Insert}_{concentration} \cdot \text{Vector}_{length}}$$

Example calculation:

Volume of trial insert needed to establish a insert:vector ratio of 3:1

$$= \frac{3}{1} \cdot \frac{190ng/\mu l \cdot 1\mu l \cdot 2kb}{210ng/\mu l \cdot 4kb}$$
$$= 1.4\mu l$$

Calculated volumes (with actual volumes used given in brackets):

Fraction 1: 0.77 µl (1µl)
Fraction 2: 0.76 µl (1µl)
Fraction 3: 2.82 µl (3µl)
Fraction 4: 7.89 µl (6.5µl)

D.4 Library titering

Only plates on which the colonies were well separated were used to calculate library titers, and an average of three plates were used to make these calculations from. As an example, the calculation of the mean titer of the Primary Library is given:

$$\text{Titer} = \frac{cfu}{\text{volume plated out}} \cdot \text{dilution factor}^{\text{-1}}$$

$$\begin{aligned}
\text{Titer 1 (plating of 1}^{\text{st}}\text{ dilution)} &= \frac{14cfu}{1\mu l} \cdot 0.005^{\text{-1}} \\
&= 2.8{\times}10^{3} cfu/\mu l \\
&= 2.8{\times}10^{6} cfu/ml \\
\text{Titer 2 (plating of 1}^{\text{st}}\text{ dilution)} &= \frac{67cfu}{5\mu l} \cdot 0.005^{\text{-1}} \\
&= 2.68{\times}10^{3} cfu/\mu l \\
&= 2.68{\times}10^{6} cfu/ml \\
\text{Titer 3 (plating of 2}^{\text{nd}}\text{ dilution)} &= \frac{1cfu}{10\mu l} \cdot 0.000025^{\text{-1}} \\
&= 4.0{\times}10^{3} cfu/\mu l \\
&= 4.0{\times}10^{6} cfu/ml \\
\text{Titer 4 (plating of 2}^{\text{nd}}\text{ dilution)} &= \frac{6cfu}{50\mu l} \cdot 0.000025^{\text{-1}} \\
&= 4.8{\times}10^{3} cfu/\mu l \\
&= 4.8{\times}10^{6} cfu/ml
\end{aligned}$$

$$\begin{aligned}
\text{Mean titer} &= \frac{Titer1 + Titer2 + Titer3 + Titer4}{4} \\
&= \frac{2.8{\times}10^{6} cfu/ml + 2.68{\times}10^{6} cfu/ml + 4.0{\times}10^{6} cfu/ml + 4.8{\times}10^{6} cfu/ml}{4} \\
&= 3.57{\times}10^{6} cfu/ml
\end{aligned}$$

D.5 Library amplification: Volume of transformation reaction needed

The volume of transformation reaction needed to amplify a library was calculated in accordance with its mean titer, and with the objective of obtaining 20 000 colony forming units (cfu) per plate:

$$\text{Volume} = \frac{\text{cfu wanted per plate}}{\text{titer in cfu}/ml}$$

and the following titer values:

Primary Library:	3.57×10^6 cfu$/ml$
Fractionated Library 1:	2.2×10^5 cfu$/ml$
Fractionated Library 2:	2.4×10^5 cfu$/ml$
Fractionated Library 3:	1.25×10^6 cfu$/ml$
Fractionated Library 4:	1.19×10^6 cfu$/ml$

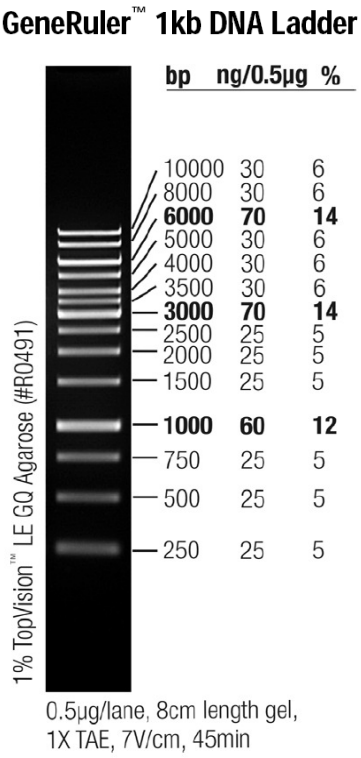
Example calculation:

Volume of transformation mixture needed to amplify the Primary Library:

$$\begin{aligned}\text{Volume} &= \frac{20000\text{cfu}}{3.57\times10^6\text{cfu}/ml} \\ &= \frac{2\text{cfu}}{357\text{cfu}/ml} \\ &= 0.0056ml \\ &= 5.6\mu l\end{aligned}$$

Appendix E

GeneRuler™ 1kb DNA Ladder



Appendix F

Python Scripts

Listing F.1: Python script written to extract tblastx hits from a locally stored html file using command-line interface.

```
import sys

#function to obtain the hitname, e-value and score for a
    querty sequence for whih a match(hit) was found
def getvalues(hit):
    global evalues,scores,names
    names.append(hit[hit.find(">")+1:hit.find("</a")])
    scores.append(hit[hit.rfind("</a>")-3:hit.rfind("</a
        >")].strip())
    evalues.append(hit[-13:-5].strip())


#function to find out which query sequences had matches
    in the database...
def gethits(filename):
    global Hits, NoHits, counthits, countnohits,
        description
    Hits = []
    NoHits = []
    description = []
    counthits = 0
```

```
    countnohits = 0

    list=open(filename,"r").readlines()

    for k in range(len(list)):
        if list[k].startswith("<b>Query=</b>")==True:
            if list[k+3].endswith("done\n"):
                NoHits.append(list[k][14:16])
                countnohits += 1
            elif list[k+3].endswith("done<PRE>\n"):
                Hits.append(list[k][14:16])
                counthits += 1
                getvalues(list[k+9])
            elif list[k+7].endswith("done<PRE>\n"):
                Hits.append(list[k][14:16])
                print list[k][14:16]
                counthits += 1
                getvalues(list[k+13])
            else:
                pass
        else:
            pass


#Backbone of the program:
evalues = []
scores = []
names = []

gethits(sys.argv[1])
tbx = open("TBLASTXresults","a")

#Write over-all stats for library to file...
tbx.write(sys.argv[1][-30:-16]+" ("+str(counthits+
    countnohits)+" sequences)\tNoHits("+str(countnohits)+
    ")\tHits("+str(counthits)+"):\n")
```

```
#Write particulars of each hit to file...
for k in range(len(Hits)):
    str(evalues[k]).strip()
    if evalues[k].startswith("e"):
        evalues[k] = "1"+evalues[k]
        if float(evalues[k]) < 1e-04:
            tbx.write(Hits[k]+"\t"+names[k]+"\t"+evalues
                [k]+"\t"+scores[k]+"\n")
        else:
            pass
    elif float(evalues[k]) < 1e-04:
        tbx.write(Hits[k]+"\t"+names[k]+"\t"+evalues[k]+
            "\t"+scores[k]+"\n")
    else:
        pass

tbx.write("\n\n\n")
tbx.close()
```

Listing F.2: Python script to search for targets in a blastx results html file.

```
import sys

#a list of Entrez Gene Identifiers and common names
    pertaining to genes that had been proven to be
    essential for insect development/viability...
targets = ['CYP6AE14','P450 monooxygenase','
    monooxygenase','COPI coatomer','chitinase','chitin','
    dronc','drice','tubulin','vacuolar ATPase','ATPase','
    caspase','CG6223','F38E11_5','CG8055','CG11276','
    CG2934-PA','CG3762','CG3416','F37C12_9','M03F4.2','
    CG2331','CG12770','CG6141','CG2746','CG1913_PA','
    CG3180','CG14542','CG9277','CG5440-PA','CG12055-PA','
    CG2934-PA','CG8385-PE','CG5343-PA','CG10067-PA','
    CG9357-PA','CG1913-PA','CG1088','CG16944-PD','CG8669-
    PA','CG1810','CG2331-PA','CG6141-PB','CG2746','CG8103
    -PA','CG9277']

hits = []

#load content of locally stored html file to memory...
list = open(sys.argv[1],"r").readlines()

#combinatorially iterate through target list and html
    file to find matches with proposed target genes...
for item in targets:
    for k in range(len(list)):
        if list[k].startswith('&gt;<input name='):
            z = list[k][list[k].find('blast_rank='):]
            if item in z:
                sublist = list[k:]
                for y in range(len(sublist)):
                    if sublist[y].startswith(' Score =')
                        :
                        score = sublist[y]
                    else:
```

```
                pass
            x = z.find(">")
            hit = (item,z[x:z.find('</a><a name=')],
                z[z.rfind('</a>')+4:],score)
            if hit in hits:
                pass
            else:
                hits.append((item,z[x:z.find('</a><a
                    name=')],z[z.rfind('</a>')+4:],
                    score))

print "\n\n"
#print any matches found to screen...
for x in hits:
    print x[0],"\t",x[1],"\t",x[2],"\t",x[3],"\n"
```

Listing F.3: Python script for automated generation of GenBank and dbEST submission files.

```
import sys,os

#create 'Publication File'...
f1 = open("Publication File","w")
f1.write("TYPE: Masters Dissertation\nTITLE:\
   nConstruction of a cDNA library for the vine mealybug
   , Planococcus ficus (Signoret)\nAUTHOR:\nHolm K.,
   Burger J.T.\nYEAR: 2008\nSTATUS: 2\n||")
f1.close()

#create 'Library File'...
f2 = open("Library File","w")
f2.write("TYPE: Lib\nNAME: Fractionated Library 4\
   nORGANISM: Planococcus ficus (Signoret)\nSEX: Mixed
   sex\nTISSUE: Whole body\nSTAGE: Mixed (nymphs -
   adults)\nVECTOR: pDNR-LIB\nV_TYPE: Plasmid\nRE_1:
   SfiI\nRE_2: SfiI\nDESCR:\nGenes expressed in the vine
    mealybug, Planococcus ficus (Signoret). Source:
   Breeding colony obtained from ARC Infruitec-
   Nietvoorbij (Western Cape, South Africa) and
   maintained on Cucurbita moschata, L.\n||")
f2.close()

#create 'Contact File'...
f3 = open("Contact File","w")
f3.write("TYPE: Cont\nNAME: Holm K., Burger J.T\nFAX:
   +21 21 8085833\nTEL: +21 21 8085885\nEMAIL:
   koraholm@gmail.com, jtb@sun.ac.za\nLAB: Vitis
   laboratory\nINST: Department of Genetics\nADDR:
   Department of Genetics, Stellenbosch University,
   Private Bag X1, Matieland 7602, South Africa\n||")
f3.close()
```

```
#create 'EST File' with records for all ESTs of the
    library...
f4 = open("EST File","w")

file = open(sys.argv[2],"r").readlines() #open
    concatenated fasta file
ids = open(sys.argv[1],"r").readlines()  #open file with
    hit descriptions

for i in range(len(file)):

    #for every sequence, create a record with the
        required fields...
    if file[i].startswith(">"):
        f4.write("TYPE: EST\nSTATUS: New\nCONT_NAME:
            Holm K., Burger J.T.\nCITATION:\nConstruction
             of a cDNA library for the vine mealybug,
            Planococcus ficus (Signoret)\nLIBRARY:
            Fractionated Library 4\nEST#: PfF4_"+file[i
            ][1:].replace("_M13F","")+"INSERT: "+str(len(
            file[i+1])-2)+"\nSEQ_PRIMER: M13 forward\
            nP_END: 5'\nDNA_TYPE: cDNA\nPUBLIC: \n")

        #Add putative identities to the records of
            sequences for which a hit had been found...
        for x in ids:
            id = file[i][1:].replace("_M13F","").replace
                ("\n","").replace("\r","")
            if x.startswith(id):
                s = x[4:]
                descr = s[s.find("&")+1:s.find("{")-8]
                if descr.startswith("CG"):
                    f4.write("PUT_ID: Similar to
                        Drosophila "+descr+"\n")
                else:
```

```
                    f4.write("PUT_ID: Putative "+descr+"
                        \n")
            else:
                pass

        #add the relevant fasta sequence to each record
            ...
        f4.write("SEQUENCE:\n")
        count = 0
        for l in range((len(file[i+1])/60)+1):
            f4.write(file[i+1][count:count+60].replace("
                \n","").replace("\r","")+"\n")
            count += 60
        f4.write("||\n")

    else:
        pass

f4.close()
```

Appendix G

Blast Results

The nucleotide sequences of all ESTs created in this study are available from the GenBank under the following accession numbers:

Primary Library	GE325203 - GE325297
Fractionated Library 1	GE325298 - GE325315
Fractionated Library 2	GE325316 - GE325328
Fractionated Library 3	GE325329 - GE325352
Fractionated Library 4	GE325353 - GE325370

NOTE: Access to potential target sequences has been retained until further investigations have been completed.

Table G.1: Sequences with matches in the non-redundant databases of the NCBI, but not in RefSeq (E-value $\leqslant 10^{-5}$).

Clone ID	Accession nr.	E-value	Description
3N	ref|XP_971645.2|	8.00E-05	PREDICTED: similar to Krueppel-like factor 5 [Tribolium castaneum]
3Q	ref|XP_624115.1|	9.00E-06	PREDICTED: similar to S-adenosylmethionine decarboxylase CG5029-PA, isoform A, partial [Apis mellifera]
D5	dbj|BA000019.2|	1.00E-05	Hypothetical protein alr1329 [Nostoc sp. PCC 7120]
L2	gb|AY691420.1|	1.00E-06	Planococcus ficus cytochrome b (cytb) pseudogene, partial sequence
M1	ref|XP_391943.1|	1.00E-16	PREDICTED: similar to CG6414-PA [Apis mellifera]
M3	ref|XP_966863.1|	5.00E-12	PREDICTED: similar to CG3625 CG3625-PB isoform 1 [Tribolium castaneum]
R4	gb|DQ238218.1|	1.00E-29	Planococcus ficus cytochrome c oxidase subunit 1 (COI) gene
S4	ref|XP_973426.1|	2.00E-16	PREDICTED: similar to CG10407 CG10407-PA [Tribolium castaneum]

Table G.2: Sequences for which a match was found in a query against all GenBank non-mouse and non-human EST entries (E-value $\leqslant 10^{-5}$).

Clone ID	Accession nr.	E-value	Query Coverage	Source Organism
1K	DY326929.1	2.00E-88	95%	*Ocimum basilicum*
1S	EH218847.1	7.00E-174	97%	*Maconellicoccus hirsutus*
1V	EH212674.1	1.00E-50	98%	*Maconellicoccus hirsutus*
2A	CB091950.1	1.00E-113	94%	*Cycas rumphii*
2E	EH214304.1	4.00E-68	85%	*Maconellicoccus hirsutus*
2G	DY323608.1	0.0	100%	*Ocimum basilicum*
3H	CB090509.1	3.00E-169	86%	*Cycas rumphii*
3M	CV088692.1	1.00E-54	54%	*Hepatopancreas Crassostrea*
3Q	EH216122.1	4.00E-99	51%	*Maconellicoccus hirsutus*
3R	EX949563.1	5.00E-103	48%	*Gnetum gnemon*
4M	CV580624.1	4.00E-34	39%	*Lottia gigantea*
4S	EH212618.1	0.0	93%	*Maconellicoccus hirsutus*
A6	EH213831.1	7.00E-142	92%	*Maconellicoccus hirsutus*
B6	EH218300.1	1.00E-44	38%	*Maconellicoccus hirsutus*
C5	EH218604.1	3.00E-51	31%	*Maconellicoccus hirsutus*
D3	DW515966.1	1.00E-50	69%	*Gossypium hirsutum*
H2	EH215315.1	3.00E-96	33%	*Maconellicoccus hirsutus*
I2	EH217793.1	2.00E-123	54%	*Maconellicoccus hirsutus*
K3	CB090509.1	4.00E-124	60%	*Cycas rumphii*
L3	FG803539.1	0.0	100%	*Anolis carolinensis*
M5	EH216107.1	2.00E-157	90%	*Maconellicoccus hirsutus*
Q6	ES501694.1	2.00E-63	78%	*Artemia franciscana*
R4	EH217095.1	0.0	98%	*Maconellicoccus hirsutus*
S5	DY326929.1	0.0	98%	*Ocimum basilicum*
U3	DY326213.1	0.0	99%	*Ocimum basilicum*

Table G.3: Sequences for which matches were obtained from the RefSeq databases of the NCBI (E-value $\leqslant 10^{-5}$). Descriptions of the Molecular Functions and Biological Processes were extracted from Flybase or Entrez Gene.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
1F	33214	Ribosomal protein LP1	***Molecular function:*** Structural constituent of ribosome. ***Biological processes:*** Translation; translational elongation. ***KEGG Pathway:*** Ribosome.
1K	4708	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	***Entrez Gene Summary:*** The protein encoded by this gene is a subunit of the multisubunit NADH:ubiquinone oxidoreductase (complex I). Mammalian complex I is composed of 45 different subunits. This protein has NADH dehydrogenase activity and oxidoreductase activity. It plays a important role in transfering electrons from NADH to the respiratory chain. The immediate electron acceptor for the enzyme is believed to be ubiquinone. Hydropathy analysis revealed that this subunit and 4 other subunits have an overall hydrophilic pattern, even though they are found within the hydrophobic protein (HP) fraction of complex I.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
1L	32083	Discs large 1 (DLGA)	***Molecular function:*** Guanylate kinase activity; epidermal growth factor receptor binding; protein binding; structural molecule activity. ***Biological processes:*** Described with 28 unique terms, many of which group under: anatomical structure development; regulation of biological process; macromolecule localization; protein localization; cell morphogenesis; cell proliferation; organelle organization and biogenesis; instar larval development; primary metabolic process; intercellular junction assembly. ***KEGG Pathway:*** Purine Metabolism.
1S	37646	CG3800 (Zinc-dependent nucleic acid binding protein)	***Molecular function:*** Nucleic acid binding; zinc ion binding. ***Biological processes:*** Unknown.
2E	33214	Ribosomal protein LP1	***Molecular function:*** Structural constituent of ribosome. ***Biological processes:*** Translation; translational elongation. ***KEGG Pathway:*** Ribosome.
2F	33223	Mediator complex subunit 15	***Molecular function:*** RNA polymerase II transcription mediator activity. ***Biological processes:*** Transcription initiation from RNA polymerase II promoter.
3A	35963	CG8801 (GTP-binding protein)	***Molecular function:*** GTP binding. ***Biological processes:*** Unknown.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
3D	37922	Nucleosome remodeling factor - 38kD	***Molecular function:*** Inorganic diphosphatase activity; magnesium ion binding. ***Biological processes:*** Chromatin remodeling; nucleosome mobilization; transcription; ecdysone receptor-mediated signaling pathway; phosphate metabolic process. ***KEGG Pathway:*** Oxidative phosphorylation.
3E	38447	Succinyl coenzyme A synthetase alpha subunit	***Molecular function:*** Succinate-CoA ligase (GDP-forming) activity; succinate-CoA ligase (ADP-forming) activity; binding; ATP citrate synthase activity. ***Biological processes:*** Tricarboxylic acid cycle. ***KEGG Pathway:*** Citrate cycle (TCA cycle); Propanoate metabolism.
3I	33835	Eukaryotic initiation factor 4a	***Molecular function:*** RNA helicase activity; translation initiation factor activity; RNA cap binding; mRNA binding; single-stranded DNA-dependent ATP-dependent DNA helicase activity; ATP-dependent helicase activity; ATP binding. ***Biological processes:*** Dorsal/ventral axis specification; translational initiation; imaginal disc growth; instar larval development; DNA unwinding during replication; regulation of alternative nuclear mRNA splicing, via spliceosome; mitotic spindle elongation; mitotic spindle organization and biogenesis.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
3J	34884	Vasa	***Molecular function:*** RNA helicase activity; ATP-dependent helicase activity; RNA binding; protein binding; ATP binding. ***Biological processes:*** Described with 16 unique terms, many of which group under: regulation of metabolic process; gamete generation; anterior/-posterior axis specification; anatomical structure development; cellular macromolecule metabolic process; oocyte axis determination; positive regulation of cellular biosynthetic process; oocyte anterior/posterior axis determination; pole cell development; germ-line cyst formation; karyosome formation; intracellular mRNA localization; mRNA metabolic process. ***KEGG Pathway:*** Purine metabolism.
3M	47218	Ribosomal protein S14a	***Molecular function:*** Structural constituent of ribosome. ***Biological processes:*** Translation.
3O	3772367	Mitochondrial ribosomal protein L53	***Molecular function:*** Unknown. ***Biological processes:*** Unknown.
3S	37744	Lethal (2) essential for life	***Molecular function:*** Unknown. ***Biological processes:*** Embryonic development; response to heat.
3U	33751	CG14034 (Protein with phospholipase activity)	***Molecular function:*** Phospholipase activity. ***Biological processes:*** Phospholipid metabolic process; lipid metabolic process.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
3V	32661	Inflated	***Molecular function:*** Cell adhesion molecule binding; receptor activity; protein binding. ***Biological processes:*** Described with 17 unique terms, many of which group under: anatomical structure development; cell adhesion; open tracheal system development; cell motility; organ development; cell-cell adhesion; organ morphogenesis; maintenance of protein localization; cell-substrate adhesion; heterophilic cell adhesion; imaginal disc-derived wing margin morphogenesis; macromolecule localization; muscle cell differentiation; gut development; regulation of cell morphogenesis.
3W	37646	CG3800 (Zinc-dependent nucleic acid binding protein)	***Molecular function:*** Nucleic acid binding; zinc ion binding. ***Biological processes:*** Unknown.
3X	37451	Misexpression suppressor of KSR 2 (MESK2)	***Molecular function:*** Unknown. ***Biological processes:*** Unknown.
4A	43255	Neprilysin 5	***Molecular function:*** Metalloendopeptidase activity; neprilysin activity; zinc ion binding. ***Biological processes:*** Proteolysis.
4B	36031	CG12140 (Electron-transferring-flavoprotein dehydrogenase)	***Molecular function:*** Electron-transferring-flavoprotein dehydrogenase activity. ***Biological processes:*** Oxidative phosphorylation.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
4F	40441	CG7458 (Protein with secondary active organic cation transmembrane transporter activity)	***Molecular function:*** Secondary active organic cation transmembrane transporter activity. ***Biological processes:*** Transport.
4J	34546	CG6230 (Cation transport ATPase)	***Molecular function:*** ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism; hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances; ATP binding. ***Biological processes:*** Cation transport; metabolic process.
4K	42100	CG5873 (Peroxidase)	***Molecular function:*** Peroxidase activity; heme binding. ***Biological processes:*** Response to oxidative stress.
4L	32083	Discs large 1 (DLGA)	***Molecular function:*** Guanylate kinase activity; epidermal growth factor receptor binding; protein binding; structural molecule activity. ***Biological processes:*** Described with 28 unique terms, many of which group under: anatomical structure development; regulation of biological process; macromolecule localization; protein localization; cell morphogenesis; cell proliferation; organelle organization and biogenesis; instar larval development; primary metabolic process; intercellular junction assembly.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
4M	38456	Ubiquitin-63E	***Molecular function:*** Protein binding. ***Biological processes:*** Described with 11 unique terms, many of which group under: cellular macromolecule metabolic process; primary metabolic process; biopolymer modification; response to stress; organelle organization and biogenesis; protein modification process; transport; establishment and/or maintenance of chromatin architecture; catabolic process; regulation of metabolic process; cellular process.
4N	34884	Vasa	***Molecular function:*** RNA helicase activity; ATP-dependent helicase activity; RNA binding; protein binding; ATP binding. ***Biological processes:*** Described with 16 unique terms, many of which group under: regulation of metabolic process; gamete generation; anterior/-posterior axis specification; anatomical structure development; cellular macromolecule metabolic process; oocyte axis determination; positive regulation of cellular biosynthetic process; oocyte anterior/posterior axis determination; pole cell development; germ-line cyst formation; karyosome formation; intracellular mRNA localization; mRNA metabolic process. ***KEGG Pathway:*** Purine metabolism.
4Q	37427	CG10069 (Glycerol-3-phosphate transporter)	***Molecular function:*** Glycerol-3-phosphate transmembrane transporter activity. ***Biological processes:*** Transport.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
4R	40556	CG9855 (RING-finger protein)	***Molecular function:*** Protein binding; zinc ion binding. ***Biological processes:*** Unknown.
4S	31521	Actin 5C	***Molecular function:*** Structural constituent of cytoskeleton; neurotransmitter transporter activity; ATP binding; protein binding. ***Biological processes:*** Cytoskeleton organization and biogenesis; sperm individualization; cytokinesis; phagocytosis, engulfment; neurotransmitter transport; visual behavior. ***KEGG Pathway:*** Cell Communication.
4T	40443	CG7470 (Oxidoreductase Transferase)	***Molecular function:*** Delta1-pyrroline-5-carboxylate synthetase activity; glutamate 5-kinase activity; glutamate-5-semialdehyde dehydrogenase activity. ***Biological processes:*** Proline biosynthetic process.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
4U	38231	Alpha Spectrin	***Molecular function:*** Cytoskeletal protein binding; actin binding; calcium ion binding; calmodulin binding; microtubule binding. ***Biological processes:*** Described with 20 unique terms, many of which group under: anatomical structure development; synaptic transmission; organelle organization and biogenesis; gamete generation; regulation of cellular component organization and biogenesis; germ-line cyst formation; regulation of developmental process; cell-cell signaling; ovarian follicle cell development; cell adhesion; fusome organization and biogenesis; instar larval development; plasma membrane organization and biogenesis; ovarian nurse cell to oocyte transport.
4V	3771968	Muscle-specific protein 300	***Molecular function:*** Cytoskeletal protein binding; actin binding; double-stranded RNA binding. ***Biological processes:*** Cytoskeleton organization and biogenesis; mesoderm development; cytoplasmic transport, nurse cell to oocyte; establishment of nucleus localization; ovarian ring canal stabilization; actin filament organization.
A3	33429	CG18641 (Lipase)	***Molecular function:*** Lipase activity. ***Biological processes:*** Lipid metabolic process.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
A5	33781	Helicase at 25E	***Molecular function:*** RNA helicase activity; RNA splicing factor activity, transesterification mechanism; ATP-dependent RNA helicase activity; ATP binding; nucleic acid binding. ***Biological processes:*** Nuclear mRNA splicing, via spliceosome; mRNA export from nucleus; regulation of alternative nuclear mRNA splicing, via spliceosome; mitotic spindle organization and biogenesis. ***KEGG Pathway:*** Purine metabolism.
A6	43693	CG1746 (ATP synthesis coupled proton transporter)	***Molecular function:*** Hydrogen-exporting ATPase activity, phosphorylative mechanism; hydrogen ion transporting ATPase activity, rotational mechanism; hydrogen ion transporting ATP synthase activity, rotational mechanism. ***Biological processes:*** Proton transport; ATP synthesis coupled proton transport. ***KEGG Pathway:*** Oxidative phosphorylation.
B4	246560	CG30354 (Protein with ubiquinol-cytochrome-c reductase activity)	***Molecular function:*** Ubiquinol-cytochrome-c reductase activity. ***Biological processes:*** Mitochondrial electron transport, ubiquinol to cytochrome c. ***KEGG Pathway:*** Oxidative phosphorylation.
D2	31528	CG6048 (Serine protease)	***Molecular function:*** Trypsin activity; serine-type endopeptidase activity. ***Biological processes:*** Proteolysis.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
D3	34149	Ribosomal protein S13	***Molecular function:*** Structural constituent of ribosome. ***Biological processes:*** Translation; mitotic spindle elongation; mitotic spindle organization and biogenesis. ***KEGG Pathway:*** Ribosome.
D6	5641	Legumain	***Entrez Gene Summary:*** This gene encodes a cysteine protease that has a strict specificity for hydrolysis of asparaginyl bonds. This enzyme may be involved in the processing of bacterial peptides and endogenous proteins for MHC class II presentation in the lysosomal/endosomal systems. Enzyme activation is triggered by acidic pH and appears to be autocatalytic. Protein expression occurs after monocytes differentiate into dendritic cells. A fully mature, active enzyme is produced following lipopolysaccharide expression in mature dendritic cells. Overexpression of this gene may be associated with the majority of solid tumor types. This gene has a pseudogene on chromosome 13. Several alternatively spliced transcript variants have been described, but the biological validity of only two has been determined. These two variants encode the same isoform.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
E3	31849	Oligosaccharyltransferase 48kD subunit	***Molecular function:*** Dolichyl-diphosphooligosaccharide-protein glycotransferase activity. ***Biological processes:*** Protein amino acid N-linked glycosylation; protein amino acid N-linked glycosylation via asparagine. ***KEGG Pathway:*** Glycan structures - biosynthesis 1; N-Glycan biosynthesis.
H2	33834	Chickadee	***Molecular function:*** Actin binding; phosphatidylinositol-4,5-bisphosphate binding. ***Biological processes:*** Described with 14 unique terms, many of which group under: anatomical structure development; organelle organization and biogenesis; actin filament-based process; cellular macromolecule metabolic process; gamete generation; cell division; central nervous system development; organ development; cell cycle; sexual reproduction; system development; ovarian nurse cell to oocyte transport; karyosome formation; anterior/-posterior axis specification.
I2	33288	Smell impaired 21F	***Molecular function:*** Unknown. ***Biological processes:*** Olfactory behavior; response to chemical stimulus.
I6	40780	CG17919 (Phosphatidylethanolamine binding protein)	***Molecular function:*** Phosphatidylethanolamine binding. ***Biological processes:*** Unknown.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
J1	40069	CG3819 (Endonuclease)	***Molecular function:*** Endonuclease activity; nucleic acid binding. ***Biological processes:*** Unknown.
J5	326232	Transport and Golgi organization 5 (Tango5)	***Molecular function:*** Unknown. ***Biological processes:*** Golgi organization and biogenesis.
K2	5740554	CG34448 (Protein with triacylglycerol lipase activity)	***Molecular function:*** Triacylglycerol lipase activity. ***Biological processes:*** Lipid metabolic process.
K5	36973	CG4847 (Cathepsin)	***Molecular function:*** Cysteine-type peptidase activity; cathepsin K activity. ***Biological processes:*** Proteolysis.
M5	37628	Ribosomal protein L23	***Molecular function:*** Protein binding; structural constituent of ribosome. ***Biological processes:*** Translation; mitotic spindle elongation; mitotic spindle organization and biogenesis. ***KEGG Pathway:*** Ribosome.
N5	40443	CG7470 (Protein involved in the proline biosynthetic process)	***Molecular function:*** Delta1-pyrroline-5-carboxylate synthetase activity; glutamate 5-kinase activity; glutamate-5-semialdehyde dehydrogenase activity. ***Biological processes:*** Proline biosynthetic process.
O5	38794	Ribosomal protein L18	***Molecular function:*** Structural constituent of ribosome. ***Biological processes:*** Translation; mitotic spindle elongation; mitotic spindle organization and biogenesis. ***KEGG Pathway:*** Ribosome.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
O6	31536	CG5941 (RNA-binding protein)	***Molecular function:*** RNA binding. ***Biological processes:*** Unknown.
P1	5740813	CG34447 (Lipase)	***Molecular function:*** Triacylglycerol lipase activity. ***Biological processes:*** Lipid metabolic process.
Q1	35910	CG8193 (Phenoloxidase)	***Molecular function:*** Monophenol monooxygenase activity; oxygen transporter activity. ***Biological processes:*** Defense response; metabolic process; transport.
Q6	31914	CG1354 (GTP-binding protein)	***Molecular function:*** GTP binding; hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides. ***Biological processes:*** Unknown.
R2	40946	CG2767 (Aldo/keto reductase)	***Molecular function:*** Alcohol dehydrogenase (NADP+) activity. ***Biological processes:*** Unknown.
R3	33981	Small ubiquitin-related modifier (SUMO) / sentrin / smt3	***Molecular function:*** Protein binding. ***Biological processes:*** Anterior/posterior pattern formation; protein modification process; protein import into nucleus; phagocytosis, engulfment.
R6	41270	Rrp46	***Molecular function:*** 3'-5'-exoribonuclease activity; RNA binding. ***Biological processes:*** mRNA processing.
S1	33751	CG14034 (Lipase)	***Molecular function:*** Phospholipase activity. ***Biological processes:*** Phospholipid metabolic process; lipid metabolic process.

Clone ID	Gene ID	Gene Description	Molecular Function and Biological Processes
S5	4708	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 2, 8kDa	***Entrez Gene Summary:*** The protein encoded by this gene is a subunit of the multisubunit NADH:ubiquinone oxidoreductase (complex I). Mammalian complex I is composed of 45 different subunits. This protein has NADH dehydrogenase activity and oxidoreductase activity. It plays a important role in transfering electrons from NADH to the respiratory chain. The immediate electron acceptor for the enzyme is believed to be ubiquinone. Hydropathy analysis revealed that this subunit and 4 other subunits have an overall hydrophilic pattern, even though they are found within the hydrophobic protein (HP) fraction of complex I.
U4	37846	eIF-5A	***Molecular function:*** Translation initiation factor activity; translation regulator activity. ***Biological processes:*** Translational initiation; autophagic cell death; salivary gland cell autophagic cell death.

Bibliography

[1] Collard, T (MD of Cape-base direct wine marketer, Wine of the Month Club) Press Release: Sept 27 (2006). URL `http://www.fin24.co.za/`.

[2] Martelli G (2000). Major Graft-Transmissible Diseases of Grapevines: Nature, Diagnosis and Sanitation. American Society for Enology and Viticulture (ASEV), Washington State Convention and Trade Center, Seattle, Washington, June 19-23.

[3] Robinson A, Franz G, Atkinson P (2004) Insect transgenesis and its potential role in agriculture and human health. Insect Biochemistry and Molecular Biology 34:113–20.

[4] Cabaleiro C, Segura A, Garcia-Berrios J (1999) Effects of Grapevine Leafroll-Associated Virus 3 on the Physiology and Must of *Vitis vinifera* L. cv. Albarino Following Contamination in the Field. American Journal of Enology and Viticulture 50:40–44.

[5] Fuchs MF (2007) Grape Leafroll Disease. Factsheet, Integrated Pest Management, Cornell University. URL `www.nysipm.cornell.edu/factsheets/grapes/diseases/grape`$_{l}eafroll.pdf$.

[6] Martelli G, Agranovsky A, Bar-Joseph M, Boscia D, Candresse T, *et al.* (2002) The family Closteroviridae revised. Archives of Virology 147:2039–44.

[7] Pietersen G (2004) Spread of Grapevine Leafroll Disease in South Africa – A Difficult, but Not Insurmountable Problem. WineLand, June :110–113.

[8] Cabaleiro C, Segura A (2006) Temporal Analysis of Grapevine leafroll associated virus 3 Epidemics. European Journal of Plant Pathology 114:441–446.

[9] Viticulture Survey Report (2006) American Vineyard Foundation. URL `http://avf.org/survey.html`.

[10] Farrell R (2005) RNA Methodologies. 3rd edition. Amsterdam: Elsevier Academic Press.

[11] Sambrook J, Russel D (2001) Molecular Cloning: A Laboratory Manual, volume 2. 3rd edition. New York: Cold Spring Harbor Laboratory Press.

[12] Creator™ SMART™ cDNA Library Construction Kit User Manual, CLONTECH Laboratories, Inc. Protocol #PT3577-1, version #PR12792 URL `www.clontech.com`.

[13] Ben-Dov Y (1994) A Systematic Catalogue of the Mealybugs of the World (Insecta: Homoptera: Coccoidea: Pseudococcidae and Putoidae) with data on Geographical Distribution, Host Plants, Biology and Economic Importance. Andover: Intercept Limited.

[14] De Lotto G (1975) Notes on the vine mealybug (Homoptera: Coccidea: Pseudococcoidae). Journal of the Entomological Society of South Africa 38:125–130.

[15] Walton VM (2003) Development of an Integrated Pest Management system for vine mealybug, *Planococcus ficus* (Signoret), in vineyards of the Western Cape Province, South Africa. Ph.D. thesis, University of Stellenbosch.

[16] Kriegler PJ (1954) 'n Bydrae tot die kennis van *Planococcus citri* (Risso) (Homoptera: Pseudococcidae). Master's thesis, University of Stellenbosch, South Africa.

[17] Whitehead VB (1957) A study of the Predators and Parasites of *Planococcus citri* (Risso) (Homoptera) on Vines in the Western Cape Province, South Africa. Master's thesis, Rhodes University, Grahamstown.

[18] Walton VM, Pringle KL (2004) Vine mealybug, *Planococcus ficus* (Signoret) (Hemiptera : Pseudococcidae), a key pest in South African vineyards . A Review. South African Journal for Enology and Viticulture 25:54–62.

[19] Grape: Information about Relative Toxicities of Insecticides and Miticides Used in Grapes to Natural Enemies and Honey Bees (2006). URL `http://www.ipm.ucdavis.edu/PMG/r302900111.html`.

[20] Walton V, Pringle K (2005) Developmental biology of the vine mealybug, *Planococcus ficus* (Signoret) (Homoptera: Pseudococcidae), and its parasitoid *Coccidoxenoides perminutus* (Timberlake) (Hymenoptera: Encyrtidae). African Entomology 13:143–147.

[21] Walton VM, Marais E (2003) Mealybug on Grapevines. DFPT Research Pamphlet, ARC Infruitec-Nietvoorbij, Stellenbosch.

[22] Cabaleiro C, Segura A (1997) Some characteristics of the transmission of grapevine leafroll associated virus 3 by Planococcus citri Risso. European Journal of Plant Pathology 103:373–378.

[23] Kruger K, Saccaggi D, Douglas N (2006). Grapevine leafroll associated virus 3 – vector interactions: Transmission by the mealybugs Planococcus ficus and Pseudococcus longispinus (Hemiptera: Pseudococcidae). 15th Meeting of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG), Stellenbosch, South Africa, 3-7 April, pp. 130-131.

[24] Minafra A, Hadidi A (1994) Sensitive detection of grapevine virus A, B, or leafroll-associated III from viruliferous mealybugs and infected tissue by cDNA amplification. Journal of Virological Methods 47:175–188.

[25] UC IPM Online (2006) Grape: Vine Mealybug, How to Manage Pests. URL `http://www.ipm.ucdavis.edu/PMG/r302301911.html`.

[26] Protocol for control of mealybug in vineyards (production blocks). (2006) Wynboer, September issue. URL `http://www.wynboer.co.za/recentarticles/200609protokol.php3`.

[27] Vine Mealybug Update. Technical report, Grape Notes, February 2003, University of California Cooperative Extension.

[28] Daane KM, Bentley JW, Walton VM, Malakar-Kuenen R, Millar JG, *et al.* (2006). New controls investigated for vine mealybug. URL `http://calag.ucop.edu/0601JFM/pdfs/VineMealybug.pdf`.

[29] Walton VM, Daane KM, Pringle KL (2004) Monitoring *Planococcus ficus* in South African vineyards with sex pheromone-baited traps. Crop Protection 23:1089–1096.

[30] Walton VM, Pringle KL (1999) Effects of pesticides used on table grapes on the mealybug parasitoid *Coccidoxenoides peregrinus* (Timberlake) (Hymenoptera : Encyrtidae). South African Journal for Enology and Viticulture 20:31–34.

[31] Cloyd R, Dickinson A (2006) Effect of insecticides on mealybug destroyer (Coleoptera: Coccinellidae) and parasitoid *Leptomastix dactylopii* (Hymenoptera: Encyrtidae), natural enemies of citrus mealybug (Homoptera: Pseudococcidae). Journal of Economic Entomology 99:1596–604.

[32] Walton VM, Pringle KL (2004) A Survey of Mealybugs and Associated Natural Enemies in Vineyards in the Western Cape Province, South Africa. South African Journal for Enology and Viticulture 25:23–25.

[33] Hinkens DM, S MJ, Millar JG (2001) Identification and synthesis of the sex pheromone of the vine mealybug, *Planococcus ficus*. Tetrahedron Letters 42:1619–1621.

[34] Walton VM, Daane KM, Bentley WJ, Millar JG, Larsen TE, *et al.* (2006) Pheromone-based mating disruption of *Planococcus ficus* (Hemiptera: Pseudococcidae) in California vineyards. Journal of Economic Entomology 99:1280–1290.

[35] Daane KM, Bentley JW, Walton VM, Malakar-Kuenen R, Millar JG, *et al.* (2003). Current Status of the Vine Mealybug, *Planococcus ficus*, in California. URL `http://ucce.ucdavis.edu/files/filelibrary/1198/10998.pdf`.

[36] Haviland D, Bentley W, Daane K (2005) Hot-water treatments for control of *Planococcus ficus* (Homoptera: Pseudococcidae) on dormant grape cuttings. Journal of Economic Entomology 98:1109–1115.

[37] Alphey N, Coleman P, Donnelly C, Alphey L (2007) Managing insecticide resistance by mass release of engineered insects. Journal of Economic Entomology 100:1642–9.

[38] Christou P, Capell T, Kohli A, Gatehouse J, Gatehouse A (2006) Recent developments and future prospects in insect pest control in transgenic crops. Trends in Plant Science 11:302–308.

[39] Fischhoff D, Bowdish K, Perlak F, Marrone P, McCormick S, *et al.* (1987) Insect tolerant transgenic tomato plants. Nature Biotechnology 5:807–813.

[40] Ferry N, Edwards M, Gatehouse J, Capell T, Christou P, *et al.* (2006) Transgenic plants for insect pest control: a forward looking scientific perspective. Transgenic Research 15:13–19.

[41] Gordon K, Waterhouse P (2007) RNAi for insect-proof plants. Nature Biotechnology 25:1231–1232.

[42] Mao Y, Cai W, Wang J, Hong G, Tao X, *et al.* (2007) Silencing a cotton bollworm P450 monooxygenase gene by plant-mediated RNAi impairs larval tolerance of gossypol. Nature Biotechnology 25:1307–1313.

[43] Baum J, Boggaert T, Clinton W, Heck G, Feldmann P, *et al.* (2007) Control of coleopteran insect pests through RNA interference. Nature Biotechnology 25:1322–1326.

[44] Price D, Gatehouse J (2008) RNAi-mediated crop protection against insects. Trends in Biotechnology 26:393–400.

[45] Chandra H (1963) Cytogenetic studies following high dosage paternal irradiation in the mealybug *Planococcus citri*. Chromosoma 14:330–346. doi: 10.1007/BF00326819.

[46] The NEW Genome Sequencer FLX System – More Flexibility, More Applications, Roche Applied Science. URL `www.genome-sequencing.com`.

[47] Emrich S, Barbazuk W, Li L, Schnable P (2007) Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research 17:69–73.

[48] Kimmel A (1987) Selection of clones from libraries: Overview. In: Berger SL and Kimmel, AR (editors), Methods in Enzymology 152, Orlando: Academic Press .

[49] Bothwell A, Yancopoulos G, Alt F (1990) Methods for Cloning and Analysis of Eukaryotic Genes. Boston: Jones and Bartlett.

[50] The Basics: RNA Isolation. Technical resources (general articles), Ambion. URL `http://www.ambion.com/techlib/basics/ rnaisol/index.html`.

[51] Chomczynski P (1992) Solubilization in formamide protects RNA from degradation. Nucleic Acids Research 20:3791–3792.

[52] Working with RNA. Technical Bulletin 154, Ambion. URL `http://www.ambion.com/techlib/tb/tb`$_1$59.*html*.

[53] Tips for successful RNA amplification (Palmer, M). Technotes11(3), Ambion. URL `http://www.ambion.com/techlib/tn/113/7.html`.

[54] The Basics: RNase Control. Technical resources (general articles), Ambion. URL `http://www.ambion.com/techlib/basics/rnasecontrol/index.html`.

[55] Ying S, editor (2003) Generation of cDNA Libraries: Methods and Protocols, volume 221 of *Methods in Molecular Biology*. Humana Press.

[56] Soares M, Bonaldo M, Jelene P, Su L, Lawton L, *et al.* (1994) Construction and characterization of a normalized cDNA library. Proceedings of the National Academy of Sciences 91:9228–32.

[57] Draper M, August P, Connolly T, Packard B, Call K (2002) Efficient Cloning of Full-Length cDNAs Based on cDNA Size Fractionation. Genomics 79:603–607.

[58] M-MLV Reverse Transcriptase, RNase H Minus. Technical Resources, Promega Corporation, revised 5/05. URL `http://www.promega.com/tbs/9pim368/9pim368.html`.

[59] Krug M, SL B (1987) First-strand cDNA synthesis primed with oligo(dT). In: Berger SL and Kimmel, AR (editors), Methods in Enzymology 152, Orlando: Academic Press .

[60] Gubler U (1987) Second-strand cDNA synthesis: mRNA fragments as primers. In: Berger SL and Kimmel, AR (editors), Methods in Enzymology 152, Orlando: Academic Press .

[61] Get the 5'-End. Technotes 7(3), Ambion. URL `http://www.ambion.com/techlib/tn/73/731.html`.

[62] Scotto-Lavino E, Du G, Frohman M (2006) 5' end cDNA amplification using classic RACE. Nature Protocols 1:2555–2562.

[63] Hu Y, Balaskas E, Gill P, Zeibdawi A, Smeenk C, *et al.* (1998). Terminal deoxynucleotidyl transferase (TDT) activity of retroviral reverse transcriptase (RT) and its association with DNA strand transfer. 5th Conference on Retroviruses and Opportunistic Infections, Chicago IL, Feb 1-5, 135 (abstract no. 299).

[64] Ouhammouch M, Brody E (1992) Temperature-dependent template switching during in vitro cDNA synthesis by the AMV-reverse transcriptase. Nucleic Acids Research 20:5443–5450.

[65] Schmidt W, Mueller M (1999) CapSelect: a highly sensitive method for 5'CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. Nucleic Acids Research 27:e31.

[66] Jendrisak J, Young R, Engel J (1987) Cloning cDNA into λgt10 and λgt11. In: Berger SL and Kimmel, AR (editors), Methods in Enzymology 152, Orlando: Academic Press .

[67] Dower W, Miller J, Ragsdale C (1988) High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Research 16:6127–6145.

[68] Ausubel F (2004) Purification of Bacteriophage Clones. Current Protocols in Molecular Biology, unit 6.5. New York: John Wiley & Sons.

[69] Munroe D, Loebbert R, Bric E, Whitton T, Prawitt D, *et al.* (1995) Systematic Screening of an Arrayed cDNA Library by PCR. Proceedings of the National Academy of Sciences 92:2209–2213. doi:10.1073/pnas.92.6.2209.

[70] Yu W, Andersson B, Worley K, Muzny D, Ding Y, *et al.* (1997) Large-Scale Concatenation cDNA Sequencing. Genome Research 7:353–358.

[71] Haerry T, O'Connor M (2002) Isolation of *Drosophila* Activin and Follistatin cDNAs using novel MACH amplification protocols. Gene 291:85–93.

[72] Hoskins R, Stapleton M, George R, Yu C, Wan K, *et al.* (2005) Rapid and efficient cDNA library screening by self-ligation of inverse PCR products (SLIP). Nucleic Acids Research 33:e185.

[73] Chirgwin J, Przybyla A, MacDonald R, Rutter W (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–9.

[74] Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Research 22:4673–4680.

[75] Online Appendix - Ribosomal RNA Sizes. Technotes 11 (1), Ambion. URL `http://www.ambion.com/techlib/append/ribosomalRNAsizes.html`.

[76] Fujiwara H, Ishikawa H (1986) Molecular mechanism of introduction of the hidden break into the 28S rRNA of insects: implication based on structural studies. Nucleic Acids Research 14:6393–6401.

[77] Absorbance Profile of RNA Before and After Treatment with TURBO DNA-*free*™. Technical report, Ambion. URL `http://www.ambion.com/catalog/supp/absorbance.html`.

[78] Chenchik A, Zhu Y, Diatchenko L, Li R, Hill J, *et al.* (1998) Generation and use of high quality cDNA from small amounts of total RNA by SMART PCR. RT-PCR methods for gene cloning and analysis. In: Siebert, P and Larrick, J (editors), BioTechniques Books, MA, USA :305–319.

[79] Wang J, Jemielity S, Uva P, Wurm Y, Graff J, *et al.* (2007) An annotated cDNA library and microarray for large-scale gene-expression studies in the ant *Solenopsis invicta*. Genome Biology 8:R9. doi:10.1186/gb-2007-8-1-r9.

[80] Lehane M, Aksoy S, Gibson W, Kerhornou A, Berriman M, *et al.* (2003) Adult midgut expressed sequence tags from the tsetse fly *Glossina morsitans morsitans* and expression analysis of putative immune response genes. Genome Biology 4:R63. doi:10.1186/gb-2003-4-10-r63.

[81] Rubin G, Hong L, Brokstein P, Evans-Holm M, Frise E, *et al.* (2000) A *Drosophila* complementary DNA resource. Science 287:2222–2224.

[82] Stapleton M, Carlson J, Brokstein P, Yu C, Champe M, *et al.* (2002) A *Drosophila* full-length cDNA resource. Genome Biology 3:research0080.1–80.8. doi:10.1186/gb-2002-3-12-research0080.

[83] Wellenreuther R, Schupp I, Poustka A, Wiemann S, (2004) SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. BMC Genomics 5:36.

[84] Chen B, Janes H, editors (2002) PCR Cloning Protocols. Methods in Molecular Biology, 2 edition. Totowa, NJ: Humana Press.

[85] Carninci P, Shibata S, Hayatsu N, Itoh M, Shiraki T, *et al.* (2001) Balanced-Size and Long-Size Cloning of Full-Length, Cap-Trapped cDNAs into Vectors of the Novel λ-FLC Family Allows Enhanced Gene Discovery Rate and Functional Analysis. Genomics 77:79–90.

[86] Sheng Y, Mancino V, Birren B (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. Nucleic Acids Research 23:1990–1996. doi:10.1093/nar/23.11.1990.

[87] Wentzel L, Nobbs T, Halford S (1995) The *Sfi*I restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. Journal of Molecular Biology 248:581–595.

[88] Halford S (2001) Hopping, jumping and looping by restriction enzymes (Novartis Medal Lecture). Biochemical Society Transactions 29:363–373.

[89] Cortesero A, Stapel J, Lewis W (2000) Understanding and Manipulating plant Attributes to Enhance Biological Control. Biological Control 17:35–49.

[90] Ferry N, Edwards M, Gatehouse J, Gatehouse A (2004) Plant–insect interactions: molecular approaches to insect resistance. Current Opinion in Biotechnology 15:155–161.

[91] Schuler M (1996) The Role of Cytochrome P450 Monooxydases in Plant-Insect Interactions. Plant Physiology 112:1411–1419.

[92] Musser R, Hum-Musser S, Eichenseer H, Peiffer M, Ervin G, *et al.* (2002) Herbivory: caterpillar saliva beats plant devenses - a new weapon emerges in the evolutionary arms race between plants and herbivores. Nature 416:599–600.

[93] Moran P, Thompson G (2001) Molecular responses to aphid feeding in *Arabidopsis* in relation to plant defensive pathways. Plant Physiology 125:1074–1085.

[94] Walling L (2000) The myriad plant responses to herbivores. Journal of Plant Growth Regulation 19:195–216.

[95] Espinoza C, Vega A, Medina C, Schlauch K, Cramer G, *et al.* (2007) Gene expression associated with compatible viral diseases in grapevine cultivars. Functional Integrative Genomics 7:95–110.

[96] Cid M, Pereira S, Cabaleiro C, Faoro F, Segura A (2007) Presence of Grapevine leafroll-associated virus 3 in primary salivary glands of the mealybug vector *Planococcus citri* suggests a circulative transmission mechanism. European Journal of Plant Pathology 118:23–30.

[97] Van der Heuvel J, Bruyere A, Hogenhout S, Ziegler-Graff V, Brault V, *et al.* (1997) The N-terminal region of the luteovirus read through domain determines virus binding to Buchnera GroEL and is essential for virus persistence in the aphid. Journal of Virology 71:7258–7265.

[98] Morin S, Ghanim M, Zeidan M, Czosnek H, Verbeek M, *et al.* (1999) A GroEL homologue from endosymbiotic bacteria of the whitefly *Bemisia tabaci* is implicated in the circulative transmission of Tomato yellow leaf curl virus. Virology 25:75–84.

[99] Bennion B, Dagget V (2003) The molecular basis for the chemical denaturation of proteins by urea. Proceedings of the National Academy of Sciences 100:5142–5147.

[100] Moosavi-Movahedi A (2005) Thermodynamics of Protein Denaturation by Sodium Dodecyl Sulfate. Journal of the Iranian Chemical Society 2:189–196.

[101] Monera O, Kay C, Hodges R (1994) Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. Protein Science 3:1984–1991.

[102] Loosemore M, Pratt R (1976) The irreversible cleavage of histidine residues by diethylpyrocarbonate (ethoxyformic anhydride). FEBS Letters 72:155–158.

[103] Burns JA, Butler JC, Moran J, Whitesides GM (1991) Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. The Journal of Organic Chemistry 56:2648–2650.

[104] Thompson E, O'Donnel I (1961) Quantitative reduction of disulphide bonds in proteins using high concentrations of mercaptoethanol. Biochimica et biophysica acta 28:447–449.

[105] Valetti C, Sitia R (1994) The differential effects of dithiothreitol and 2-mercaptoethanol on the secretion of partially and completely assembled immunoglobulins suggest that thiol-mediated retention does not take place in or beyond the Golgi. Molecular Biology of the Cell 5:1311–1324.

[106] Puskas R, Manley N, Wallace D, Berger S (1982) Effect of Ribonucleoside-Vanadyl Complexes on Enzyme-Catalyzed Reactions Central to Recombinant Deoxyribonucleic Acid Technology. Biochemistry 21:4602–4608.

[107] Dieffenbach C, Dveksler G, editors (2003) PCR Primer: A Laboratory Manual. 2nd edition. New York: Cold Spring Harbor Laboratory Press.

[108] Ebeling W, Heinrich N, Klockow M, Metz H, Orth H, *et al.* (1974) Proteinase K from *Tritirachium album* limber. European Journal of Biochemistry 47:91.

[109] Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology 166:1557–1580.