Development of molecular techniques to identify mealybugs (Hemiptera: Pseudococcidae) of importance on grapevine in South Africa

Davina Luisa Saccaggi

Development of molecular techniques to identify mealybugs (Hemiptera: Pseudococcidae) of importance on grapevine in South Africa

Davina Luisa Saccaggi

Submitted in partial fulfillment of the requirements for the degree of

Magister Scientiae

October 2005

Supervisor: Dr. K. Krüger
Co-supervisor: Prof. G. Pietersen

Department of Zoology and Entomology
Faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria

DECLARATION

I, Davina Luisa Saccaggi, declare that the thesis which I hereby submit for the degree of Magister Scientiae at the University of Pretoria is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

_____

Signature

_____

Date

1982:

In wine grapes [mealybugs] are not regarded as important pests. Their pest status in wine grapes may well be underestimated.

Annecke and Moran

1983:

Adults and instars of *P. longispinus* were able to transmit GVA from naturally infected vines to both herbaceous plants and virus-free grapes.

Rosciglione *et al.*

2002:

The vine mealybug is an economic pest of vineyards in the Mediterranean regions of Europe, Africa, and in the Middle East, as well as in South Africa, Pakistan, Argentina, and the southeastern United States.

Bettiga

Acknowledgements

Many individuals and organizations contributed to the completion of this work. Their input is sincerely appreciated.

Project funding was provided by the University of Pretoria, Winetech and THRIP (Technology and Human Resources for Industry Programme).

A number of individuals assisted with sample collection and manipulation as well as data analyses. N. Douglas helped extensively with field collections of mealybugs. Mealybug specimens from colonies of *P. citri*, *P. ficus* and *P. viburni* were kindly provided by J. Meyer (Agricultural Research Council – Plant Protection Research Institute, ARC-PPRI), K. Krüger (University of Pretoria) and V. Walton (Stellenbosch University), respectively. Dr. Walton also allowed me to use his field-collected mealybugs to test the multiplex PCR. I. Millar (Biosystematics Division, ARC-PPRI) identified sub-samples of all mealybug collections used in this study. C. MacDonald provided extensive training in molecular methods during the initial stages of the project. G. Malherbe assisted with trouble-shooting and optimization during the latter part of the project. Many other students in the department, in particular H. Brettschneider and M. Warren, were a source of encouragement and often acted as "sounding boards" for new ideas.

My supervisors, Dr. Kerstin Krüger and Prof. Gerhard Pietersen, were always willing to provide expert opinion, insight, support and encouragement. Their enthusiasm made the work load lighter and fueled my continued interest in the subject.

My family – both by blood and by marriage – showed continued interest, encouragement and patience during every stage of this study. My husband, in particular, demonstrated endless patience and interest in a subject entirely foreign to him, and proved an ongoing source of inspiration as he looked at tasks from a different perspective.

Finally, in acknowledgement of the privilege and blessing it is to be able to pursue knowledge, I give thanks to my Heavenly Father. The glory of God truly is intelligence.

Summary

Mealybugs (Hemiptera: Pseudococcidae) cause severe damage to many commercial crops, including grapevine. This is largely because of their ability to transmit various grapevine viral diseases, in particular grapevine leafroll-associated viruses (GLRaVs). Grapevine leafroll is one of the most wide-spread grapevine diseases worldwide. Managing the field-spread of grapevine leafroll disease requires, amongst others, stringent mealybug control. Mealybug monitoring and control methods rely on timely and accurate identification of the species present. However, proper identification of mealybug species is problematic, time-consuming and requires an expert taxonomist. In most cases, only adult females can be reliably identified morphologically. Immature insects, males and damaged specimens cannot be assigned to species. In this study, a molecular method was developed to rapidly and accurately distinguish three mealybug species associated with grapevine, namely the vine mealybug *Planococcus ficus* (Signoret), the citrus mealybug *Planococcus citri* (Risso) and the longtailed mealybug *Pseudococcus longispinus* (Targioni-Tozzetti). During the development of this identification method, a number of tasks were undertaken. Firstly, rapid and reliable DNA extraction methods were tested for mealybug DNA. Two rapid extraction methods were adapted and tested, namely the direct buffer method and the spot-PCR method. These methods reliably extracted DNA even from very small or damaged individuals, and could be performed in 15-20 minutes and three hours, respectively. Secondly, mealybug mitochondrial DNA from the cytochrome *c* oxidase subunit 1 (CO I) gene was amplified and sequenced. It was found that DNA from the 3'-end of CO I showed minimal intraspecific variation (<1%), but sufficient interspecific variation (7-12%) to clearly delineate species. This region was then used to develop three species-specific forward primers, which were used in conjunction with a common universal reverse primer. These primers were all used in a multiplex PCR to differentially amplify DNA from each of the three species. The primers were designed such that each yielded a DNA product of different length which could be separated by electrophoresis on an agarose gel. In this manner the identity of the species could be determined. The entire identification protocol (including extraction, PCR and electrophoresis) could be completed in approximately four hours. All amplified

specimens in a blind trial were correctly identified, regardless of size or condition of the specimen. The protocol is simple enough to be implemented in any molecular laboratory. This represents a considerable improvement over currently available techniques for mealybug identification, and is certain to be of great use in diagnostic identification of mealybugs in vineyards and export consignments.

CONTENTS

List of Tables

List of Figures

List of Boxes

CHAPTER ONE

A general introduction to mealybugs, grapevines, molecular techniques, detection and identification methods

Chapter One

A general introduction to mealybugs, grapevines, molecular techniques, detection and identification methods

Mealybugs

Mealybugs are small insects, comprising the family Pseudococcidae within the Hemiptera. The family consists of more than 2000 described species in 270 genera (Downie and Gullan 2004). Female mealybugs generally have three nymphal instars (Gullan 2000). The adult female is small (1-3 mm), ovoid and strongly resembles the nymphal stages. Male mealybugs undergo four instars (Gullan 2000). At the end of the second instar they cease feeding, and start spinning a waxy cocoon (Annecke and Moran 1982). Within this cocoon they complete the second and third instars, and emerge as tiny (0.5 mm) winged adults which do not feed and are short-lived (Gullan and Kosztarab 1997). Under optimum conditions, females can lay up to 240 eggs (El-Minshawy *et al.* 1974) and each life cycle takes approximately four to six weeks to complete, depending on species (Annecke and Moran 1982). Thus population levels can increase rapidly. These dense populations cause severe damage to plants, and mealybugs are a pest on a number of commercial crops. In South Africa, fifteen mealybug species have been listed as pest species (Moran 1983). For this reason, characterization and control of mealybug populations is of utmost concern in many agricultural sectors.

Impact on grapevine

On grapevine, mealybugs cause both direct and indirect crop damage. At large population levels, mealybugs can withdraw enough nutrients from the vine to severely weaken the plant. Mealybugs secrete large amounts of honeydew as a waste product. This provides a substrate on which sooty mould can grow, coating the leaves and stem of the vine and lowering the photosynthetic ability of the plant (Hattingh *et al.* 1998, Gullan *et al.* 2003). Of greater concern, however, is their ability to transmit viruses and other pathogens (Roivainen 1980). This ability to transmit viruses is particularly damaging to the grapevine industry.

Mealybugs transmit a number of grapevine viral diseases, including grapevine viruses A and B (GVA and GVB) and a number of the grapevine leafroll-associated viruses (GLRaVs). Grapevine leafroll is the most wide-spread disease of grapevine, occurring in all vine-growing areas of the world (Goheen and Cook 1959, Martelli 1986). It has been estimated that up to 62% of world-wide viral losses are due to grapevine leafroll (Krake *et al.* 1999). To date, nine serologically unrelated viruses of the family Closteroviridae (Martelli *et al.* 2002; Alkowni *et al.*, 2004) have been associated with grapevine leafroll symptoms (Engelbrecht and Kasdorf 1985, Krake *et al.* 1999, Dovas and Katis 2003). The most wide-spread of these viruses, and the one most commonly associated with leafroll symptoms is grapevine leafroll-associated virus 3 (GLRaV-3) (Cabaleiro *et al.* 1999). Leafroll viruses are phloem-based and affect the transport system of the plant (Martelli 1986). Leafroll is associated with the characteristic downward rolling of the leaf margins and discolouration of the interveinal leaf surface, while the leaf veins remain green (Martelli 1986, Krake *et al.* 1999). Symptoms become especially apparent during autumn, when infected vineyards take on a red or yellow colouration, depending on the cultivar (Krake *et al.* 1999). Grapevine leafroll lowers the photosynthetic ability of the vine, thereby causing an over-all weakening of the plant (Bertamini *et al.* 2004). Economic losses are caused by a reduction in production, delay in ripening, and reduction in pigmentation and sugar content of the grapes (Martelli 1986, Cabaleiro *et al.* 1999, Krake *et al.* 1999).

Grapevine viruses were first shown to be transmissible by mealybugs in 1983, when Rosciglione *et al.* showed that healthy vines could be infected with grapevine virus A via the mealybug vector *Pseudococcus longispinus* (Targioni-Tozzetti). Since then, a number of other mealybugs and soft scale insects (Hemiptera: Pseudococcidae and Coccidae) have been shown to transmit grapevine viruses. These are listed in detail in Table 1.1. Of particular concern in South Africa are three mealybug species, namely the vine mealybug *Planococcus ficus* (Signoret), the citrus mealybug *Planococcus citri* (Risso) and the longtailed mealybug *Pseudococcus longispinus*.

In South Africa GLRaV-3 has been shown to be transmissible by the mealybugs *Planococcus ficus* (Engelbrecht and Kasdorf 1990) and *Pseudococcus longispinus* (D. Saccaggi and K. Krüger, unpubl.). Both species are known to occur on grapevine, although in South Africa *P. ficus* is the major pest. *Planococcus citri* is of

concern because of its historical association with grapevine in South Africa (Annecke and Moran 1982). In the past, *P. citri* was reported extensively in vineyards. However, in 1975 a survey conducted in the Western Cape province concluded that the mealybug present in vineyards was in reality *P. ficus* that had been wrongly identified. They concluded that *P. citri* was "in fact rare on vines" (Annecke and Moran 1982). *Planococcus citri* has not been reported from South African vineyards again, although it occurs on vines in other areas of the world (e.g. Golino *et al.* 2002). This leaves the question as to whether *P. citri* does occur on South African vines or not. This question is made even more problematic by the difficulty in distinguishing *P. ficus* from *P. citri* morphologically (De Lotto 1975). The two species are highly similar, as will be discussed below, and mis-identifications are often made. In the scientific literature, there are a number of cases where *P. ficus* has been mis-identified as *P. citri*. For instance, Cabaleiro and Segura (1997a and 1997b) reported on GLRaV-3 transmission by a mealybug they called *P. citri*, when in fact the mealybug was later found to be *P. ficus* (Ben-Dov and German 2003). Of course, this complicates comparisons between different studies, as one is often unsure of the correct identity of the mealybug species used. These types of mis-identifications hamper scientific research and sometimes necessitate the duplication of work. Therefore, correct identification of mealybug species is of crucial importance.

Table 1.1. Mealybugs and soft scale insects implicated in the field-spread of grapevine viruses.

Family	Species	Virus transmitted [1]	Country	Reference
Pseudococcidae	*Pseudococcus longispinus*	GLRaV-3	New Zealand	Petersen and Charles 1997
		GLRaV-3	USA	Golino *et al.* 2002
		GLRaV-3	South Africa	Saccaggi and Krüger unpubl.
		GVA	Italy	Rosciglione *et al.* 1983
		GVA	Italy	La Notte *et al.* 1997b
	Planococcus ficus	GLRaV-3	South Africa	Engelbrecht and Kasdorf 1990
		GLRaV-3	Spain	Cabaleiro and Segura [2] 1997a
				Cabaleiro and Segura [2] 1997b
		GLRaV-3	USA	Golino *et al.* 2002
		GVA	South Africa	Engelbrecht and Kasdorf 1990a
	Planococcus citri	GLRaV-3	USA	Golino *et al.* 2002
	Pseudococcus maritimus	GLRaV-3	USA	Golino *et al.* 2002

[1] GLRaV-1 and GLRaV-3: grapevine leafroll-associated virus type 1 and type 3. GVA and GVB: grapevine virus A and B.
[2] In these publications *Planococcus ficus* was mis-identified and reported as *Planococcus citri* (Ben-Dov and German 2003).

Table 1.1. (cont.) Mealybugs and soft scale insects implicated in the field-spread of grapevine viruses.

Family	Species		Virus transmitted [1]	Country	Reference	
Pseudococcidae	*Pseudococcus*	*viburni*	GLRaV-3	USA	Golino *et al.*	2002
			GVA	Italy	Garau *et al.* [3]	1995
			GVB	Italy	Garau *et al.* [3]	1995
	Pseudococcus	*calceolariae*	GLRaV-3	New Zealand	Petersen and Charles	1997
	Heliococcus	*bohemicus*	GLRaV-3	France	Sforza *et al.*	2000
					Sforza *et al.*	2003
			GLRaV-1	France	Sforza *et al.*	2000
	Phenacoccus	*aceris*	GLRaV-3	France	Sforza *et al.*	2000
					Sforza *et al.*	2003
			GLRaV-1	France	Sforza *et al.*	2000
					Sforza *et al.*	2003

[1] GLRaV-1 and GLRaV-3: grapevine leafroll-associated virus type 1 and type 3. GVA and GVB: grapevine virus A and B.
[3] In this publication *Pseudococcus viburni* was called by a junior synonym, *Pseudococcus affinis* (Ben-Dov and German 2003).

Table 1.1. (cont.) Mealybugs and soft scale insects implicated in the field-spread of grapevine viruses.

Family	Species		Virus transmitted [1]	Country	Reference	
Coccidae	*Pulvinaria*	*vitis*	GLRaV-3	Italy	Belli *et al.*	1994
	Parthenolecanium	*corni*	GLRaV-1	France	Sforza *et al.*	2000
				Italy	Sforza *et al.*	2003
	Neopulvinaria	*innumerabilis*	GLRaV-1 [4]	Italy	Sforza *et al.*	2003
	Coccus	*hesperidum*	GLRaV-3	South Africa	Krüger	unpubl.
	Coccus	*longulus*	GLRaV-3	South Africa	Krüger	unpubl.

[1] GLRaV-1 and GLRaV-3: grapevine leafroll-associated virus type 1 and type 3. GVA and GVB: grapevine virus A and B.
[4] transmitted only when associated with GVA

Mealybug identification

Mealybugs are traditionally identified by subtle morphological differences between adult females (Millar 2002, but see Gullan 2000). For field identification, undamaged young adult females, before they start reproducing, can be tentatively assigned to species. Among the three species studied here, *P. longispinus* is recognizable by the two elongated posterior waxy filaments on the adult female. *Planococcus ficus* and *P. citri*, however, are more difficult to identify. These two species are morphologically very similar, and field identification relies solely on the number of short waxy filaments around the edge of the body: 17 on each side in *P. ficus* and 18 in *P. citri* (Annecke and Moran 1982). In general, female mealybugs show a reduction in useful morphological characteristics (Downie and Gullan 2004), and cannot be accurately and reliably identified without the aid of an experienced mealybug taxonomist (Gullan and Kosztarab 1997). These identifications rely on tedious and painstaking slide-mountings of specimens and inspection of minute cuticular features under a high-definition microscope (De Lotto 1975, Millar 2002, Watson and Kubiriba 2004).

The vast majority of species descriptions and identification keys for mealybugs are based on features of the adult female (Millar 2002). Adult males and nymphs cannot be identified at present. Likewise, damaged specimens often lack identifying characteristics. Due to the small size of mealybugs, specimens become damaged very easily during collection, preservation, transportation or preparation for slide-mounting. These damaged specimens may lack the characteristics necessary for identification to species level.

Despite these difficulties, correct identification of mealybug species is essential. Due to their pest status on many crops, fast, accurate and reliable identification of the mealybug species present is of utmost importance for the implementation of control measures. For instance, pheromone traps are baited with species-specific pheromones, which are used to monitor mealybug numbers for pest control. Incorrect identification of the mealybug species in the field would lead to inefficient trapping and monitoring, and a possible mealybug outbreak. In addition, since only males are caught in the traps, the identity of the species cannot be confirmed directly. Further field surveys must be

carried out to corroborate the identity of the infesting species. When relying on morphological identification, this becomes a long, tedious and often costly process. Another area in which mealybug identification is of importance is during the export of fresh produce. Mealybugs which are encountered on fruit due for export must be identified before the fruit can be shipped. This process requires that any nymphs found are reared, and adult females are collected and sent for professional identification. Until such time as an identification can be made, the fruit is held in cold storage at a large cost to both the producer and the shipper. Rearing of nymphs takes four to six weeks, and then an identification can only be made if a female mealybug is reared. If the nymphs die, only adult males emerge, or the specimens are damaged, no identification can be made and the consignment must be destroyed or sold locally at a reduced rate.

For these reasons, it has become necessary to find an alternative method of mealybug identification that is rapid as well as accurate, and can extend to the identification of damaged, immature and male specimens. This method needs to be reliable and simple enough to be performed in any laboratory, with an accuracy comparable to that of an experienced taxonomist.

A possible species identification method that fulfils all the above criteria is that of molecular identification. This approach utilises genetic variation between species to identify diagnostic nucleotide "characters" to differentiate species. This involves finding nucleotide differences that are stable between species, but do not vary significantly within a species. A molecular identification approach has been successfully applied in a number of species groups, especially those for which morphological identification is time-consuming or problematic (Besansky *et al.* 2003). However, the initial time and costs involved in the development of such a technique are often prohibitive, and therefore are not typically undertaken for most species. Those insect species for which molecular identification methods have been developed are usually of medical (Marcilla *et al.* 2001, Koekemoer *et al.* 2002a and 2002b, Weeto *et al.* 2004), forensic (Harvey *et al.* 2003a and 2003b, Zehner *et al.* 2004) or agricultural (Beuning *et al.* 1999, Otranto and Stevens 2002, Gariepy *et al.* 2005) importance. These will be dealt with in more detail below when molecular identification systems are discussed.

Basic DNA-manipulation techniques

Development of a DNA-based identification technique requires finding stable nucleotide differences between species. This in turn requires the availability of DNA sequence data. However, as is often the case with insects, not enough molecular data are currently available for mealybugs to be able to design a reliable DNA-based identification technique. Therefore, in this study DNA sequence data must be obtained for each of the three mealybug species in order to form a basis for investigation. To do this, some basic DNA manipulation techniques are employed, which will be briefly described and discussed below.

DNA extraction

In order to use DNA for analyses, it must be separated from the surrounding body and cell contents. Extracting DNA from a sample usually involves digestion of proteins and lysis of the cell membranes in a buffered solution (Flavell and Barr 1990). The buffered solution stabilizes the DNA helix, retaining the quality of DNA for further analyses. This extraction step is often followed by purification of the DNA. Purified DNA remains stable for longer periods, allowing it to be stored for long-term use. However, purification may also increase loss or damage of DNA. Small amounts of DNA can be lost at each transfer or washing step, and large DNA molecules may break up during the purification process. Therefore, some extraction methods omit the purification procedure in the interest of larger DNA yields and simplicity of the technique, but at a cost to long-term quality and storage of the DNA.

Many different DNA extraction protocols have been developed. This wide range of both commercial and in-house extraction methods emphasises the importance of suitable DNA extraction techniques to each study. Protocols vary widely, and many methods are specific to the type of organism or tissue used (e.g. Mercier *et al.* 1990, Löffler *et al.* 1997, Csaikl *et al.* 1998). Some extraction methods are optimised for simplicity and speed, while others are adapted for qualitative or quantitative DNA extraction. Each technique has applicability in different studies and to different organisms. Therefore it

is of importance to review, test and adapt extraction methods to uniquely suit the study at hand.

PCR amplification

The polymerase chain reaction (PCR) was developed by R. K. Saiki, K. B. Mullis and others in the late 1980s (Saiki *et al.* 1985, Mullis and Faloona 1987, Saiki *et al.* 1988) and has radically changed molecular research and diagnostics (Kocher and Wilson 1991, Caterino *et al.* 2000). The PCR involves the *in vitro* synthesis of large amounts of DNA copies from a single starting molecule (Kocher and Wilson 1991). PCR uses short, single strands of DNA (usually 18-30 nucleotides), called oligomers or primers, to select a DNA region of specific interest. Once the primers are annealed to the DNA, thermostable *Taq* polymerase (first purified from the thermophilic organism *Thermus aquaticus*) builds a complementary strand extending from the primer by incorporating free deoxynucleoside triphosphate (dNTP; base + deoxyribose sugar + phosphate) molecules in the reaction mix. Two primers that anneal on complementary strands are used, with *Taq* extending the region between them. The reaction mixture is cycled between different temperature optima for the different stages of reaction of denaturation, annealing and elongation. This process is repeated a number of times and the DNA thus produced increases exponentially. This large DNA product can then be used in any further analyses. The ability to replicate DNA simply and efficiently by PCR has contributed greatly to the ease with which DNA data can be utilised, and many biological studies now include a molecular component (Bermingham and Luettich 2003).

DNA sequencing

In order to obtain data on which a subsequent molecular technique can be based, the nucleotide sequence of the DNA is needed. This process is termed sequencing. Although sequencing methods are now automated, the underlying principles for sequencing have remained unchanged since Sanger *et al.* (1977) developed a method for sequencing using chain-terminating inhibitors. In this method, a dNTP is used which contains a functional group which inhibits further reaction with other dNTPs. The most

common inhibitors used are the 2',3'-dideoxynucleoside triphosphates (ddNTPs). These lack a 3'-hyroxyl group on the deoxyribose sugar, which has the effect of terminating the growing chain between the sugar and phosphate groups when it is incorporated in place of a normal dNTP (Howe and Ward 1991). If a mixture of normal and inhibiting dNTPs is used in a PCR reaction, a number of DNA chains will be produced, each of a different length, each ending with an inhibiting ddNTP. In the original paper by Sanger *et al.* (1977), four separate reactions were run, one for each dNTP (i.e for deoxyadenine, -thymine, -cytocine and -guanine triphosphate). Each reaction contained an inhibiting ddNTP. The DNA fragments formed by the chain-termination method were separated by electrophoresis on a gel, revealing the different sized products. The DNA sequence could then be "read" from bottom to top. This method was first combined with initial DNA amplification via PCR by Wrischnik *et al.* (1987), and has since remained essentially unchanged, except that all dideoxy-reactions are now performed together in a single reaction mix. Nucleotide-specific fluorescent markers are attached to each ddNTP, such that each of the four nucleotides fluoresces with a different light (Howe and Ward 1991). This enables a single reaction to be performed for all four nucleotides. The resulting DNA fragments are then separated by capillary electrophoresis, and the sequence "read" by the identifying the fluorescent markers.

DNA-based identification methods

DNA-based identification methods vary widely. Many methods were developed before the large-scale use of PCR or automated sequencing methods. Currently, most DNA-based identification techniques include the use of a PCR, either in initial amplification of the DNA, or as a diagnostic tool in itself. Ultimately, all DNA-based identification techniques rely on finding stable nucleotide differences between species which are used as diagnostic markers to separate species.

DNA fragment mapping

Restriction enzymes, restriction fragment length polymorphism (RFLP) (e.g. Marfurt *et al.* 2003), random amplified polymorphic DNA (RAPD) (Williams *et al.*

1990) and amplified fragment length polymorphisms (AFLP) (Vos *et al.* 1995) have been used extensively to analyse genetic variation (Caterino *et al.* 2000). Some of these techniques are still in use today. These techniques all rely on cleavage or amplification of fragments of DNA at specific short nucleotide sequences, sometimes in a PCR reaction. This produces fragments of different lengths, each ending or starting with the same short sequence. Nucleotide changes which result in a change of restriction or primer site will produce a different number of fragments or fragments of different lengths. These fragments are then separated, usually by electrophoresis, and the resulting pattern interpreted. Often, the pattern thus generated shows clear differences between species and can then be used as a diagnostic tool (e.g. Malgorn and Coquoz 1999).

There are two main advantages to these techniques. Firstly, the nucleotide sequence of the DNA strand does not need to be known beforehand. This decreases the amount of time and money involved in the initial design of the protocol. Secondly, they detect sequence variation over large regions of DNA, often across the whole genome (Caterino *et al.* 2000). The likelihood of detecting variation is therefore very high. However, there are a number of disadvantages in the use of these techniques, especially when applied to species identification. They work over very large regions of DNA, and although this results in high detection ability, it also yields high uncertainty. One cannot know where in the genome the variability arises, nor if the variation is indeed stable between species and not simply an artifact of intraspecific variation. Variation that does not result in a change in restriction or primer site will also go undetected (Gasser and Chilton 2001). When applied to large-scale diagnostic use in different laboratories, the reproducibility of these techniques is low. Small changes in the reaction conditions can dramatically alter the results (see Stevens and Wall 1995, Benecke 1998). This can sometimes entirely obliterate a pattern that was supposed to be diagnostic for that species (Kengne *et al.* 2001).

Although DNA fragment mapping has sometimes been used for direct species identifications (e.g. Marfurt *et al.* 2003), its reliability for this purpose is usually low.

DNA conformation polymorphism

A more recent technique that has been used is double- or single-strand conformation polymorphism (DSCP or SSCP). Conformation polymorphism relies on the differing mobilities of different sizes and structures (i.e. conformations) of DNA fragments (Ren 2000, Gasser and Chilton 2001). Different conformations are encountered in double-stranded DNA (dsDNA) when nucleotide mutations cause a change in the curvature of the double helix (Atkinson and Adams 1997). SSCP relies on denaturation of the dsDNA molecule (separation into single-stranded DNA – ssDNA). The denatured DNA is then kept in an environment where the two DNA stands cannot re-anneal. In this situation, the ssDNA twists in on itself, forming secondary and tertiary structures through internal hydrogen bonding. Nucleotide differences between ssDNA fragments can lead to different conformations (Ren 2000, Anderson *et al.* 2003). Mobilities of the differing dsDNA or ssDNA conformations can be compared by separation by electrophoresis, enabling detection of nucleotide variation (de Albuquerque and Costa 2003).

DSCP is particularly sensitive when dealing with AT-rich genomes, as long stretches of As or Ts alter the curvature of the DNA helix (Atkinson and Adams 1997). SSCP is very sensitive, and in some cases even a single nucleotide difference can lead to a different conformation, and therefore a different mobility (Sentinelli *et al.* 2000, de Souza Menezes *et al.* 2003). This depends to a large extent on the reaction conditions, and many different conditions may have to be tested before a high sensitivity is reached (e.g. Markino *et al.* 2000, Andersen *et al.* 2003). However, SSCP also has a degree of uncertainty. The folding characteristics of ssDNA are poorly understood at present, and the conformation of an ssDNA molecule cannot be predicted from its nucleotide sequence (Bettinaglio *et al.* 2002). Therefore, one cannot be sure that a genetic change will be reflected in the mobility of ssDNA. SSCP is also time-consuming and requires specialized equipment and expertise (Sentinelli *et al.* 2000).

Although SSCP has been used for accurate species identification (e.g. Kong *et al.* 2003), it is more suited to large-scale screening of samples for possible nucleotide variation (Caterino *et al.* 2000, e.g. Salas *et al.* 2001, de Gruijter *et al.* 2002).

Phylogenetic analysis

Phylogenetic analysis is widely used to elucidate relationships between individuals, using cogent sequence data. Usually, this relationship data is used in evolutionary and phylogeographic studies (Sperling 1999). More seldom, it is used to identify unknown specimens.

In a phylogenetic analysis, data from related individuals will align most closely to each other. When both known and unknown specimens are used, the relationships of the unknown specimens relative to the known individuals can be found. If these are of the same species, the unknown specimens will align most closely with con-specific individuals, and thus a species identification can be made. However, if these are only of related species they will still align to those most closely related, but an exact species identification cannot be made.

Species identification by means of phylogenetic analysis requires a database of sequence data from known specimens. Sometimes this data can be obtained from previous publications, and are often stored in a central database such as Genbank. However, if the DNA sequence data is not available (as is often the case in insects), reference sequences must be obtained by DNA sequencing. As with other DNA-based identification systems, this initial time and cost may be prohibitive, and is commonly only undertaken for insects of economic importance (e.g. Zehner *et al.* 2004). In addition, DNA sequence data from the specimen must be obtained and a phylogeny computed. This can be a lengthy process (Wells and Sperling 2001).

Species identification by phylogenetic analysis can be extremely accurate, provided the initial database is large enough to cover all unknown species, or the unknown species can be identified to a level where they will be covered by the analysis (Wells and Sperling 2001). The accuracy of this identification technique makes it ideal for use in legal forensic diagnostics, where the identification may undergo close legal scrutiny. It has been used to this purpose, most notably to identify flies (Diptera) of forensic importance during legal proceedings (Wells and Sperling 2001, Harvey *et al.* 2003a, 2003b, Zehner *et al.* 2004) However, the time involved in making this robust identification is not well suited to large-scale diagnostic identifications, where high-throughput of samples is of paramount importance.

Diagnostic PCR

The polymerase chain reaction is used in almost all genetic studies, usually to generate enough DNA product to be used in further analyses – such as cloning, sequencing and DNA fingerprinting (Bermingham and Luettich 2003). More recently, however, it has been used as a tool in itself.

Diagnostic PCR for identification relies on the design of species-specific primers. Under optimal conditions, these primers will anneal to the species in question, but not to any other. How these primers are used in the PCR varies, and this determines the type of diagnostic PCR.

Species-specific PCR involves the use of species-specific primers that are designed in pairs and used separately. Each species-specific primer pair will identify one species of the group, and each primer pair is used in a separate reaction. Since separate PCRs are used, this identification process can become very long and tedious, especially for large species groups. The design of primers is also sometimes problematic due to the high specificity requirements, and often a number of primers need to be tested before a suitable combination is found. Despite these difficulties, species-specific PCR is accurate and is commonly used as an identification technique.

Species-specific PCR has been used in insect groups to identify species. An example of a species-specific PCR applicable to this study is that developed by Beuning *et al.* (1999). Species-specific primers were developed to identify four New Zealand mealybug species. The four species could be separated by PCR-amplifying each sample with four separate sets of primers, each in a separate reaction. Each primer set only amplified a single species, and the species could then be identified.

Multiplex PCR involves using a number of specific primers together in the same reaction mix. Depending on the purpose of the multiplex PCR, two different types of primers can be designed. Firstly, primers may be designed to amplify any or all specimens present in a sample without differentiating them. This is often used in large-scale screening of clinical samples. If samples are found positive, further tests are performed to determine the identity of the species present. Secondly, primers may be designed to differentiate the species involved, either within the same sample or in different samples. This is usually accomplished by designing the primers to yield DNA

products of different lengths, which can then be separated by electrophoresis. In both types of multiplex reaction, primers are not necessarily designed in pairs. Commonly, species-specific forward (or reverse) primers are designed and used with a single common reverse (or forward) primer. This decreases the number of chemical compounds present in the reaction mix, and therefore decreases non-specific activity. Although multiplex PCRs are clearly faster than conventional species-specific PCRs, they are difficult to design and optimize. In particular, the primers are difficult to design, since they must have very similar thermodynamic and kinetic properties, yet not interact with each other. If primers are to differentiate species, another constraint on primer design is that the primers must be separated from each other by a minimum number of nucleotides in the original DNA sequence. Despite these difficulties, if a multiplex PCR can be designed and optimized properly, it is the fastest method of DNA-based species identification currently available.

Although multiplex PCR is used routinely to identify some parasitic (e.g. Chansiri *et al.* 2001, Kho *et al.* 2003) and bacterial (e.g. Yoon *et al.* 1996, Liu *et al.* 2003, Kwon *et al.* 2004) infections in clinical samples, its use in insect identification has only recently been realized (Kengne *et al.* 2001, Koekemoer *et al.* 2002b, Gariepy *et al.* 2005). The study by Koekemoer *et al.* (2002b) and the subsequent testing of the method by Weeto *et al.* (2004) provide an excellent example of the potential of multiplex PCR for identifying species during large-scale diagnostic work. During the study a universal forward primer and species-specific reverse primers were designed. These were all placed in the same reaction mix, along with extracted DNA from a single mosquito. Each combination of forward primer and species-specific reverse primer yielded amplified DNA of a different length. These amplified fragments could then be separated on an agarose gel for direct identification of the species. In this way they were able to separate all six species of the *Anopheles funestus* Giles group (Diptera: Culicidae) with 100% accuracy in those cases in which the test could be completed (868 of 900 tests were completed; DNA from the other 32 specimens could not be amplified at all) (Weeto *et al.* 2004).

Choice of genetic region

In any molecular study, careful choice of a genetic region is of crucial importance. This determines the procedures used in DNA isolation, amplification, sequencing and manipulation, as well as the usefulness of the information obtained (Brower and DeSalle 1994, Hwang and Kim 1999, Rokas *et al.* 2002). In this study, mitochondrial DNA of the cytochrome *c* oxidase subunit 1 (CO I) gene was used in analyses. This choice will be discussed below.

Properties of mitochondrial DNA

A number of properties make mitochondrial DNA (mtDNA) ideal as a molecular marker, and mtDNA has been used in many genetic studies (Zhang and Hewitt 1997). However, as with any genetic region, there are also some properties of mtDNA that make it difficult to use. These drawbacks should also be kept in mind when working with mtDNA.

Mitochondrial DNA is a small circular molecule, varying in length from about 14 000 to 17 000 base pairs, which is far smaller than the genomic chromosomes (Hwang and Kim 1999). Mitochondria also occur in large numbers inside a cell. The small molecule is less likely to be torn or become damaged during extraction or subsequent manipulation of the DNA. Even if some copies are damaged, the large number of mitochondria present in each cell ensures that many undamaged copies will still be extracted. This is especially useful when working with small, degraded or precious specimens. The ease of extraction and manipulation make mtDNA genes a popular choice for many studies (Brower and De Salle 1994).

The mitochondrial DNA molecule is extremely compact and contains only one non-coding region: the control loop (called the D-loop in mammals and the AT-rich region in insects) (Zhang and Hewitt 1997). Outside the control loop, very few non-coding nucleotides are found (Avise 1991). For instance, Clary and Wolstenholme (1985) found a maximum of 31 non-coding nucleotides between the mitochondrial coding genes in *Drosophila yakuba* Burla (Diptera: Drosophilidae). Since the size and position of coding regions are well documented (Hwang and Kim 1999), universal or conserved

primers can be developed for DNA amplification in diverse organisms (Kocher *et al.* 1989).

Mitochondria generally undergo no recombination and are maternally inherited (Birky *et al.* 1983, but see Zhang and Hewitt 1997, Wallis 1999). This makes mtDNA ideal for determining simple patterns of relatedness. The fact that mitochondria undergo no recombination lowers the complexity of analytical tools needed to interpret molecular data.

When considering the compactness and economy of the mtDNA molecule, it might be assumed that mitochondrial nucleotide sequence is highly conserved (Avise 1991). Surprisingly, just the opposite is usually found, and it has been shown that animal mtDNA evolves at a faster rate than nuclear DNA (nDNA) (Brown *et al.* 1979, Avise 1991, Lin and Danforth 2004). Many fascinating theories exist as to why this is so. However, they are unfortunately beyond the scope of this study, and will not be discussed here (see Avise 1991 and refs therein). The fast evolutionary rate of mtDNA makes it well suited to studies at lower taxonomic levels. However, the relative rate of mtDNA mutation in insects has been questioned (Zhang and Hewitt 1997). Where some studies have found faster evolutionary rates in insect mitochondrial genes (e.g Rokas *et al.* 2002), others have not (Powell 1986). It is unsure whether insect mtDNA does indeed evolve faster than nDNA (Zhang and Hewitt 1997).

Arguably the most important characteristic to bear in mind when working with insect mtDNA is its tendency to have a large majority of the bases adenine and thymine (A and T) and a corresponding reduction in the number of guanine and cytocine (G and C) bases. Most nuclear genes have about an equal amount of A and T to G and C. However, mtDNA usually has at least a 55% AT-bias (Simon *et al.* 1994). This bias is extremely marked in insect mtDNA, where *Drosophila yakuba* has a 78.6% AT-bias (Clary and Woltenholme 1985) and the honeybee *Apis mellifera ligustica* Spinola (Hymenoptera: Apidae) has an extreme 84.9% AT-bias (Crozier and Crozier 1993), the highest AT-bias found to date. The implications of this strong AT-bias are numerous. Firstly, transversions (A↔T) are expected to dominate over transitions (G↔A or C↔T) (Crozier and Crozier 1993). This must be taken into account in phylogenetic studies, where most models assume a transition bias in mutations. Having a predominance of

only two nucleotides also means that loci become saturated very quickly and there is a large chance of multiple substitutions at a single locus going unnoticed (Satta *et al.* 1987, Rokas *et al.* 2002). Substitutions are always easier to detect when the four bases occur in approximately equal frequencies (Marcilla *et al.* 2001). A strong AT-bias also makes many molecular techniques and analyses more difficult (e.g. Mooers and Holmes 2000). For instance, design of proper primers for PCR, in which 40-60% GC content is the ideal, is very difficult for a strongly AT-biased genome.

Another complication to be kept in mind when working with mtDNA is the possible existence of pseudogenes (Caterino *et al.* 2000). These are segments of mtDNA that have become inserted into the nuclear genome during evolutionary history. Here they no longer function as coding genes, and are therefore free to accumulate a large number of mutations. When amplifying mtDNA, it is possible to co-amplify or preferentially amplify a pseudogene. This would show a far higher and more random mutation rate than the mitochondrial gene, and therefore confound analyses. Unfortunately, there is very little way of being absolutely sure of the gene one is amplifying, although a number of indicators can suggest the presence of a pseudogene (Zhang and Hewitt 1996). These are given in Box 1.1. Pseudogenes have been reported in some insect groups, including aphids (Sunnucks and Hales 1996) and Orthopterans (Vaughan *et al.* 1999). At present, it is unknown how common this phenomenon is in insects.

Cytochrome *c* oxidase subunit 1

Cytochrome *c* oxidase subunit 1 (CO I) is a mitochondrial gene coding for a protein involved in the metabolic pathway of the cell. As such, it is extremely important and therefore conserved across taxa. Its amino acid sequence evolves slowly, and this has been used to separate species groups at higher taxonomic levels (Hwang and Kim 1999). However, due to the degeneracy in the mitochondrial DNA code, the nucleotide content of CO I evolves fairly rapidly (without changing the amino acid sequence), and is often used to separate species or even populations (e.g. Brown *et al.* 1997, Trewick 2001). It has even been suggested that the CO I gene be used to develop genetic "barcodes" for all organisms, and that these be used in all identifications (Hebert *et al.* 2003a and 2003b). The structure and organization of CO I and the genes surrounding it are highly

conserved. This makes it easy to design conserved PCR primers for the region. Simon *et al.* (1994) listed a total of 14 conserved primers for CO I and the flanking genes. This availability of universal primers makes CO I simple to amplify and sequence with minimal initial effort. All these reasons make CO I an ideal choice when designing a study at the species level, and when working with an organism that has not previously been characterized molecularly.

Box 1.1. Nuclear copies of mitochondrial genes

(adapted from Zhang and Hewitt 1996)

There is a possibility that a pseudogene is being amplified if the study encounters the following anomalies:

1. More than one band, or different bands, are constantly produced during PCR amplification.
2. Background peaks or sequence ambiguities are constantly found when sequencing.
3. The DNA sequence contains data which will unexpectedly change the polymerase translation of the sequence, such as unusual frameshifts, insertions/deletions or stop codons.
4. The DNA sequence is particularly more divergent than expected.
5. Phylogenetic analysis results in unusual, unexplained or contradictory tree topology.

Objectives of the study

This study forms part of a large project dealing with the vectors of grapevine leafroll in South Africa. As part of that project, this study examined three mealybug species associated with grapevine in South Africa, namely the vine mealybug *Planococcus ficus*, the citrus mealybug *Planococcus citri* and the longtailed mealybug *Pseudococcus longispinus*. A molecular technique was developed to identify these species. This technique had an accuracy comparable to that of an experienced mealybug taxonomist for adult females, but exceeded that in its sensitivity and reliability in identifying immature, damaged and male specimens. The technique was rapid, yet simple enough to be implemented in any molecular laboratory as part of a pre-existing monitoring and control strategy.

Therefore, the main objective of this study was to develop a DNA-based identification technique for the three mealybug species that is rapid, accurate, sensitive, reliable and simple. In order to accomplish this objective, the following tasks were undertaken:

a) Tested and adapted DNA extraction methods for mealybug which were rapid and reliable and could be implemented during large-scale diagnostic work.

b) Developed and optimized PCRs for the amplification of mealybug DNA from the mitochondrial CO I region.

c) Obtained DNA sequence data from the mealybug mitochondrial CO I region.

d) Compared intraspecific and interspecific variation in the mealybug mitochondrial CO I region.

e) Found a region within the CO I gene showing minimal intraspecific variation, but enough interspecific variation to delineate the three mealybug species.

f) Developed a suitable DNA-based diagnostic technique for identification of the three mealybug species which fulfiled the criteria listed above.

g) Tested this DNA-based identification technique for reliability to correctly identify the three mealybug species in all cases.

Development of this technique resulted in more rapid identification of mealybugs which extended to immature, damaged and male specimens. This can be put into use to improve and streamline mealybug control strategies in the South African wine industry.

CHAPTER TWO

Rapid DNA extraction from small insects for PCR analyses

Chapter Two

Rapid DNA extraction from small insects for PCR analyses

Abstract

As the number of molecular studies increases, the need for reliable extraction techniques will continue to grow. Molecular diagnostic procedures, in particular, require swift and reliable extraction protocols. In order to optimise these procedures, rapid and efficient DNA extraction protocols are needed. In this study, two rapid DNA extraction techniques (direct buffer and spot-PCR) are presented, adapted and tested in comparison to two well-known traditional extraction methods (salting out and phenol/chloroform) for DNA from small insects, using mealybugs (Hemiptera: Pseudococcidae) as an example. The two rapid techniques decrease processing time and increase reliability when compared to the traditional extraction methods. The protocols can be completed in 15 min and 3 h, respectively, and DNA is extracted very efficiently (up to 100% of small specimens tested, compared to 78% and 8% for the salting out and phenol/chloroform methods). These methods are simple and can easily be implemented as part of any pre-existing molecular diagnostic procedure.

Introduction

With the recent surge in development of various molecular techniques, sequence data are becoming more important in many branches of biology. The quality of this type of data is dependent on the quality of the molecular method used in analyses, which in turn is dependent on the integrity of the DNA or RNA obtained from a sample. The quality of the DNA of a sample is dependent on its age, preservation and the method of DNA extraction used. The DNA extraction method employed impacts on the quality and quantity of DNA available for use.

Until recently, DNA extraction protocols were developed with the objective of obtaining high-quality, large molecular weight, purified DNA (Sambrook *et al.* 1989). This DNA would then be stored and used repeatedly in long-term research projects. The length of time taken to complete the protocol was irrelevant when compared to the length of time for which extracted DNA was stored. These types of DNA extraction protocols are still in common use in many molecular studies.

However, recently a number of molecular diagnostic tools have been developed, particularly for medically and economically important organisms (e.g. Yoon *et al.* 1996, Chansiri *et al.* 2001). These largely include diagnostic screening and identification tools for clinical samples. In the field of entomology, DNA-based identifications have only been used for a few organisms. Those insects for which molecular identification methods have been developed are usually of medical (Marcilla *et al.* 2001, Koekemoer *et al.* 2002a, 2002b, Weeto *et al.* 2004), forensic (Harvey *et al.* 2003a, 2003b, Zehner *et al.* 2004) or agricultural (Beuning *et al.* 1999, Otranto and Stevens 2002, Gariepy *et al.* 2005) importance. Rapid identifications are important for speedy decision-making with regard to legal proceedings, medical treatments or pest and vector control. However, despite the development of streamlined molecular diagnostic techniques, development of DNA extraction methods has not kept pace. Many diagnostic studies still employ long and tedious extraction protocols, which clearly decrease the efficiency of the diagnostic technique. With a continually increasing demand for prompt identifications, as well as the recent suggestion that all biological identifications should include a molecular component (Hebert *et al.* 2003a, 2003b), fast, efficient and sensitive DNA extraction protocols are essential.

In this study, two rapid DNA extraction techniques were tested for speed, reliability and sensitivity in extracting DNA from small specimens, using a number of different buffers. Both methods involve mechanical rupturing of the insect body followed by a single, short boiling step in a buffered solution. The first method is a direct buffer boiling method employed by Koekemoer *et al.* (2002b), originally using STE buffer. The second is a spotting method previously used only for viral RNA extraction from plant samples (La Notte *et al.* 1997). During the course of this study, we tested the same

technique for its ability to extract insect DNA. The two extraction techniques were then compared to two traditional in-house extraction techniques often used for insects.

As a case study, the techniques were tested for extraction of mitochondrial DNA from mealybugs (Hemiptera: Pseudococcidae). Mealybugs are very small insects (nymphs and males: 0.5 mm, adult females: 1-3 mm) which are commercial and phytosanitary pests on a number of agricultural crops. During molecular studies, traditional DNA extraction methods are often not able to extract sufficient DNA from mealybugs for subsequent analyses, yet further molecular investigation is needed on this economically important pest. For these reasons, we tested, adapted and compared rapid DNA extraction techniques for reliability and sensitivity when extracting DNA from small specimens.

Materials and Methods

Insects

Three mealybug species were used in this study, namely the vine mealybug *Planococcus ficus* (Signoret), the citrus mealybug *Planococcus citri* (Risso) and the longtailed mealybug *Pseudococcus longispinus* (Targioni-Tozzetti) (105, 64 and 57 individuals, respectively). All species were used for the ten extraction methods tested. *Planococcus ficus* was obtained from field-collected samples from vineyards in the Western Cape Province, South Africa and from a laboratory colony maintained at the University of Pretoria, South Africa. *Planococcus citri* and *P. longispinus* were obtained from laboratory colonies maintained at the Agricultural Research Council – Plant Protection Research Institute (ARC-PPRI) and the University of Pretoria, respectively. Once collected, mealybugs were stored in absolute ethanol at ambient temperature and transferred to -20ºC upon returning to the laboratory (Post *et al.* 1993, Fukatsu 1999, Tayutivukul *et al.* 2003).

The direct buffer method

A single whole insect was crushed in 50-200 µL TNES (50 mM Tris-HCl, pH 7.5, 0.4 M NaCl, 20 mM EDTA, 0.5% SDS), STE (0.1 M NaCl, 10 mM Tris, pH 8.6,

1 mM EDTA), GES (0.1 M glycine, pH 9, 50 mM NaCl, 1 mM EDTA, 1% β-mercaptoethanol, 0.5% Triton X-100) or CTAB (100 mM Tris-HCl, pH 8, 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 0.2% β-mercaptoethanol) buffer. The sample was incubated at 94ºC for 12 min and cell debris precipitated by spinning at 13 000 rpm for 1 min. Extracted DNA was stored at -20ºC.

The spot-PCR method

A single whole insect was crushed on a positively charged nylon membrane soaked in a 50 mM NaOH and 2.5 mM EDTA solution, then allowed to dry. A small portion (ca. 3 mm^2) of the spotted membrane was cut out and placed in 10-50 µL TNES, STE, GES, or CTAB buffer (described above). The sample was then incubated at 95ºC for 10 min and cooled on ice. Extracted DNA was stored at -20ºC.

The phenol/chloroform method

DNA from a single whole insect was extracted using a modification of a general procedure for extraction with phenol (Sambrook *et al.* 1989, Sambrook and Russell 2001). The insect was crushed and incubated at 40ºC in 0.6 mg/mL Proteinase K and 300 µL TNES buffer for 4-18 h. DNA was then purified by washing with organic solvents: once with a chloroform : isoamly mix (24:1 v/v); once with a chloroform : phenol mix (1:1 v/v); and once with chloroform only. DNA was then precipitated with absolute ethanol. Extracted DNA was stored at -20ºC.

The salting out method

DNA from a single whole insect was extracted using the protocol of Sunnucks and Hales (1996) with minor adjustments, including the following: the insect was incubated at 40ºC in 0.6 mg/mL Proteinase K and TNES buffer; and samples were left for at least 1 h at -20ºC during precipitation of the DNA with absolute ethanol. Extracted DNA was stored at -20ºC.

PCR conditions

Performance of the extraction methods was tested by PCR amplification of the CO I gene with conserved primers C1-J-2183 (alias Jerry) and TL2-N-3014 (alias Pat) (Simon *et al.* 1994). Products were used in the following 25 µL reaction mix: NH_4 Reaction Buffer (final concentration: 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl, 0.01% Tween-20), 4 mM $MgCl_2$, 50 µM each dNTP, 0.4 µM each primer, 0.05 units/µL BIOTAQ™ DNA Polymerase (BioLine GmbH, Luckenwalde, Germany), and 1-2 µL extracted DNA.

PCR thermal cycling conditions for consisted of initial denaturation at 94ºC for 1 min, then 35 cycles with denaturation at 94ºC for 45 sec, annealing at 48ºC for 45 sec and extension at 72ºC for 1 min, with a final extension at 72ºC for 3 min. Samples were cooled to at least 10ºC, removed and stored at 4ºC until needed.

Results

The direct buffer method

The direct buffer method was extremely fast, taking only 15-20 min to complete. DNA extracted using the STE and GES buffers amplified well in PCR. These buffers extracted DNA from 96% (48 of 50) and 89% (8 of 9) of samples tested, respectively (Table 2.1), even at decreased buffer volumes. The TNES and CTAB buffers failed to extract DNA or inhibited the PCR reaction. To test this, amplified DNA from a previous successful PCR was added to the reaction mix. Under these circumstances, reaction mixes containing TNES buffer amplified very weakly, whereas those containing CTAB did not amplify at all. Extracted DNA from the STE and GES buffers remained stable for 4-6 weeks, after which degradation of the DNA could be seen in the PCR amplified product by smearing and many smaller bands.

The spot-PCR method

The spot-PCR method could be completed in approximately three hours, depending on how long the membrane was left to dry. The GES, TNES and STE buffers worked well at higher buffer volumes (30-50 µL) with larger insects. GES and STE buffer extracted DNA successfully in 86% (12 of 14) and 100% (9 of 9) of cases, respectively.

At lower buffer volumes (10-20 µL) or when using very small or parasitized mealybugs, TNES buffer failed to extract DNA (Table 1), leading to an overall success rate of only 20%. The CTAB buffer failed to extract any DNA or inhibited the PCR reaction. Adding previously amplified DNA did not result in an amplification product. Extracted DNA could only be stored for approximately two weeks, after which degradation of the DNA was noticed.

Table 2.1. Comparison of methods for extraction of DNA from mealybugs.

Method	**Buffer**	**Success [1] (sample size)**	**Processing time**	**DNA storage period**
Salting out	TNES	78% (39 of 50)	2 days	10 mo
Phenol/chloroform	TNES	8% (1 of 12)	2 days	--
Direct buffer	STE	94% (48 of 50)	20 min	4-6 wk
	GES	89% (8 of 9)	20 min	4-6 wk
	TNES	0% (0 of 50)	20 min	--
	CTAB	0% (0 of 10)	20 min	--
Spot-PCR	GES	86% (12 of 14)	3 h	2 wk
	TNES	20% (2 of 10)	3 h	2 wk
	STE	100% (9 of 9)	3 h	2 wk
	CTAB	0% (0 of 12)	3 h	--

[1] Success of the extraction was determined by PCR.

The phenol/chloroform method

Using this method, DNA was extracted and amplified in only one case (Table 2.1). This product did not amplify consistently in the PCR.

The salting out method

The salting out method took a minimum of about eight hours and was usually conducted over a period of two days. The extractions were left overnight at either the incubation (40ºC) or cooling (-20ºC) steps. The salting out method worked successfully in 78% of cases (44 of 51; Table 2.1). The specimens from which DNA could not be extracted were usually small first or second instar nymphs. Extracted DNA could be stored for up to 10 months with minimal degradation of the DNA.

Discussion

In this study, two rapid, largely unknown protocols for DNA extraction from insects were adapted and tested. These were then compared to two extraction techniques which are widely used for insect DNA extraction.

When compared to the more traditional extraction methods, the rapid protocols presented here performed reliably, extracting DNA from a higher percentage of specimens than either of the traditional methods. This failure of the traditional methods to extract DNA in all cases was unexpected, and especially so in the case of the phenol/chloroform purification, which is used extensively in insects (e.g. Vaughan 1999, Marcilla *et al.* 2001, Harvey *et al.* 2003a), as well as a large number of other animal groups (e.g. Kocher *et al.* 1989). The most likely reason for this is loss of DNA during the many transfer steps, exacerbated by the fact that mealybugs are very small insects in the first place (~ 0.5 mm nymphs, 1-3 mm adult females). It was noted that when these methods failed to extract DNA, the specimen used was usually a small first or second instar nymph or a parasitized individual. If loss of DNA is to account for occasional failure of extractions, then any method involving purification and/or transfer steps should be used with caution when working with very small or precious samples. DNA loss such as this would be unfortunate in diagnostic work, where identification of every

specimen, regardless of how small, is essential. Plainly, extraction methods that minimise loss of DNA are vital when developing a streamlined molecular technique for identification.

In both the direct buffer and spot-PCR extractions, GES and STE buffers performed well (86 – 100% success), while TNES and CTAB buffers did not (0 – 20% success). When PCR mixes with the latter two buffers were spiked with previously amplified DNA, the reaction amplified poorly (in the case of TNES) or not at all (in the case of CTAB). This implies some inhibition of reaction. The most likely cause of this is the high concentration of EDTA (20 mM EDTA) in these two buffers, when compared to the far lower concentration in the GES and STE buffers (1 mM EDTA). EDTA is a metal ion chelator, meaning that it removes and inactivates metal ions present in solution (Ebbing 1996). In PCR, metal ions are involved in the catalyzation of polymerization of the DNA strand by the enzyme (Steitz 1999). Thus, inactivation of the metal ions would hamper the reaction, thereby decreasing or completely inhibiting the PCR (Wiedbrauk *et al.* 1995, Abu Al-Soud and Rådström 2001). The observed failure of PCR amplification with samples containing a high EDTA concentration could be due to EDTA's chelating effect, and would explain why even spiked samples amplified poorly or not at all.

Both the direct buffer and spot-PCR extraction techniques can be completed quickly (15 minutes and three hours, respectively). The direct buffer method, in particular, lends itself to large-scale, rapid laboratory processing of samples. The spot-PCR would be ideal for field collections and small specimens, due to its reliability with three different buffers. In the original publication, La Notte *et al.* (1997) report that the dried spotted membrane can be stored at room temperature for a month. This increases the applicability of the technique for field-work, as dry samples at room temperature can be stored and transported far more easily than liquids or samples requiring cold storage. Thus samples may be collected, directly spotted onto the membrane, and DNA extracted upon return to the laboratory.

As no DNA quality difference could be seen in the final PCR product from all successful extraction methods, initial PCR quality did not depend on the presence or absence of a purification step. Therefore, when extracting DNA for amplification by PCR, there is no intrinsic need to include a purification step. In fact, a purification step

both lengthens the process and increases the chances of losing DNA from small specimens. However, purification of the DNA allows long-term storage of the extracted product. The rapid extractions presented here, which undergo no purification, do not remain stable for more than a few weeks. In most diagnostic applications, where the most important outcome is rapid identification of a sample, and samples are discarded once this identification has been made, long-term storage of the extracted DNA is unnecessary and a purification step would be superfluous. Purification protocols are best used only during long-term research projects, where DNA may be needed repeatedly.

The two rapid, sensitive and simple DNA extraction techniques presented here are ideal for diagnostic work. They are much faster and more reliable than the more widely used traditional methods, being able to extract DNA from a greater percentage of specimens. They are also extremely simple, and can be easily implemented as part of a pre-existing molecular diagnostic procedure.

CHAPTER THREE

Determination of nucleotide sequence from three mealybug species

Chapter Three
Determination of nucleotide sequence from three mealybug species

Abstract

When developing a DNA-based identification method, accurate nucleotide sequence data are required. In this chapter, mitochondrial DNA was amplified and sequenced from the cytochrome *c* oxidase subunit 1 (CO I) gene of three mealybug species, namely *Planococcus ficus*, *Planococcus citri* and *Pseudococcus longispinus*. A total of 1218, 1227 and 763 bp were obtained from *P. ficus*, *P. citri*, and *P. longispinus* respectively. The nucleotide sequence obtained showed very little intraspecific variation (<1%) and moderate interspecific variation (7-12%). These values were comparable to those from mealybug mitochondrial sequences published in previous studies. The data obtained here will be of use in further development of a DNA-based identification technique.

Introduction

In order to develop an identification method, stable molecular markers are needed to distinguish between species. In mealybugs (Hemiptera: Pseudococcidae), not enough molecular data are available to develop molecular markers or a DNA-based identification system. Therefore, nucleotide sequence of the three mealybug species in question needed to be obtained. This study focussed on nucleotide sequence from the mitochondrial cytochrome *c* oxidase subunit 1 (CO I) gene from the three mealybug species in question, namely *Planococcus ficus* (Signoret), *Planococcus citri* (Risso) and *Pseudococcus longispinus* (Targioni-Tozzetti).

Currently, very little mealybug sequence data are available. Those that are available on Genbank (National Centre for Biotechnology Information (NCBI), a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH)), are mainly from mealybugs outside this study (e.g. *Ferrisia* spp.: Gullan *et al.* 2003) or from nuclear or ribosomal genes (Beuning *et al.* 1999, Cook *et al.* 2002,

Downie and Gullan 2004, Mohan and Chandra, unpubl., Genbank accession number AF156881). These sequences are not of use in this study, since most of the regions are too conserved to distinguish between species. Genbank contains only 24 mealybug mitochondrial sequences, most of which are from mealybugs outside the two genera covered in this study. Only three published nucleotide sequences are available for the CO I gene of any of the three species used in the present study. Two are short sequences (385bp) from the mid-section of CO I of *P. citri* and *P. longispinus*, respectively (Gullan *et al.* 2003). The other is a longer segment of the CO I and CO II gene from *P. citri* (Thao *et al.* 2002). No sequence data for *P. ficus* are available.

Since sequence data relevant to the present study are unavailable, DNA from the three species needed to be amplified and sequenced. Ideally, this would be done by using PCR primers from previous studies on the same or closely related species. However, since no prior sequence data are available for this study, other methods must be used to amplify and sequence a gene region. One possibility is to build a randomly-primed synthetic-DNA library. DNA fragments (clones) are inserted into bacteria vectors and each clone is sequenced using primers specific to the bacteria. The overlapping fragments are then combined into a cDNA library to create a genetic map of the species. This data can then be used to design primers to amplify a gene region of interest. However, this method is costly and time-consuming. Another method, which is far simpler and cheaper, is to use published universal primer pairs to amplify a gene region (Brower and DeSalle 1994, Hillis *et al.* 1996). A comprehensive list of universal primers for amplification of insect mitochondrial DNA (mtDNA) is found in Simon *et al.* (1994). These universal primers are directed against more conserved DNA regions flanking variable regions. The variability of certain regions is similar in many insects. A region of known variability can therefore be chosen and amplified using universal primers.

This chapter describes the methods used to amplify and sequence mtDNA from the CO I gene of *P. ficus*, *P. citri* and *P. longispinus* using universal primers. The sequence data obtained will be analysed with the intent of finding a genetic region of suitable variability for species identification.

Materials and methods

DNA was extracted from whole mealybugs as discussed previously in Chapter Two. Mealybugs of all life stages were used. DNA from all successful extraction methods was used for PCR amplification, sequencing and analyses as described below.

Polymerase chain reaction (PCR)

Each 25 µL PCR reaction mix consisted of the following: 2.5 µL 10× NH_4 Reaction Buffer (160 mM $(NH_4)_2SO_4$, 670 mM Tris-HCl, pH 8.8, 0.1% (v/v) Tween-20), 4 mM magnesium chloride ($MgCl_2$), 50 µM each dNTP, 0.4 µM each primer, 0.05 units/µL BIOTAQ™ DNA Polymerase (BioLine GmbH, Luckenwalde, Germany) and 1-2 µL extracted DNA.

Amplification reactions were run in a thermal cycler (Perkin Elmer GeneAmp® PCR System 2400 or Eppendorf Mastercycler® gradient). The specific cycle followed depended on the melting point of the primers, the predicted product length and the result of previous cycles, and will be described below.

An extra reaction without template DNA was run with each cycle as a negative control. In cases in which amplification was problematic (see discussion below), DNA from an extraction which had previously amplified well was used as a positive control.

Primers

The universal primers C1-J-1718, C1-N-2191 (alias Nancy), C1-J-2183 (alias Jerry) and TL2-N-3014 (alias Pat) were used (Simon *et al.* 1994). The primer set C1-J-1718 and C1-N-2191 bordered a region of 473 bp from the mid-section of the CO I gene. C1-J-2183 and TL2-N-3014 bordered a region of 831 bp from the 3'-end of CO I. Primer names, sequences and melting temperatures are given in Table 3.1. The primer names follow the convention of Simon *et al.* (1994), in which the number refers to the position of the 3' base relative to *Drosophila yakuba* Burla (Diptera: Drosophilidae) (Clary and Wolstenholme 1985).

Primers were synthesized by Invitrogen Life Technologies, and arrived in a lyophilised state of known amount. They were then resuspended in TLE buffer to

produce a 200 µM stock solution. TLE buffer consisted of 0.01 M Tris, pH 7.4 and 0.1 mM EDTA, pH 8. These stock solutions were diluted with sterile distilled water to 10 µM working solutions. Stock solutions were stored at -70ºC, working solutions at -20ºC. Primer melting temperatures (T_m) were usually supplied by the manufacturer. If this was not supplied, thermodynamic T_m was calculated using the thermodynamic values for nucleotide interactions given in Breslauer *et al.* (1986). This method is accurate to within 4.4 ± 3.8ºC (| predicted – observed T_m | ± standard deviation) of the experimentally observed melting temperature (Owczarzy *et al.* 1997).

Table 3.1. Details of primers used to amplify portions of the CO I gene, taken from Simon *et al.* (1994).

Name	Primer sequence (5'-3')	Melting temperature
C1-J-1718	GGAGGATTTGGAAATTGATTAGTTCC	58.5ºC
C1-N-2191	CCCGGTAAAATTAAAATATAAACTTC	53.6ºC
C1-J-2183	CAACATTTATTTTGATTTTTTGG	50.2ºC
TL2-N-3014	TCCAATGCACTAATCTGCCATATTA	58.4ºC

Amplification of the mid-section of CO I

Primers C1-J-1718 and C1-N-2191 were used to amplify the mid-section of CO I. As the conventional PCR cycle produced non-specific bands, a touchdown PCR was used to minimise the problem of non-specific primer annealing. The PCR cycle consisted of an initial denaturation temperature (T_d) of 94ºC for 3 min, then one cycle with annealing temperature (T_a) of 60ºC for 30 sec, extension temperature (T_e) of 72ºC for 15 sec, then T_d of 94ºC for 45 sec. T_a was decreased in successive cycles to 58, 56, 55 and 54ºC, respectively, for one cycle each. T_a was then decreased to the final

temperature of 53ºC for 30 cycles. Samples were cooled to at least 10ºC, removed and stored at 4ºC until needed.

Amplification of the 3'-end of CO I

Primers C1-J-2183 and TL2-N-3014 were used to amplify the 3'-end of CO I. A conventional PCR cycle was used, consisting of an initial denaturation at 94ºC for 1 min, then 35 cycles with T_d = 94ºC for 45 sec, T_a = 48ºC for 45 sec and T_e = 72ºC for 1 min, and a final extension period at 72ºC for 3 min. Samples were cooled to at least 10ºC, removed and stored at 4ºC until needed.

Visualisation of PCR products

Amplification products were separated by electrophoresis on a 1.5% agarose gel. The gel was prepared by adding agarose to TAE buffer (0.04 M Tris, 1 mM Na_2EDTA, pH 8 and 0.1% glacial acetic acid) and boiling in a microwave until dissolved. The mixture was cooled to room temperature and 6 µL 0.08 mM ethidium bromide (EtBr) was added and mixed by swirling. The gel was then poured into a plastic mould, combs were inserted, and the gel was allowed to set. Once set, the gel was placed in an electrophoresis tank filled with TAE buffer. 5 µL of the PCR product was mixed with 2 µL 1× Loading Buffer (6× Loading Buffer, Promega, contained 0.03% bromophenol blue, 0.03% xylene cyanol, 0.4% orange G, 15% Ficoll 400, 10 mM Tris-HCl, pH 7.5 and 50 mM EDTA) and loaded into the wells (Sambrook and Russell 2001). A DNA marker (Promega) was used as a size comparison. The DNA marker contained a total of eleven fragments, with ten fragments ranging from 100 bp to 1000 bp in 100 bp increments and an additional fragment of 1500 bp. The gel was run at 50 V for 12-20 min, depending on the expected fragment size. The gel was then observed under ultraviolet (UV) light to detect the presence of bands, indicating the presence of amplified DNA.

PCR product purification

Successfully amplified products were prepared for sequencing by purification with sodium acetate (NaAc). A solution consisting of 100 µL absolute ethanol, 20 µL ddH_2O

and 2 µL 3 M NaAc per reaction was prepared and added to the remaining PCR product. This mixture was vortexed for approximately 30 sec, then centrifuged for 15 min at 12 000 rpm. The supernatant was discarded, 100 µL 70% ethanol added, and the mixture centrifuged for 10 min at 12 000 rpm. This washing step was repeated, and the supernatant again discarded. The remaining pellet was then dried at 80ºC for 3-5 min and resuspended in 30 µL sterile distilled water.

Cycle sequencing

Depending on the concentration of template DNA present after purification, 2-4 µL of the purified PCR product was used in cycle sequencing. The BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used for sequencing.

For the primer set C1-J-1718 and C1-N-2191, each 10 µL cycle sequencing reaction mix consisted of the following: 1 µL 5× sequencing buffer (0.4 M Tris-HCl, pH 9 and 0.01 M $MgCl_2$), 2 µL BigDye™, 2 µL 1.6 µM primer and ddH_2O.

For the primer set C1-J-2183 and TL2-N-3014, each 10 µL reaction mix consisted of 3 µL 5× sequencing buffer, 1 µL BigDye™, 1 µL 1.6 µM primer and ddH_2O.

Reactions were run on a thermal cycler (Perkin Elmer GeneAmp® PCR System 2400) with the following cycle: initial T_d = 94ºC for 2 min, then 25 cycles with T_d = 96ºC for 10 sec, T_a for 5 sec and T_e = 60ºC for 4 min, then cooled to 4ºC for 1 min. The annealing temperature used for cycle sequencing was one degree Celsius below that used for the PCR cycle. Thus primers C1-J-1718 and C1-N-2191 were cycled at an annealing temperature of 52ºC, and primers C1-J-2183 and TL2-N-3014 at 47ºC. Amplification products were stored at 4ºC until DNA precipitation was performed.

DNA precipitation

The cycle sequencing product was precipitated in preparation for sequencing by using sodium acetate to aid in precipitation of the DNA. A solution consisting of 50 µL absolute ethanol, 10 µL ddH_2O and 2 µL 3 M NaAc per reaction was prepared and added to the cycle sequencing product. This mixture was vortexed for approximately 30 sec and centrifuged for 15 min at 12 000 rpm. The supernatant was discarded, 100 µL 70% ethanol added, and the mixture centrifuged for 10 min at

12 000 rpm. This washing step was repeated, and the supernatant again discarded. The pellet was then dried at 80ºC for 3-5 min, and stored dry at -20ºC until sent for sequencing.

DNA sequencing

DNA sequencing was performed using an automated version of the dideoxynucleotide chain termination method (Sanger *et al.* 1977, Wrischnik *et al.* 1987) on an ABI3100 PRISIM™ Genetic Analyzer (Applied Biosystems) at the DNA Sequencing Facility, University of Pretoria.

PCR products from primers C1-J-1718 and C1-N-2191 were sent for sequencing in both the forward and reverse directions from 15 and eight individuals of *P. ficus* and *P. citri*, respectively. PCR products from primers C1-J-2183 and TL2-N-3014 were sent for sequencing in both the forward and reverse directions from ten, three and four individuals of *P. ficus*, *P. citri* and *P. longispinus*, respectively.

Sequence analysis

The sequences obtained from the DNA Sequencing Facility were in the form of .abi files and were viewed in the software program Chromas Lite 2.00 (Technelysium Pty. Ltd.). Sequence quality was judged by eye and sequences of sufficient quality were preliminarily edited by discarding the bases at the beginning and end of the sequence where the information was ambiguous.

Forward and reverse sequences, respectively, for each species were then aligned using Clustal X (Thompson *et al.* 1994 and 1997, Jeanmougin *et al.* 1998). Each sequence was then individually checked, and each mis-matched base re-examined. It was found that many mis-matches were due to errors in the automated reading of the sequence. The reasons for these errors will be discussed later. The errors were corrected, and the sequences then re-aligned. From this information, forward and reverse consensus sequences were generated. From the reverse sequence, a complimentary sequence was generated and compared to the forward sequence. The combination of these sequences was presented as the consensus sequence for that species.

These consensus sequences were then compared to each other as well as to published sequences by aligning them in Clustal X. Similarities and differences were noted. Specified and degenerate bases in these sequences were named according to the conventions of the International Union of Biochemistry (IUB): G = guanine; C = cytosine; A = adenine; T = thymine; Y = pyrimidine (C/T); R = purine (G/A); M = A/C; K = G/T; S = G/C and W = A/T.

Results

Figure 3.1 shows the relative positions of the primers used in amplification and sequencing of the CO I region of the mitochondrial molecule. The forward nucleotide sequence and associated amino acid translations from each of the three mealybug species, written 5' to 3' by convention, are given in Appendix 1 and 2, respectively. Alignments of the nucleotide and amino acid sequences from these three species are given in Appendix 3 and 4, respectively.

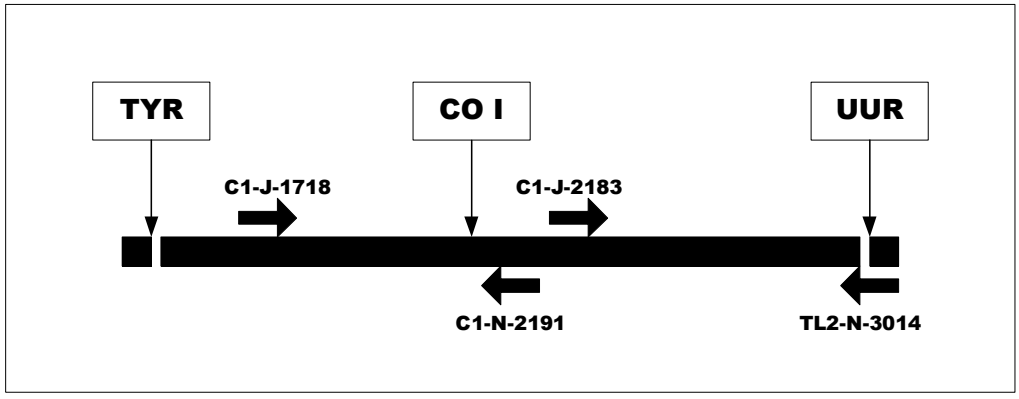

Figure 3.1. Relative positions of the forward and reverse primers amplifying the CO I gene. TYR: tRNA Tyrosine. CO I: cytochrome *c* oxidase subunit 1. UUR: tRNA Leucine.

Nucleotide sequence

***Planococcus ficus*.** Using primers C1-J-1718 and C1-N-2191, ten forward and eight reverse sequences, respectively, of sufficient quality were obtained from *P. ficus*. In the forward sequence, 424 bases were aligned with 98.9% similarity. In the reverse sequence, 422 bases were aligned with 99.8% similarity. Using primers C1-J-2183 and TL2-N-3014, ten forward reverse sequences, respectively, were obtained. In the forward sequence, 693 bases were aligned with a 99.7% similarity. In the reverse sequence, 611 bases were aligned with a 99.9% similarity. Once aligned to generate a consensus sequence, a total of 1218 bp of DNA sequence were obtained from *P. ficus* showing an over-all AT-bias of 81.7%. 458 bp were obtained from the mid-section of CO I (primer set C1-J-1718 and C1-N-2191) with an 82.1% AT-bias. 760 bp were obtained from the 3'-end of CO I (primer set C1-J-2183 and TL2-N-3014) with an 81.2% AT-bias.

***Planococcus citri*.** Using primers C1-J-1718 and C1-N-2191, three forward and six reverse sequences, respectively, of sufficient quality were obtained from *P. citri*. In the forward and reverse sequences, 421 and 451 bases were aligned, respectively. Each showed 100% similarity. Using primers C1-J-2183 and TL2-N-3014, three forward and reverse sequences, respectively, were obtained. In the forward sequence, 762 bases were aligned with a 99.4% similarity. In the reverse sequence, 681 bases were aligned with a 99.5% similarity. In the consensus sequence, a total of 1227 bp of nucleotide sequence were obtained from *P. citri* showing an over-all AT-bias of 82.7%. 465 bp were obtained from the mid-section of CO I (primer set C1-J-1718 and C1-N-2191) with an 84.5% AT-bias. 762 bp were obtained from the 3'-end of CO I (primer set C1-J-2183 and TL2-N-3014) with an 80.8% AT-bias.

***Pseudococcus longispinus*.** When using primers C1-J-1718 and C1-N-2191, nucleotide sequence data could not be obtained for *P. longispinus.* Possible reasons for this will be discussed later. Using primers C1-J-2183 and TL2-N-3014, four forward and reverse sequences, respectively, were obtained. In the forward sequence, 730 bases were aligned with a 99.9% similarity. In the reverse sequence, 741 bases were aligned with a 99.6% similarity. In the consensus sequence was 763 bp long with an 80.2% AT-bias.

Comparison of nucleotide sequences from the three mealybug species

Alignments of nucleotide and amino acid sequences from the three mealybug species are found in Appendix 3 and 4, respectively.

Intraspecific variation in all amplified nucleotide sequences was very low, varying from 98.9-100% similarity. Interspecific nucleotide similarity between *P. ficus* and *P. citri* in the mid-CO I region was 93.1% in the region of overlap. In the 3'-section of CO I, the three species showed an over-all similarity of 85.2% in the region of overlap, with *P. ficus* and *P. citri* being more similar (92.5%) than either *P. ficus* and *P. longispinus* (89.1%) or *P. citri* and *P. longispinus* (88.6%).

When analysing interspecific nucleotide substitutions, A-T transversions were very common, accounting for 48.9% of substitutions (74 of 152). This transversion was more common than C-T transitions (30.3%, 46 of 152), which was the next most common substitution. 81.6% (124 of 152) of the nucleotide substitutions occurred at 3rd codon positions, 3.9% (6 of 152) at 2nd codon positions and 14.5% (22 of 152) at 1st codon positions. Only 22 (14.5%) nucleotide changes resulted in an amino acid replacement. Three of these were due to degenerate bases in the nucleotide sequences, and the corresponding amino acids therefore not being specified. Thus a total of 130 (85.5%) silent nucleotide substitutions occurred between the three species.

Comparison to published mealybug sequences

The nucleotide sequences obtained from the three mealybug species used in this study were compared to available DNA sequences on GenBank by the BLAST search algorithm (Altschul *et al.* 1997). A summary of the BLAST search results is given in Appendix 5.

Published sequences from the study by Gullan *et al.* (2003) were 385 bp long, and showed approximately a 350 bp over-lap with the sequences obtained in the present study from the 3'-end of CO I. Published sequences from the study by Thao *et al.* (2002) were 1452-1458 bp long (CO I and CO II genes), and showed a 755 bp over-lap with the sequences obtained in the present study from the 3'-end of CO I. Only three published CO I sequences are available for the mealybug species in this study: two for

P. citri (Gullan *et al.* 2003, Thao *et al.* 2002) and one for *P. longispinus* (Gullan *et al.* 2003). No nucleotide sequence data are available for *P. ficus*.

A comparison of the three *P. citri* sequences (350 bp alignment) showed an 88.9% similarity. A pairwise comparison showed the *P. citri* sequence from this study and that from Gullan *et al.* (2003) to be more similar (98.3% identity) than either that from this study and Thao *et al.* (2002) (89.4% identity), or that from Gullan *et al.* (2003) and Thao *et al.* (2002) (91.1% identity). Differences between the two former sequences were all due to degenerate bases in the sequences, and resulted in only one possible amino acid change (unspecified due to the degenerate base). Differences between the former sequences and that from Thao *et al.* (2002) showed a large proportion of A-T transversions (63.6%, 21 of 33) and a lower amount of C-T transitions (24.3%, 9 of 33). The amino acid translation of the sequence from Thao *et al.* (2002) showed six changes when compared to the former two sequences (differences due to degenerate bases are not included in the calculation).

A comparison of the two *P. longispinus* sequences (348 bp alignment) showed 98.6% similarity, with five transitions: four pyrimidine (C/T) transitions and one purine (G/A) transition. No amino acid changes occurred.

Comparison to the genetic sequence of *Drosophila yakuba*

In order to make insect genetic studies more comparable, it has been suggested that all sequences obtained be referenced to some standard (Caterino *et al.* 2000). For insect mtDNA, this standard is *Drosophila yakuba*. The *D. yakuba* sequence was obtained from Genbank (Clary and Wolstenholme 1983 and 1985, Genbank accession number NC001322). The amplified sequences obtained in the present study were aligned to the *D. yakuba* sequence using Clustal X. The amplified sequences correspond to the CO I gene of *D. yakuba* (positions 1470 - 3010, Clary and Woltenholme 1983 and 1985). Relative to *D. yakuba*, the amplicons from the primer set C1-J-1718 and C1-N-2191 extend through positions 1749-2216 and 1693-2152 in *P. ficus* and *P. citri*, respectively. The two segments of *P. ficus* overlap by one base at position 2216. The two segments from *P. citri* show a 27 nucleotide gap. The amplicons from primer set C1-J-2183 and TL2-N-3014 extend through positions 2216-2978, 2219-2982 and 2221-2986 in *P. ficus*,

P. citri and *P. longispinus*, respectively. Overall, the sequences did not show a very high similarity to *D. yakuba*: 67.0%, 66.7% and 68.5% in *P. ficus*, *P. citri* and *P. longispinus*, respectively.

Discussion

Performance of primer set C1-J-1718 and C1-N-2191

The primer set C1-J-1718 and C1-N-2191 gave poor results when amplifying and sequencing mealybug DNA.

As noted briefly before, the original amplification product yielded a number of bands due to non-specific primer annealing. Increasing the annealing temperature resulted in inhibition of the reaction. Therefore a touchdown PCR was performed, which in most cases yielded a single product.

In order to obtain amplification in most samples, the primers needed to be heated briefly (approx. 5 min) at 94ºC and cooled rapidly on ice. The high temperature prevented the occurrence of primer-primer complexes, making more primer available for annealing to the DNA once the PCR reaction commenced.

In some cases, PCR was not successful, most notably in *P. longispinus*, which was never amplified using this primer set. In attempting to amplify *P. longispinus*, the forward primer C1-J-1718 also was matched with reverse primer TL2-N-3014. Although this combination was able to amplify a specimen of *P. ficus*, *P. longispinus* could still not be amplified. This was probably due to primer C1-J-1718 not annealing to the target *P. longispinus* DNA, most likely due to mis-match between the primer and the DNA. This in surprising, as the universal primers anneal to conserved nucleotide regions, and can tolerate a high degree of mis-match before annealing is completely inhibited (Kocher *et al.* 1989, Sommer and Tautz 1989). In order to substantiate this, the entire CO I gene would need to be amplified and sequenced. This, however, is beyond the scope of this study. Future studies would benefit from investigating this possibility by sequencing across position 1718 to determine the possibility of a primer-DNA mis-match. There are currently no mealybug DNA sequences in any published study which cover this region. If this section is indeed very different in *P. longispinus* (and possibly

other mealybugs) it could have implications for genetic studies of mealybugs in the areas of phylogenetics, evolution and functionality.

In those samples that did amplify, the product was prepared and sent for sequencing. Sequencing of the mid-section of CO I with primers C1-J-1718 and C1-N-2191 was also problematic, and a number of sequences were of such poor quality as to render them unusable. In total, 22 products were sequenced with both the forward and reverse primers. Of these, only thirteen forward sequences (primer C1-J-1718) and twelve reverse sequences (primer C1-N-2191) could be used for analyses. Usable sequence length (i.e. length after discarding ambiguous bases at the beginning and end of the sequence) was between 421 and 451 bases long.

Thus, a number of problems were encountered during amplification and sequencing of the mid-sections of CO I with primers C1-J-1718 and C1-N-2191. The cause of these problems could be due to a number of factors. The most likely reasons are related to the extremely high A+T content of the sequence, as discussed below.

Adenine and thymine form only two hydrogen bonds when pairing, as compared to guanine and cytosine, which form three (Watson and Crick 1953a and 1953b). Thus the A-T bond is weaker and less specific than the G-C bond. This could account for the non-specificity and decreased annealing probability of the primers. It has been hypothesised that DNA polymerases in an A-T-rich genome are specialised for A-T bonds, and do not function optimally with G-C bonds (Clary and Wolstenholme 1985, Wolstenholme and Clary 1985). Conversely, it may be possible that a polymerase that has evolved in an unbiased genome (i.e. an equal number of all four nucleotides) would function sub-optimally when faced with an extremely AT-biased gene region. If this is the case, sub-optimal extension could account for the poor quality of many of the amplification products and sequences. In addition, commercial sequencing mixes (as used here) contain equal amounts of the four nucleotides. In a sequence that is AT-rich, a far larger proportion of As and Ts are used from the mixture than Cs and Gs. This has two main consequences. Firstly, the amount of As and Ts in the mixture will be exhausted more quickly than expected, thus preventing any further amplification due to lack of the correct nucleotides. Secondly, a large number of Cs and Gs will remain unused in the reaction mix. When using labelled nucleotides for sequencing, these

unused nucleotides can precipitate with the DNA, and contribute to a CG background in the sequence chromatograph (G. Malherbe, pers. comm.). This confirms what was observed in the present study, and was largely the cause of the errors in the automated reading of the sequence. Although commercial sequencing kits with higher proportions of As and Ts are available, they are expensive and seldom significantly increase the quality of the sequence (R. Zipvel, pers. comm.). Therefore, in this study, a larger amount of BigDye™ was used when sequencing this region, so as to provide more As and Ts. To minimise co-precipitation of the remaining Cs and Gs during DNA precipitation, all reactions were carried out at room temperature instead of on ice. These procedures improved the quality of the sequences obtained.

Performance of primer set C1-J-2183 and TL2-N-3014

The primer set C1-J-2183 and TL2-N-3014 yielded good quality amplification and sequencing products of mealybug DNA. Extracted DNA was always amplified cleanly and high-quality sequences were obtained. Of the seventeen products sequenced, all yielded good quality forward and reverse sequences. Although a PCR product of approximately 830 bp was obtained, the sequencer was not able to accurately read the latter part of the sequence. Therefore a large number of ambiguous bases needed to be discarded from the end of each sequence. Despite this, good quality sequences of between 611 and 762 bases in length were obtained.

Intra- and inter-specific variation

The CO I mealybug sequences obtained showed little, if any, intraspecific nucleotide variation (98.9-100% identity), but enough interspecific variation to clearly delineate species (85.2-93.1% identity). This low variability is due to the conserved nature of the CO I gene. Higher levels of interspecific variation would be undesirable, as this raises the possibility of multiple substitutions at a single site, which would obscure the data (Brower and DeSalle 1994, Simon *et al.* 1994).

These levels of variation are comparable to the intra- and interspecific variation found in the CO I sequences from other mealybugs. Four nucleotide sequences from the mealybug species *Ferrisia gilli* (Gullan *et al.* 2003) showed a 98.8% intraspecific

similarity when aligned in Clustal X. However, nine sequences from *F. virgata* (Gullan *et al.* 2003) showed a lower similarity of 95.4%. Interspecific variation within the mealybug genus *Ferrisia* showed a 95.9% similarity between species (Gullan *et al.* 2003).

Amongst the three mealybug species, a large number of A-T transversions were observed (48.9% of all substitutions). However, with the exception of the published *P. citri* sequence from Thao *et al.* (2002), no A-T transversions were observed within species between the published and amplified sequences. It is interesting to note that in the only intraspecific comparison in which A-T transversions were observed (i.e. with *P. citri* from Thao *et al* 2002), the sequences showed an over-all identity more similar to that obtained in interspecific comparisons than intraspecific comparisons. In theory, transitions (pyrimidine (C-T) or purine (G-A) substitutions) should occur more readily than transversions due to the more similar structure of the bases involved (Watson and Crick 1953b). However, this trend is expected to be reversed in AT-rich genomes, as transitions would disrupt the high AT-bias (Crozier and Crozier 1993). The seemingly opposite trends observed in this study between and within species are surprising. The fact that no intraspecific transitions were observed may be due to the low variation observed in CO I. A high number of transitions may only be observed over longer evolutionary time periods, as has been suggested by comparisons within and between *Drosophila* species: within species or between sibling species, transitions dominate (Satta *et al.* 1987), but between more distantly related species, A-T tranversions are more common (Woltenholme and Clary 1985).

AT-bias in insect mitochondrial genomes

AT-bias in a genome has a number of consequences for molecular analyses, as has been discussed in detail in Chapter One. Here, some of these properties will be discussed specifically as they relate to the mealybug CO I sequences obtained. The difficulty in PCR amplification and sequencing of the mealybug DNA from the mid-section of CO I is most likely related to the high AT-content of the region. These difficulties were discussed above.

All mealybug sequences showed a strong AT bias, which is common in insect mtDNA. Simon *et al.* (1994) reviewed various insect mitochondrial gene sequences, and reported an AT bias of between 67% (silverfish) and 82% (honey bee), compared to the average AT bias in mammalian mtDNA of 55-64%. In the honeybee (*Apis mellifera ligusta* Spinola (Hymenoptera: Apidae)) mitochondrial genome, Crozier and Crozier (1993) reported an over-all AT-bias of 84.9%, and a slightly lower AT-bias of 83.3% in protein-coding genes (such as CO I). This AT-bias is the highest yet reported for any mitochondrial genome. The average AT-bias found in the three mealybug species (81.5%), although lower than that of the honeybee, is nevertheless much higher than that found in *D. yakuba* (67.9%). It has been shown that AT-bias is significantly associated with rate of genetic evolution (Lin and Danforth 2004). Thus, the high AT-bias observed in these three mealybug species may indicate a faster evolving genome and a more derived lineage.

Conclusion

In this chapter, the nucleotide sequence of the CO I gene from three mealybug species has been reported and analysed. The nucleotide content of the mid- and 3'-end of the CO I gene shows very little intraspecific variation (98.9-100% identity) and moderate interspecific variation (85.2-93.1% similarity). This level of variation is ideal when designing a species identification method. Too much intraspecific variation would result in false identifications due to populational differences, and too little interspecific variation would not allow for clear differentiation between species. Therefore the nucleotide data presented here can be used in the further development of a DNA-based identification protocol for *P. ficus*, *P. citri* and *P. longispinus*.

In addition, in obtaining the nucleotide sequence data, a number of interesting trends were noticed. Most notably, the AT content of the CO I gene from all three mealybug species was very high (average 81.5%). This led to difficulties in amplification and sequencing of one of the DNA regions, for which modifications were made to methods. These modifications, as well as others, may be necessary when working with most strongly AT-biased genomes. The strong AT bias also suggests a

more rapidly evolving genome, and comparative studies between mealybugs and other similar insects may yield fascinating results. However, pursuit of these questions is beyond the scope of the present study, and is best left to researchers in the fields of molecular function and evolution. For the purposes of this study, the nucleotide sequence data obtained will be used to develop a DNA-based mealybug identification method for large-scale use in the wine industry.

CHAPTER FOUR

A multiplex PCR assay for the simultaneous identification of three mealybug species (Hemiptera: Pseudococcidae)

Chapter Four

A multiplex PCR assay for the simultaneous identification of three mealybug species (Hemiptera: Pseudococcidae)

Abstract

Molecular species identifications are becoming more wide-spread in ecological studies, particularly with regard to insects for which morphological identification is difficult or time-consuming. In this study we describe the development and application of a single-step multiplex PCR for the identification of three mealybug species (Hemiptera: Pseudococcidae) associated with grapevine: *Planococcus ficus* (vine mealybug), *Planococcus citri* (citrus mealybug) and *Pseudococcus longispinus* (longtailed mealybug). Mealybugs are pests on many commercial crops, including grapevine, in which they transmit viral diseases. Morphological identification is time-consuming, requires a high level of expertise, and can usually only identify adult females. The multiplex PCR developed here is rapid, reliable, sensitive, accurate and simple. The entire identification protocol (including DNA extraction, PCR and electrophoresis) can be completed in approximately four hours. Specimens from which DNA could be extracted were always correctly identified (100% accuracy; 90% (192 of 213) DNA extraction success). In addition, the procedure is simple enough to be implemented in any molecular laboratory. The principles described here can be extended to any organism for which rapid, reliable identification is needed.

Introduction

Mealybugs (Hemiptera: Pseudococcidae) are cosmopolitan pests on many commercial plants. They feed on the phloem of the plant, lowering plant vigor and causing fruit scarring when heavy infestations are present. Mealybugs also secrete honeydew, a substrate on which sooty mould grows. In addition, mealybugs act as

vectors of a number of plant pathogens, including grapevine leafroll viruses (e.g. Engelbrecht and Kasdorf 1990).

Mealybugs belong to the family Pseudococcidae, consisting of more than 2000 described species in 270 genera (Downie and Gullan 2004). The adult female is small (1-3 mm in length), ovoid and resembles the nymphal stages. The adult male is winged, non-feeding and short-lived (Downie and Gullan 2004). Identification of mealybugs is problematic, as they are small and may be superficially very similar (Gullan and Kosztarab 1997). Identification relies on features only visible under a microscope, and specimens therefore need to be painstakingly slide-mounted, a process which can take a number of days and requires familiarity with microscopic cuticular features (Millar 2002, Watson and Kubiriba 2005). This makes accurate identification of mealybugs very difficult. In addition, only mealybugs for which detailed keys are available can be identified. These keys usually describe the adult female only. Therefore, nymphs and males cannot at present be identified.

Despite these difficulties in mealybug identification, accurate and timely identification is critical in many sectors of agriculture. For instance, pheromone traps, which are used in monitoring of infestation levels in the field, are species-specific. Therefore correct initial identification of the mealybug species present is important. In addition, since only the males are caught in the traps, a direct confirmation of the species trapped is impossible, and a field-survey must be conducted to look for adult females to corroborate the identification of the infesting species. Another agricultural sector in which timely identification is crucial in the import and export of fresh produce. When mealybugs are encountered on a consignment of fruit, the produce must be held in quarantine in cold stage until the mealybugs can be identified. If nymphs are found, they must be reared to maturity, and then only adult females can be identified. This rearing process may take up to six weeks (Beuning *et al.* 1999). If nymphs cannot be reared, or only males are found, no identification can be made and the consignment must be destroyed, causing severe losses to the exporter/importer.

For these reasons, an accurate and rapid mealybug identification technique is needed which can identify damaged, immature and male specimens in addition to females. This study aimed to develop a molecular identification method for the species

of mealybug associated with grapevine in South Africa. This technique needed to be fast, accurate, reliable and sensitive enough to identify nymphs and male mealybugs, and simple enough to be implemented in any molecular laboratory. A few DNA-based methods for insect species identification have been developed in the past, (e.g. Zehner *et al.* 2004), but many do not fulfill the required criteria.

Only one DNA-based method has previously been developed for mealybug identification (Beuning *et al.* 1999). That study used a PCR-based approach to identify four New Zealand mealybug species (*Pseudococcus viburni* (Signoret), *P. calceolariae* (Maskell), *P. longispinus* (Targioni-Tozzetti) and *P. similans* (Lidgett)). This method employed variation in the internal transcribed spacer (ITS) regions of the nuclear ribosomal DNA to develop species-specific PCR primers. These primers were then used in a series of PCRs, each employing a separate set of species-specific primers, to identify each of the four species. Although this process is significantly faster than traditional morphological identification, it is still time-consuming. A faster identification method is that of the multiplex (or cocktail) PCR, in which a number of species-specific primers are used in a single PCR reaction. This method has been used in clinical diagnosis of infections and strains present in a single sample (Bermingham and Luettich 2003). It has only recently been applied to insect identification. For instance, Koekemoer *et al.* (2002) developed a multiplex PCR to identify six species of the *Anopheles funestus* Giles (Diptera: Culicidae) mosquito group in a single PCR reaction. The species-specific primers employed in this study were designed to yield DNA products of different lengths that could be separated by electrophoresis, thus determining the identity of the sample.

Therefore, in this study the development of a multiplex PCR for the identification of the three mealybug species most commonly associated with grapevine in South Africa was undertaken. These species are the vine mealybug *Planococcus ficus* (Signoret), the citrus mealybug *Planococcus citri* (Risso) and the longtailed mealybug *Pseudococcus longispinus*. All three species occur on vines in all grape-growing areas of the world (e.g. La Notte *et al.* 1997b, Golino *et al.* 2002), although their pest status in different areas may differ. In South Africa, *P. ficus* is the major pest in vineyards, and has been shown to transmit grapevine viruses. *Pseudococcus longispinus* is a highly polyphagous

pest also occurring in vineyards. It has also been shown to transmit grapevine viruses. *Planococcus citri* has historically been reported on South African vines, although the original species identification has been questioned. It does occur in other wine-growing areas of the world (e.g Golino *et al.* 2002). *Planococcus citri* and *P. ficus* are very closely related, and are morphologically very similar. Mis-identifications of these two species are common (Ben-Dov and German 2003). Therefore, an accurate and reliable identification technique for these three species is essential to further research.

Materials and methods

Mealybugs

Mealybugs were obtained from field collections and laboratory colonies as listed in Table 4.1. To confirm their identity, sub-samples of all mealybug collections were sent to I.M. Millar of the Biosystematics Division of the Plant Protection Research Institute of the Agricultural Research Council (ACR-PPRI), South Africa. For initial DNA amplification and sequencing, and optimization and testing of the multiplex PCR, mealybug specimens of known species were used. Ten, three and four individuals of *P. ficus*, *P. citri* and *P. longispinus*, respectively, were used to obtain sequence data. During testing of the multiplex PCR and application to field collections, specimens were of unknown species (n = 30 and n = 185, respectively). Collected mealybugs were stored in absolute ethanol at ambient temperature and transferred to -20ºC upon return to the laboratory (Post *et al.* 1993, Fukatsu 1999, Tayutivutikul *et al.* 2003). Both nymphs and adult mealybugs were used in analyses.

DNA extraction

DNA was extracted from whole mealybugs using either the direct buffer (Koekemoer *et al.* 2002b) or spot-PCR method (La Notte *et al.* 1997a), with the following adaptations: 50-100 µL STE buffer was used in the direct buffer method, 30 µL STE buffer was used in the spot-PCR and no primers were added during extraction. 2 µL of the extraction was used in subsequent PCRs.

Table 4.1. Collection information for mealybug specimens used during optimization and testing of the multiplex PCR.

Species	Location	Collection	Host plant	Number of specimens[1]
P. ficus	Paarl, Western Cape	vineyard	*Vitis vinifera* (grapevine)	20
	Stellenbosch, Western Cape	vineyards	*Vitis vinifera* (grapevine)	115
	Pretoria, Gauteng	lab colony	*Cucurbita maschata* (butternut)	8
P. citri	Stellenbosch, Western Cape	farm	*Ipomoea batatas* (sweet potato)	6
	ARC-PPRI [2], Gauteng	lab colony	*Cucurbita maschata* (butternut)	19
P. longispinus	Stellenbosch, Western Cape	vineyards	*Vitis vinifera* (grapevine)	9
	Paarl, Western Cape	vineyard	*Vitis vinifera* (grapevine)	1
	Stellenbosch, Western Cape	garden	*Agapanthus praecox* (African lily)	1
	UP[3], Gauteng	lab colony	*Alocasia macrorrhizos* (elephant ear)	16
P. viburni	Stellenbosch, Western Cape	vineyard	*Sonchus oleraceus* (sow thistle)	4
	SUN[4], Western Cape	lab colony	*Cucurbita maschata* (butternut)	15
	Pretoria, Gauteng	garden	*Salvia* sp. (sage)	3

[1] Sample size was limited by availability of mealybugs
[2] Agricultural Research Council, Plant Protection Research Institute
[3] University of Pretoria
[4] Stellenbosch University

PCR amplification

The universal primers C1-J-2183 (alias Jerry) and TL2-N-3014 (alias Pat) (Simon *et al.* 1994) were used to amplify 831 bp of the mitochondrial cytochrome *c* oxidase subunit 1 (CO I) gene. PCR amplification was performed in a 25 µL reaction mix consisting of NH_4 Reaction Buffer (final concentration: 16 mM $(NH_4)_2SO_4$, 67 mM Tris-HCl, 0.01% Tween-20), 4 mM $MgCl_2$, 50 µM each dNTP, 0.4 µM each primer, 0.05 units/µL BIOTAQ™ DNA Polymerase (BioLine, Luckenwalde, Germany) and 1-2 µL extracted DNA.

Thermocycling conditions consisted of an initial denaturation at 94ºC for 1 min, then 35 cycles of 94ºC for 45 sec, annealing at 48ºC for 45 sec and extension at 72ºC for 1 min, with final extension at 72ºC for 3 min. PCR products were visualized under UV light on a 1.5% agarose gel stained with ethidium bromide (EtBr).

DNA sequencing and analyses

Products were cycle sequenced with the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), using ¼ of the manufacturer's recommended reaction mix. Products were then precipitated and sequenced on an ABI3100 PRISIM™ Genetic Analyzer (Applied Biosystems) at the DNA Sequencing Facility, University of Pretoria. DNA sequences thus obtained were edited by eye in Chromas Lite 2.00 (Technelysium Pty. Ltd.), then aligned within and between species using Clustal X (Thompson *et al.* 1994 and 1997, Jeanmougin *et al.* 1998). A consensus sequence was generated for each species, using data from both the forward and reverse primers.

Species-specific primer design

Using the alignment generated by Clustal X, three species-specific forward primers were designed in regions where the mealybug sequences differed. These primers were designed such that they could be matched with the universal reverse primer TL2-N-3014. Primers were designed to cover at least 3 bp changes between species, have a unique C- or G-base at the 3'-end, and yield products that could easily be separated on an agarose gel. Primers were named according to the convention of Simon *et al.* (1994). Primer names, sequences and other selected parameters are given in Table 4.2.

Multiplex PCR

The multiplex PCR reaction mix was essentially the same as that used for the universal primers, with the exception that primer C1-J-2427 was used at the slightly higher concentration of 0.5 µM. All other primer concentrations were at 0.4 µM. Thermocycling conditions consisted of an initial denaturation at 94ºC for 2 min, then 35 cycles of 94ºC for 40 sec, 64ºC for 40 sec and 72ºC for 50 sec, with a final extension at 72ºC for 2 min. Samples were cooled to at least 10ºC before being removed and stored at 4ºC until needed. Products from the multiplex PCR were visualized under UV light on a 2.5% agarose gel stained with EtBr.

Multiplex PCR controls

All PCRs were run with an external negative control, which contained all reagents except DNA. During optimization of the multiplex PCR, specimens of known species were used. In further application of the multiplex PCR, a known specimen of each species was used as a positive control. The species-specific primers were tested for cross-reactivity, both separately and in the multiplex reaction mix.

Testing and application of the multiplex PCR

The accuracy of the multiplex PCR was tested using mealybugs of known origin in blind trials (n = 30). Specificity was tested using specimens of the obscure mealybug *Pseudococcus viburni* (n = 22, see Table 4.1). The multiplex PCR was used to identify field-collected mealybugs of all three species (n = 140).

Table 4.2. Details and thermodynamic properties of species-specific primers designed for use in the multiplex PCR.

Name [1]	Specificity	Sequence (5'-3')	Primer Length	Melting temperature	Product length [2]
C1-J-2260	*P. ficus*	TCAAATTATAAATCAAGAAAGGGGAAAAC	29	59.0ºC	754 bp
C1-J-2427	*P. citri*	TAATTATTGCTATTCCTACAAGAATTAAAATC	32	58.5ºC	587 bp
C1-J-2608	*P. longispinus*	TTTGTTGTAGCACATTTTCATTATGTAC	28	58.8ºC	406 bp

[1] Primers are named according to the convention of Simon *et al.* (1994).
[2] Product length is when used in conjunction with TL2-N-3014 (Simon *et al.* 1994).

Results

DNA sequence data

In the PCR, a single fragment of approximately 830 bp was amplified. Sequencing of this fragment yielded usable nucleotide sequence of approximately 760 bp. Intraspecific variation in this region was very low (<1%), and a consensus sequence was therefore easily generated. A pair-wise comparison of the nucleotide sequence from the three mealybug species showed *P. ficus* and *P. citri* to be more similar (92.5% identity) than either *P. ficus* and *P. longispinus* (89.1% identity) or *P. citri* and *P. longispinus* (88.6% identity) (appendix 3). These sequences mapped onto the 3'-end of the mitochondrial CO I gene from positions 2216 to 2986 of the *Drosophila yakuba* Burla (Diptera: Drosophilidae) mitochondrial genome (Clary and Wolstenholme 1983 and 1985). BLAST searches of these sequences on Genbank revealed a high similarity (92-100%) to other mealybug sequences from the CO I gene (appendix 5). The consensus sequences from each of the three mealybug species were submitted to Genbank: *P. ficus* accession number DQ238220; *P. citri* accession number DQ238221; and *P. longispinus* accession number DQ238222.

Multiplex PCR and primers

When used in conjunction with primer TL2-N-3014, the species-specific primers yielded amplicons of 754, 587 and 406 bp from *P. ficus*, *P. citri* and *P. longispinus*, respectively (see Table 4.2 and Figure 4.1). Once optimized at the conditions stated, these primers consistently yielded single amplicons of the specified size from each species. Occasionally, an extra non-specific band was observed in the PCR product. However, this band was always lighter than the specific one and did not occur at the same position as that from one of the other primers. It therefore never interfered with interpretation of the results.

No cross-reactivity of primers was observed at optimum conditions. When each species-specific primer was tested separately for reactivity with other species, none occurred when reactions were near optimum (T_a = 64 ± 2ºC). Within the multiplex reaction mix, cross-reactivity of primers close to the optimum annealing temperature

(64 ± 1ºC) was never observed. When the multiplex PCR was tested on *P. viburni* no amplification product was obtained.

Blind trials with 30 specimens were set up to test the multiplex PCR. Twenty nine specimens amplified and all were correctly identified (Figure 4.2). The specimen which did not amplify was subsequently re-tested using both the multiplex primers and the conserved CO I primers (C-J-2183 and TL2-N-3014). No amplicon could be obtained.

Application of the multiplex PCR

The multiplex PCR was used to identify field-collected mealybugs. These mealybugs had been stored in absolute ethanol at ambient temperature for approximately five months before being transferred to -20ºC. Many specimens were very small or heavily parasitized. However, DNA was extracted and amplified from 87% (122 of 140) of individual mealybugs analyzed. The identification of these specimens was in agreement with a sub-sample sent for morphological identification.

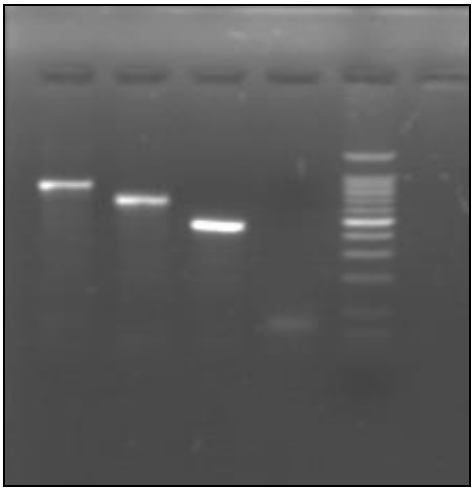

Figure 4.1. Multiplex PCR amplification of mealybug DNA. Lane 1: *P. ficus*; lane 2: *P. citri*; lane 3: *P. longispinus*; lane 4: negative control; lane 5: DNA size marker.

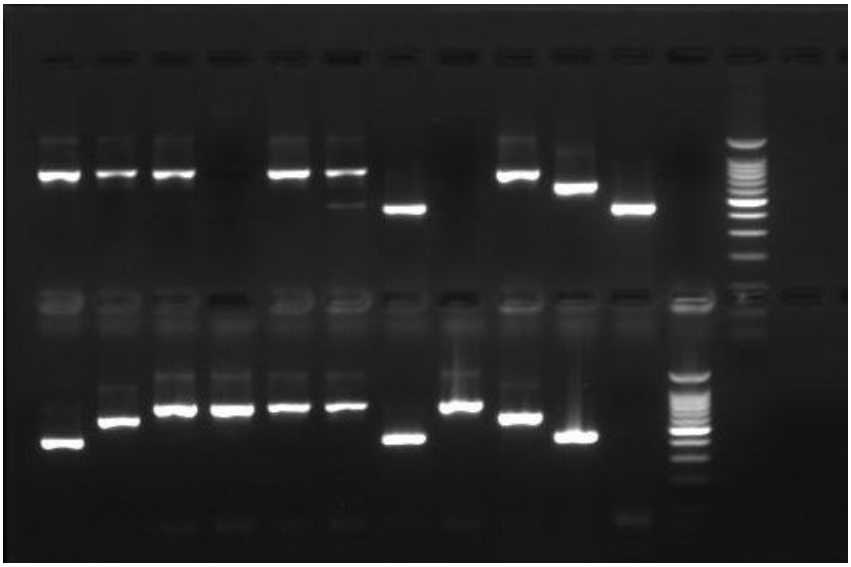

Figure 4.2. Multiplex PCR amplification of mealybug DNA to identify unknown specimens. The last five lanes in each row are PCR controls (*P. ficus*, *P. citri*, and *P. longispinus*) and DNA size marker. Row 1, lanes 1-3 and 5-6: *P. ficus*; lane 7: *P. longispinus*; lanes 4 and 8: unamplified specimens. Row 2, lanes 1 and 7: *P. longispinus*, lane 2: *P. citri*, lanes 3-6: *P. ficus*.

Discussion

In this study, a multiplex PCR has been developed for the identification of three species of mealybug associated with grapevine. When combined with the direct buffer extraction method, the entire protocol took approximately four hours to complete. Due to the nature of PCR, a large number of samples can be run concurrently, thereby further shortening over-all processing time. The method reliably extracted and amplified DNA even from small and damaged specimens. In the blind trials, every specimen that amplified was correctly identified.

However, it should always be borne in mind that no identification method is infallible. Sources of error in any method should be identified, eliminated if possible, and incorporated into further decision-making if not. Possible sources of error are primer-bias, false negatives due to high intraspecific variation and false positives due to low interspecific variation. These sources of error have been addressed in this study by optimization and thorough testing of the multiplex PCR. Although primer-bias was

initially observed, a change in relative primer concentrations eliminated all cross-reactivity. This multiplex PCR is unlikely to yield false negative results for one of the three species in question, due to the conserved nature of the CO I gene used in this study and the observed lack of intraspecific variation, even between different populations. Likewise, false positives are unlikely to occur, since this multiplex has been tested with five different mealybug species and has shown no cross-reactivity. In addition, *P. ficus* and *P. citri*, and *P. longispinus* and *P. viburni*, respectively, are very closely related to each other (Ben-Dov and German 2003, Downie and Gullan 2004). Thus the testing of each of these mealybugs pairs serves as a control. Since the species-specific primers showed no cross-reactivity to these most closely related species, they are unlikely to show cross-reactivity to any more distantly-related mealybugs. The present study, however, did not prove this and it still remains to be tested.

The accuracy and reliability of the technique is comparable to that obtained with other molecular identification studies. For example, the method developed by Beuning *et al.* (1999) for mealybug identification (employing primers in separate reactions) correctly identified 23 of 24 (96%) specimens in a blind trial. The last specimen was incorrectly identified as a different species within the species group. The multiplex PCR developed by Koekemoer *et al.* (2002) amplified 868 of 900 mosquito specimens when tested by Weeto *et al.* (2004). All amplified specimens were correctly identified. The multiplex developed by Gariepy *et al.* (2005) correctly identified all voucher specimens (20 of 20) and was sensitive enough to detect parasitoid eggs in the host three days post-parasitism. In comparison, in this study the multiplex PCR correctly identified all amplified specimens (29 of 30 (97%)) specimens in blind trials. The last specimen did not amplify. When the multiplex PCR was applied to field-collected samples, it performed extremely well, amplifying and correctly identifying 122 of 140 (87%) specimens, despite the fact that they had been stored in alcohol for over five months at ambient temperature. Those from which DNA could not be extracted or amplified were usually parasitized individuals, tiny first instar nymphs or eggs from which DNA could not be extracted. Over-all, the multiplex PCR was able to amplify 90% (192 of 213) of specimens, regardless of size, sex or condition of the specimen. All amplified specimens were correctly identified.

The speed, reliability, sensitivity and accuracy of this technique represents a considerable improvement over currently available identification protocols. In addition, the simple procedure can be implemented in any molecular laboratory and does not require extensive taxonomic or molecular experience. This identification protocol will aid in accurately and rapidly identifying mealybug specimens found on export consignments and in vineyards. Timely and accurate identification is essential to management of these pest species. In addition, it is envisioned that this multiplex PCR will be implemented in scientific research to provide accurate species identification for biological and ecological studies on mealybugs.

CHAPTER FIVE

Conclusion

Chapter Five
Conclusion

Background

This study was initiated as part of a large-scale project on the vectors of grapevine leafroll viruses in South African vineyards. As part of that project, three mealybug (Hemiptera: Pseudococcidae) species associated with grapevine in South Africa were examined: the vine mealybug *Planococcus ficus* (Signoret), the citrus mealybug *Planococcus citri* (Risso) and the longtailed mealybug *Pseudococcus longispinus* (Targioni-Tozzetti). These species have been recorded on grapevine in South Africa and in other grape-growing areas of the world, and have been shown to transmit grapevine leafroll viruses.

However, control of mealybugs in commercial vineyards is hampered by difficulties in accurate identification. Currently, only undamaged adult females can be assigned to species. Damaged specimens, nymphs and males cannot be identified. Therefore, the purpose of this research was to develop an identification technique suitable for accurate and rapid identification of these three mealybug species, whether encountered as adults or nymphs. This was of particular concern for the identification of *P. ficus* and *P. citri*, as they are very similar and are frequently confused. In this study a DNA-based technique was developed for mealybug identification which is rapid, simple, accurate, sensitive and reliable, and can be implemented routinely for identification of large sample sizes.

Development of a DNA-based identification technique

In order to develop a reliable DNA-based identification method for mealybugs, a number of tasks were undertaken.

DNA extraction

Two rapid and reliable DNA extraction methods were developed and tested for diagnostic use. The direct buffer method could be completed in 15 – 20 min, and the spot-PCR method took approximately 3 hrs. Both methods extracted DNA from most specimens tested (96% and 100% for direct buffer and spot-PCR with STE buffer, respectively). This is a considerable improvement over the traditional extraction methods tested, which took 1-2 days and extracted DNA from a maximum of 78% (39 of 50) of samples.

Nucleotide sequence data

A gene region of suitable variability was sought. This involved amplification and sequencing of two regions of the mitochondrial cytochrome *c* oxidase subunit 1 (CO I) gene. The 3'-region of this gene yielded low intraspecific variability, but sufficient interspecific variability to clearly and accurately delineate species.

Multiplex PCR

Using the nucleotide sequence data obtained, a multiplex PCR was designed for the simultaneous identification of the three mealybugs species. Species-specific primers were designed and used together in a single reaction mix to differentially amplify DNA from each of the three mealybug species. The species-specific DNA products were each of a different length, and could thus be separated by electrophoresis on an agarose gel, allowing easy identification of the species.

The result of these methods is an accurate identification of each of the three mealybug species within approximately four hours. The method is simple, and can be implemented in any molecular laboratory with basic facilities for heating, thermal cycling and agarose gel electrophoresis. The sensitivity, accuracy and reliability of the technique is high, allowing correct identification of adult, damaged or immature specimens of both sexes. This method has the potential to be implemented as a commercial diagnostic technique.

This multiplex PCR allows for quicker and more accurate identification of the mealybug species associated with grapevine in South Africa, permitting more prompt

and accurate decisions regarding pest monitoring and control, and quarantine of imported or exported products. This simple, accurate identification will aid in scientific research on mealybugs and mis-identifications will hopefully become a thing of the past.

Application of the multiplex PCR

Although the multiplex PCR has been used to test field-collected samples, it has yet to be implemented commercially as a diagnostic procedure. Application of the multiplex PCR in different laboratories and to samples collected from different areas under different conditions will test the robustness of the PCR. As the multiplex PCR has already been used to test a large number of samples (197 individual mealybugs) from widely-spaced locations, it is improbable that any difficulties will arise during the routine use of this method for species identification.

This multiplex PCR has been designed in such a way that it can be expanded to include other mealybug species. It could then be used in other crop systems to identify mealybugs of importance. If new mealybug species are added to the analyses, sequence data would have to be obtained for each new species, and species-specific primers designed. These primers would have to be similar to the present ones in all thermodynamic properties, and each yield DNA product of a different length. Theoretically, given the size of the CO I gene (approx. 1500 bp), up to 15 mealybug species with DNA products differing by 100 bp could be included in the analysis. In practice, however, the number is likely to be less than this due to the difficulty of designing suitable primers in regions of high enough variability.

The principles and procedures used in development of the multiplex PCR in this study can be applied to any species group. Thus, using the same principles, a multiplex PCR can be developed for identification of any species. Multiplex PCR has seldom been used for insect identification, and the possibilities of this method are numerous. As a diagnostic tool, multiplex PCR would be a great asset wherever implemented. For instance, management of insect pests on commercial crops would benefit from fast, accurate identification. The use of a single-step multiplex PCR could also be used in

molecular identification of medically or forensically important species, allowing rapid action to be taken.

Future research

During the course of this study, a number of interesting questions have arisen. Although investigation of these questions is beyond the scope of this project, they are discussed briefly here to provide an outline for future research in these areas.

Mis-priming in *P. longispinus*

When amplifying *P. longispinus* mtDNA from the mid-section of CO I, it was found that primers C1-J-1718 and C1-N-2191 could not amplify a DNA product, and that primer C1-J-1718 did not anneal when used in conjunction with TL2-N-3014. This is very interesting, as it may indicate a change in nucleotide sequence at the primer annealing site. This would be unusual, as the primers are universal, and are designed to anneal to conserved regions of the genome. A change in the primer site may have implications for studies of mealybugs in the areas of phylogenetics, evolution and functionality, and would be interesting to investigate.

AT-bias

The mealybug mitochondrial sequences obtained in this study were strongly AT-biased, and presented problems during PCR amplification and sequencing. Although this has been noted in mitochondrial DNA before, few suggestions for overcoming the problems associated with AT-bias have been made. Su *et al.* (1996) showed that a reduced extension temperature was necessary in the PCR cycle when amplifying extremely AT-rich DNA. They suggested an extension temperature no higher than 60ºC. Brehm *et al.* (2001) used a PCR cycle profile with an extension temperature of 68ºC and a modified nucleotide mix for sequencing that contained higher amounts of dTTP and dATP relative to dCTP and dGTP when amplifying and sequencing the control region of *Drosophila subobscura* (93% AT-bias). No other suggestions or modifications have been made when working with AT-biased genomes.

In the present study, when it was attempted to amplify the mid-CO I DNA fragment using a reduced extension temperature, the amplification failed. Therefore a reduced extension temperature cannot be the only solution to the problems associated with amplification of a strongly AT-biased genome. It would be interesting to investigate the outcome of various modifications when amplifying and sequencing a strongly AT-biased DNA segment, such as the insect mitochondrial control region. A study such as this could suggest and test protocols which would enhance the efficiency with which AT-biased DNA can be amplified and sequenced.

Conclusion

This study aimed to develop a molecular identification technique for three species of mealybug that was rapid, accurate, sensitive, reliable and simple. In order to accomplish this objective, rapid and reliable DNA extraction methods were tested, mtDNA from the CO I gene was amplified, sequenced and analysed, species-specific forward primers were designed for each of the three species, and these were used in a multiplex PCR for the simultaneous identification of the three mealybug species.

Thus the identification protocol developed in this study includes rapid and reliable DNA extraction techniques which are easily carried out on a large scale. The molecular identification itself consists of a multiplex PCR, using species-specific primers to amplify DNA from each of the three mealybug species. The method is accurate, sensitive and reliable, correctly identifying all amplified specimens, including small, damaged and degraded specimens. The entire protocol (including extraction, PCR and electrophoresis) can be completed in approximately four hours, and is simple enough to be implemented in any molecular laboratory by anyone with basic training in molecular methods. This technique represents a substantial improvement over currently available methods, and will aid in rapid and accurate identification of mealybugs occurring in South African vineyards.

CHAPTER SIX

References

Abu Al-Soud, W., and P. Rådström. 2001. Purification and characterization of PCR-inhibitory components in blood cells. Journal of Clinical Microbiology. 39: 485-493.

Alkowni, R., A. Rowhani, S. Daubert, and D. Golino. 2004. Partial characterization of a new Ampelovirus associated with Grapevine Leafroll Disease. Journal of Plant Pathology. 86: 123-133.

Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25: 3389-3402.

Anderson, P.S., C. Jespersgaard, J. Vuust, M. Christiansen, and L. Allan-Larsen. 2003. Capillary electrophoresis-based single strand DNA conformation analysis in high-throughput mutation screening. Human Mutation. 21: 455-465.

Annecke, D.P., and V.C. Moran. 1982. Insects and mites of cultivated plants in South Africa. Butterworth & Co., Pretoria.

Atkinson, L., and E.S. Adams. 1997. Double-strand conformation polymorphism (DSCP) analysis of the mitochondrial control region generates highly variable markers for population studies in a social insect. Insect Molecular Biology. 6: 369-376.

Avise, J.C. 1991. Ten unorthodox perspectives on evolution prompted by comparative population genetic findings of mitochondrial DNA. Annual Review of Genetics. 25: 45-69.

Belli, G., A. Fortusini, P. Casati, L. Belli, P.A. Bianco, and S. Prati. 1994. Transmission of a grapevine leafroll associated closterovirus by the scale insect *Pulvinaria vitis* L. Rivista de Patologia Vegetale. 4: 105-108.

Ben-Dov, Y., and V. German. 2003. Pseudococcidae. In Y. Ben-Dov, D.R. Miller and G.A.P. Gibson (eds.), ScaleNet: A Database of the Scale Insects of the World. http://www.sel.barc.usda.gov/scalenet/scalenet.htm.

Benecke, M. 1998. Random amplified polymorphic DNA (RAPD) typing of necrophageous insects (Diptera: Coleoptera) in criminal forensic studies: validation and use in practice. Forensic Science International. 98: 157-168.

Bermingham, N., and K. Luettich. 2003. Polymerase chain reaction and its applications. Currrent Diagnostic Pathology. 9: 159-164.

Bertamini, M., K. Muthuchelian, and N. Nedunchezhian. 2004. Effect of grapevine leafroll on the photosynthesis of field grown grapevine plants (*Vitis vinifera* L. cv. Lagrein). Journal of Phytophathology. 152: 145-152.

Besansky, N.J., D.W. Severson, and M.T. Ferdig. 2003. DNA barcoding of parasites and invertebrate disease vectors: what you don't know can hurt you. Trends in Parasitology. 19: 545-546.

Bettiga, L. 2002. The vine mealybug spreads to coastal production areas. Central Coast Vineyard News, United States Department of Agriculture, University of California, California.

Bettinaglio, P., A. Galbusera, J. Caprioli, S. Orisio, A. Perna, F. Arnoldi, S. Bucchioni, and M. Noris. 2002. Single strand conformation polymorphism (SSCP) as a quick and reliable method to genotype M235T polymorphism of angiotensinogen gene. Clinical Biochemistry. 35: 363-368.

Beuning, L.L., P. Murphy, E. Wu, T.A. Batchelor, and B.A.M Morris. 1999. Molecular-based approach to the differentiation of mealybug (Hemiptera: Pseudococcidae) species. Journal of Economic Entomology. 92: 463-473.

Birky, C.W., Jr., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics. 103: 513-527.

Brehm, A., D.J. Harris, M. Hernandez, V.M. Cabrera, J.M. Larruga, F.M. Pinto, and A.M. González. 2001. Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. Insect Molecular Biology. 10: 573-578.

Breslauer, K.J., R. Frank, H. Blöcker, and L.A. Marky. 1986. Predicting DNA duplex stability from the base sequence. Proceedings of the National Academy of Sciences USA. 83: 3746-3750.

Brower, A.V.Z., and R. DeSalle. 1994. Practical and theoretical considerations for a choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Annuals of the Entomological Society of America. 87: 702-716.

Brown, W.M., M.George, Jr., and A.C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences USA. 76: 1967-1971.

Brown, J.M., J.H. Leebens-Mack, J.N. Thompson, O. Pellmyr, and R.G. Harrison. 1997. Phylogeography and host association in a pollinating seed parasite *Greya politella* (Lepidoptera: Prodoxidae). Molecular Ecology. 6: 215-224.

Cabaleiro, C., and A. Segura. 1997a. Field transmission of grapevine leafroll associated virus 3 (GLRaV-3) by the mealybug *Planococcus citri*. Plant Disease. 8: 283-287.

Cabaleiro, C., and A. Segura. 1997b. Some characteristics of the transmission of grapevine leafroll associated virus 3 by *Planococcus citri* Risso. European Journal of Plant Pathology. 103: 373-378.

Cabaleiro, C., A. Segura, and J.J. García-Berrios. 1999. Effects of grapevine leafroll-associated virus 3 on the physiology and must of *Vitis vinifera* L. cv. Albariño following contamination in the field. American Journal of Enology and Viticulture. 50: 40-44.

Caterino, M.S., S. Cho, and F.A.H. Sperling. 2000. The current state of insect molecular systematics: a thriving tower of Babel. Annual Review of Entomology. 45: 1-54.

Chansiri, K., P. Kwoasak, C. Tananyutthawongese, W. Sukhumsirichart, N. Sarataphan, and S. Phantana. 2001. Detection of *Plasmodium falciparum* and *Wuchereria bancrofti* infected blood samples using multiplex PCR. Molecular and Cellular Probes. 15: 201-207.

Clary, D.O., and D.R. Wolstenholme. 1983. Genes for cytochrome *c* oxidase subunit I, URF2, and three tRNAs in *Drosophila* mitochondrial DNA. Nucleic Acids Research. 11: 6859-6872.

Clary, D.O., and D.R. Wolstenholme. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. Journal of Molecular Evolution. 22: 252-271.

Cook, L.G., P.J. Gullan, and H.E. Trueman. 2002. A preliminary phylogeny of the scale insects (Hemiptera: Sternorrhyncha: Coccidae) based on nuclear small-subunit ribosomal DNA. Molecular Phylogenetics and Evolution. 25: 43-52.

Crozier, R.H., and Y.C. Crozier. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics. 133: 97-117.

Csaikl, U.M., H. Bastian, R. Brettschneider, S. Gauch, A. Meir, M. Schauerte, F. Scholz, C. Sperisen, B. Vornam, and B. Ziegenhagen. 1998. Comparitive analysis of different DNA extraction protocols: a fast, universal maxi-preparation of high quality plant DNA for genetic evaluation and phylogenetic studies. Plant Molecular Biology Reporter. 16: 69-86.

De Albuquerque, D.M., and S.C.B. Costa. 2003. Genotyping of human cytomegalovirus using non-radioactive single-strand conformation polymorphism (SSCP) analysis. Journal of Virological Methods. 110: 25-28.

De Gruijter, J.M., A.M. Polderman, X.Q. Zhu, and R.B. Gasser. 2002. Screening for haplotypic variability within *Oesophagostomum bifurcum* (Nematoda) employing a

single-stand conformation polymorphism approach. Molecular and Cellular Probes. 16: 185-190.

De Lotto, G. 1975. Notes on the vine mealybug (Homoptera: Coccoidea: Pseudococcidae). Journal of the Entomological Society of South Africa. 38: 125-130.

De Souza Menezes, J., D. De Almeida Drummond Franklin, H. Seki, and F.D. Rumjanek. 2003. Single-strand conformation polymorphism of hyper-variable regions HV1 and HV2 of human mitochondrial DNA: detection by silver staining. Forensic Science International. 133: 242-245.

Dovas, C.I., and N.I. Katis. 2003. A spot multiplex nested RT-PCR for the simultaneous and genetic detection of viruses involved in the aetiology of grapevine leafroll and rugrose wood of grapevine. Journal of Virological Methods. 109: 217-226.

Downie, D.A., and P.J. Gullan. 2004. Phylogenetic analysis of mealybugs (Hemiptera: Coccoicea: Pseudococcidae) based on DNA sequences from three nuclear genes, and a review of the higher classification. Systematic Entomology. 29: 238-259.

Ebbing, D.D. 1996. General Chemistry, fifth edition. Houghton Mifflin Company, Boston.

El-Minshawy, A.M., H.H. Karam, and S.K. El-Sawaf. 1974. Biological studies on the long tailed mealybug, *Pseudococcus longispinus* (Targ. and Tozzeti). Bulletin of the Society of Entomology Egypte. 58: 385-391.

Engelbrecht, D.J., and G.G.F. Kasdorf. 1985. Association of a closterovirus with grapevines indexing positive for grapevine leafroll disease and evidence for its natural spread in grapevine. Phytopathologia Mediterranea. 24: 101-105.

Engelbrecht, D.J., and G.G.F. Kasdorf. 1990. Transmission of grapevine leafroll disease and associated closteroviruses by the vine mealybug, *Planococcus ficus*. Phytophylactica. 22: 341-346.

Flavell, A.J., and G.C. Barr. 1990. DNA sequence information and transmission, p. 165-214. In G.M. Blackburn, and M.J. Gait (eds), Nucleic Acids in Chemistry and Biology. Oxford University Press, New York.

Fukatsu, T. 1999. Acetone preservation: a practical technique for molecular analysis. Molecular Ecology. 8: 1935-1945.

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

Gariepy, T.D., U. Kuhlmann, T. Haye, C. Gillott, and M. Erlandson. 2005. A single-step multiplex PCR assay for the detection of European *Peristenus* spp., parasitoids of *Lygus* spp. Biocontrol Science and Technology. 15: 481-495.

Gasser, R.B., and N.B. Chilton. 2001. Applications of single-strand conformation polymorphism (SSCP) to taxonomy, diagnosis, population genetics and molecular evolution of parasitic nematodes. Veterinary Parasitology. 101: 201-213.

Goheen, A.C., and J.A. Cook. 1959. Leafroll (ref-leaf or rougeau) and its effects on vine growth, fruit quality, and yields. American Journal of Enology and Viticulture. 10: 173-181.

Golino, D.A., S.T. Sim, R. Gill, and A. Rowhani. 2002. California mealybugs can spread grapevine leafroll disease. California Agriculture 56: 196-201.

Gullan, P.J. 2000. Identification of the immature instars of mealybugs (Hemiptera: Pseudococcidae) found on citrus in Australia. Australian Journal of Entomology. 39: 160-166.

Gullan, P.J., and M. Kosztarab. 1997. Adaptations in scale insects. Annual Review of Entomology. 42: 23-50.

Gullan, P.J., D.A. Downie, and S.A. Steffan. 2003. A new pest species of the mealybug genus *Ferrisia* Fullaway (Hemiptera: Pseudococcidae) from the United States. Annuals of the Entomological Society of America. 96: 723-737.

Harvey, M.L., I.R. Dadour, and S. Gaudieri. 2003a. Mitochondrial DNA cytochrome oxidase I gene: potential for distinction between immature stages of some forensically important fly species (Diptera) in western Australia. Forensic Science International. 131: 134-139.

Harvey, M.L., M.W. Mansell, M.H. Villet, and I.R. Dadour. 2003b. Molecular identification of some forensically important blowflies of southern Africa and Australia. Medical and Veterinary Entomology. 17: 363-369.

Hatting, V., C.J. Cillier, and E.C.G. Bedford. 1998. Family Pseudococcidae: mealybugs, p. 112-273. In E.C.G. Bedford, M.A. Van den Berg, and E.A. De Villiers (eds), Citrus pests in the Republic of South Africa (2nd edition). Institute for Tropical and Subtropical Crops, Nelspruit.

Hebert, P.D.N., A. Cywinska, S.L. Ball, and J.R. DeWaard. 2003a. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London, series B. 270: 313-321.

Hebert, P.D.N., S. Ratnasingham, and J.R. DeWaard. 2003b. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London, series B. 270: S96-S99.

Hillis, D.M., C. Moritz, and B.K. Mabel. 1996. Molecular Systematics. Sunderland, Sinauer.

Howe, C.J., and E.S. Ward. 1991. DNA sequencing, p. 157-183. In T.A. Brown (ed.), Essential Molecular Biology: A Practical Approach, vol. 2. Oxford University Press, NewYork.

Hwang, U.-W, and W. Kim. 1999. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. The Korean Journal of Parasitology. 37: 215-228.

Jeanmougin, F., J.D. Thompson, M. Gouy, D.G. Higgins, and T.J. Gibson. 1998. Multiple sequence alignment with Clustal X. Trends in Biochemical Sciences. 23: 403-405.

Kengne, P., H.D. Trung, V. Baimai, M. Coosemans, and S. Manguin. 2001. A multiplex PCR-based method derived from random amplified polymorphic DNA (RAPD) markers for the identification of species of the *Anopheles minimus* group in Southeast Asia. Insect Molecular Biology. 10: 427-435.

Kho, W.-G., J.-Y. Chung, E.-J. Sim, M.-Y. Kim, D.-W. Kim, S. Jongwutiwas, and K. Tanabe. 2003. A multiplex polymerase chain reaction for a differential diagnosis of *Plasmodium falciparum* and *Plasmodium vivax*. Parasitology International. 52: 229-236.

Kocher, T.D., W.K. Thomas, A. Meyer, S.V. Edwards, S. Pääbo, F.X. Villablanca, and A.C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proceedings of the National Academy of Sciences USA. 86: 6196-6200.

Kocher, T.D., and A.C. Wilson. 1991. DNA amplification by the polymerase chain reaction, p. 185-207. In T.A. Brown (ed.), Essential Molecular Biology: A Practical Approach, vol. 2. Oxford University Press, New York.

Koekemoer, L.L., K. Hargreaves, R.H. Hunt, and M. Coetzee. 2002a. Identification of *Anopheles parensis* (Diptera: Culicidae) using ribosomal DNA internal transcribed spacer (ITS2) sequence variation. African Entomology. 10: 235-239.

Koekemoer, L.L., L. Kamau, R.H. Hunt, and M. Coetzee. 2002b. A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. American Journal of Tropical Medicine and Hygiene. 66: 804-811.

Kong, P., C. Hong, P.A. Richardson, and M.E. Gallegly. 2003. Single-strand-conformation polymorphism of ribosomal DNA for rapid species differentiation in genus *Phytophthora*. Fungal Genetics and Biology. 39: 238-249.

Krake, L.R., N.S. Scott, M.A. Rezaian, and R.H. Taylor. 1999. Graft-transmitted Diseases of Grapevines. CSIRO Publishing, Ollingwood.

Kwon, H.-S., E.-H. Yang, S.-W. Yeon, B.-H. Kang, and T.-Y. Kim. 2004. Rapid identification of probiotic *Lactobacillus* species by multiplex PCR using species-specific primers based on the region extending from 16S rRNA through 23S rRNA. Federations of European Microbiological Societies Microbiology Letters. 239: 267-275.

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

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

Lin, C.-P., and B.N. Danforth. 2004. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Molecular Phylogenetics and Evolution. 30: 686-702.

Liu, C., Y. Song, M. McTeague, A.W. Vu, H. Wexler, and S.M. Finegold. 2003. Rapid identification of the species of the *Bacteroides fragilis* group by multiplex PCR assays using group- and species-specific primers. Federations of European Microbiological Societies Microbiology letters. 222: 9-16.

Löffler, J., H. Hebart, U. Schumacher, H. Reitze, and H. Einsele. 1997. Comparison of different methods for extraction of DNA of fungal pathogens from cultures and blood. Journal of Clinical Microbiology. 35: 3311-3312.

Malgorn, Y., and R. Coquoz. 1999. DNA typing for identification of some species of Calliphoridae, an interest in forensic entomology. Forensic Science International. 102: 111-119.

Marcilla, A., M.D. Bargues, J.M. Ramsey, E. Magallon-Gastelum, P.M. Salazar-Schettino, F. Abad-Franch, J.-P. Dujardin, C.J. Schofield, and S. Mas-Coma. 2001. The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas Disease. Molecular Phylogenetics and Evolution. 18: 136-142.

Marfurt, J., I. Niederwieser, N.D. Makia, H.-P. Beck, and I. Felger. 2003. Diagnostic genotyping of old and new world *Leishmania* species by PCR-RFLP. Diagnostic Microbiology and Infectious Disease. 46: 115-124.

Markino, R., K. Kaneko, T. Kurahashi, T. Matsumura, and K. Mitamura. 2000. Detection of mutation of the *p53* gene with high sensitivity by fluorescence-based PCR-SSCP analysis using low-pH buffer and an automated DNA sequencer in a large number of DNA samples. Mutation Research. 452: 83-90.

Martelli, G.P. 1986. Virus and virus-like diseases of the grapevine in the Mediterranean area. FOA Plant Protection Bulletin. 34: 25-42.

Martelli, G.P., A.A. Agranovsky, M. Bar-Joseph, D. Boscia, T. Candresse, R.H.A. Coutts, V.V. Dolja, B.W. Falk, D. Gonsalves, W. Jelkmann, A.V. Karasev, A. Minafra, S. Namba, H.J. Vetten, G.G. Wisler, and N. Yochikawa. 2002. The family *Closteroviridae* revised. Archives of Virology. 147: 2039-2044.

Mercier, B., C. Gaucher, O. Feugeas, and C. Mazurier. 1990. Direct PCR from whole blood, without DNA extraction. Nucleic Acids Research. 18: 5908.

Millar, I.M. 2002. Mealybug genera (Hemiptera: Pseudococcidae) of South Africa: identification and review. African Entomology. 10: 185-233.

Mooers, A.Ø., and E.C. Holmes. The evolution of base composition and phylogenetic inference. Trends in Ecology and Evolution. 15: 365-369.

Moran, V.C. 1983. The phytophagous insects and mites of cultivated plants in South Africa: patterns and pest status. Journal of Applied Ecology. 20: 439-450.

Mullis, K.B., and F.A. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods in Enzymology. 155: 335-350.

Otranto, D., and J.R. Stevens. 2002. Molecular approaches to the study of myiasis-causing larvae. International Journal of Parasitology. 32: 1345-1360.

Owczaray, R., P.M. Vallone, F.J. Gallo, T.M. Paner, M.J. Lane, and A.S. Benight. 1997. Predicting sequence-dependent melting stability of short duplex DNA oligomers. Biopolymers. 44: 217-239.

Petersen, C.L., and J.G. Charles. 1997. Transmission of Grapevine Leafroll-associated Closteroviruses by *Pseudococcus longispinus* and *P. calceolariae*. Plant Pathology. 46:509-515.

Post, R.J., P.K. Flook, and A.L. Millest. 1993. Methods for the preservation of insects for DNA studies. Biochemical Systematics and Ecology. 21: 85-92.

Powell, J.R., A. Caccone, G.D. Amato, and C. Yoon. 1986. Rates of nucleotide substitution in *Drosophila* mitochondrial DNA and nuclear DNA are similar. Proceedings of the National Academy of Sciences USA. 83: 9090-9093.

Ravel, S., N. Monteny, D.V. Olmos, J.E. Verdugo, and G. Cuny. 2001. A preliminary study of the population genetics of *Aedes aegypti* (Diptera: Culicidae) from Mexico using microsatellite and AFLP markers. Acta Tropica. 78: 241-250.

Ren, J. 2000. High-throughput single-stand conformation polymorphism analysis by capillary electrophoresis. Journal of Chromatography B. 741: 115-128.

Roivainen, O. 1980. Mealybugs, p. 15-38. In K.F. Harris, and K. Maramorosch (eds.), Vectors of Plant Pathogens. Academic Press, New York.

Rokas, A., J.A.A. Nylander, F. Ronquist, and G.N. Stone. 2002. A maximum-likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae): implications for insect phylogenetic studies. Molecular Phylogenetics and Evolution. 22: 206-219.

Rosciglione, B., M.A. Castellano, G.P. Martelli, V. Savino, and G. Cannizzaro. 1983. Mealybug transmission of grapevine virus A. Vitis. 22: 331-347.

Saiki, R.K., S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science. 230: 1350-1354.

Saiki, R.K., D.H. Gelfand, S. Stoffel, S.J. Scharf, R. Higuchi, G.T. Horn, K.B. Mullis, and H. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 239: 487-491.

Salas, A., E.M. Rasmussen, M.V. Lareu, N. Morling, and A. Carracedo. 2001. Flourescent SSCP of overlapping fragments (FSSCP-OF): a highly sensitive method of the screening of mitochondrial DNA variation. Forensic Science International. 124: 97-103.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual, 2nd ed, vol. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sambrook, J., and D.W. Russell. 2001. Molecular Cloning: A Laboratory Manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences USA. 74: 5463-5467.

Satta, Y., H. Ishiwa, and S.I. Chigusa. 1987. Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Molecular Biology and Evolution. 4: 638-650.

Sentinelli, F., S. Lovari, M. Vitale, G. Giorgi, U. Di Mario, and M.G. Baroni. 2000. A simple method for non-radioactive PCR-SSCP using MDE™ gel solution and a midi gel format: application for the detection of variants in the GLUT1 and CTLA-4 genes. Journal of Biotechnology. 78: 201-204.

Sforza, R., V. Komar, and C. Grief. 2000. New scale insect vectors of grapevine closteroviruses. Proceedings of the 13th Meeting of the International Council for the study of Viruses and Virus-like Diseases of Grapevine (ICVG). Adelaide, Australia (March): 14-15.

Sforza, R., E. Boudon-Padieu, and C. Greif. 2003. New mealybug species vectoring grapevine leafroll-associated viruses-1 and -3 (GLRaV-1 and -3). European Journal of Plant Pathology. 109: 975-981.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America. 87: 651-701.

Sommer, R., and D. Tautz. 1989. Minimal homology requirements for PCR primers. Nucleic Acids Research. 16: 6749.

Sperling, F.A.H., A.G. Raske, and I.S. Otvoz. 1999. Mitochondrial DNA sequence variation among populations and host races of *Lambdina fiscellaria* (Gn.) (Lepidoptera: Geometridae). Insect Molecular Biology. 8: 97-106.

Steitz, T.A. 1999. DNA polymerases: structural diversity and common mechanisms. The Journal of Biological Chemistry. 274: 17395-17398.

Stevens, J., and R. Wall. 1995. The use of random amplified polymorphic DNA (RAPD) analysis for studies of genetic variation on populations of the blowfly *Lucilia sercata* (Diptera: Calliphoridae) in southern England. Bulletin of Entomological Research. 85: 549-555.

Sunnucks, P., and D.F. Hales. 1996. Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). Molecular Biology and Evolution. 13: 510-524.

Su, X.-Z., Y. Wu, C.D. Sifri, and T.E. Wellems. 1996. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. Nucleic Acids Research. 24: 1574-1575.

Tanaka, H., T. Suka, D.W. Roubik, and M. Mohamed. 2001a. Genetic differentiation among geographic groups of three honey bee species, *Apis cerana*, *A. koschevnikovi* and *A. dorsata*, in Borneo. Bulletin of the Museum of Nature and Human Activities. 6: 5-12.

Tanaka, H., D.W. Roubik, M. Kato, F. Liew, and G. Gunsalam. 2001b. Phylogenetic position of *Apis nuluensis* of northern Borneo and phylogeography of *A. cerana* as inferred from mitochondrial DNA sequences. Insectes Sociaux. 48: 44-51.

Tayutivutikul, J., W. Pongprasert, L.A. Royce, and K. Ruangrit. 2003. Comparison of preservation techniques for silkworm (*Bombyx mori* L.) DNA based on polymerase chain reaction (PCR) products. Chiang Mai University Journal. 2: 107-114.

Thao, M.L., P.J. Gullan, and P. Baumann. 2002. Secondary (*γ-Proteobacteria*) endosymbionts infects the primary (*β-Proteobacteria*) endosymbionts of mealybugs multiple times and coevolve with their hosts. Applied and Environmental Microbiology. 68: 3190-3197.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 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.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 25: 4876-7882.

Trewick, S.A. 2001. Scree weta phylogeography: surviving glaciation and implications for Pleistocene biogeography. New Zealand Journal of Zoology. 28: 291-298.

Vaughan, H.E., J.S. Heslop-Harrison, and G.M. Hewitt. 1999. The localization of mitochondrial sequences to chromosomal DNA in Orthopterans. Genome. 42: 847-880.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research. 23: 4407-4414.

Wallis, G.P. 1999. Do animal mitochondrial genomes recombine? Trends in Ecology and Evolution. 14: 209-210.

Watson, J.D., and F.H.C. Crick. 1953a. Molecular structure of nucleic acids. Nature. 171: 737-738.

Watson, J.D., and F.H.C. Crick. 1953b. Genetical implications of the structure of deoxyribonucleic acid. Nature. 171: 964-969.

Watson, G.W., and J. Kubiriba. 2005. Identification of mealybugs (Hemiptera: Pseudococcidae) on banana and plantain in Africa. African Entomology. 13: 35-47.

Weeto, M.M., L.L. Koekemoer, L. Kamau, R.H. Hunt, and M. Coetzee. 2004. Evaluation of a species-specific PCR assay for the *Anopheles funestus* group from eleven African countries and Madagascar. Transactions of the Royal Society of Tropical Medicine and Hygiene. 98: 142-147.

Wells, J.D., and F.A.H. Sperling. 2001. DNA-based identification of forensically important Chrysomyinae (Diptera: Calliphoridae). Forensic Science International. 120: 110-115.

Wiedbrauk, D.L., J.C. Werner, and A.M. Drevon. 1995. Inhibition of PCR bu aqueous and vitreous fluids. Journal of Clinical Microbiology. 33: 2643-2646.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Ingey, 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research. 18: 6531-6535.

Wolstenholme, D.R., and Clary, D.O. 1985. Sequence evolution of *Drosophila* mitochondrial DNA. Genetics. 109: 725-744.

Wrischnik, L.A., R.G. Higuchi, M. Stoneking, H.A. Erlich, N. Arnheim, and A.C. Wilson. 1987. Length mutations in human mitochondrial DNA: direct sequencing of enzymatically amplified DNA. Nucleic Acids Research. 15: 529-541.

Yoon, J.-H., S.T. Lee, Y.K. Shin, S.-B. Kim, H.-J. Kim, M. Goodfellow, and Y.-H.Park. 1996. Rapid identification of *Saccharomonospora* strains by multiplex PCR using species-specific primers within the 16S rRNA gene. Journal of Microbiological Methods. 27: 89-95.

Zehner, R., J. Amendt, S. Schütt, J. Sauer, R. Krettec, and D. Povolný. 2004. Genetic identification of forensically important flesh flies (Diptera: Sarcophagidae). International Journal of Legal Medicine. 118: 245-247.

Zhang, D.-X., and G.M. Hewitt. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. Trends in Ecology and Evolution. 11: 247-251.

Zhang, D.-X., and G.M. Hewitt. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics and Ecology. 25: 99-120.

APPENDICES

Appendix 1: Mealybug nucleotide sequences

Nucleotide sequence of *Planococcus ficus* from the mid-region of CO I, PCR amplified and sequenced with primers C1-J-1718 and C1-N-2191 (Genbank accession number DQ238218).

```
  ORIGIN
   1    CCACGATTAA ATAATTTSAG ATTTTGATTA TTAATTCCAT CTCTTATTTT AATAATATTA
  61    AATATAATTT TATCTAATAA TATTAATACT GGTTGAACTT TATATCCCCC TTTAATTAAT
 121    CAAAATTTTA TTACATTAAA TTTTATCATT TTTTCTTTAC ATTTAAATGG TATTTCATCA
 181    ATTTTTAGAT CAATCAATTT TATTTCTTCA ATTTTTATTA TTAATAATAA TAATTTTTTT
 241    TTAAATAATA TTACTCTTTA TATTTGATCT ATTATTATCA CAACTATTTT ATTAATTATT
 301    TCTATTCCTA TTTTATCTAG AGCTATTACT ATAATTATTT TAGATAAWAA TCTTAATATA
 361    AATTTTTTTA ATCCCTTAGG AAATGGTAAC CCAWTTTTAT ACCAACATTT ATTTTGATTT
 421    TTTGGGCATC CAGAAGTTTA TATTTTAATT TTACCGGG
```

Nucleotide sequence of *Planococcus ficus* from the 3'-end of CO I, PCR amplified and sequenced with primers C1-J-2183 and TL2-N-3014 (Genbank accession number DQ238220).

```
  ORIGIN
   1    GATTTGGTGC TATATCTCAA ATTATAAATC AAGAAAGGGG AAAACTAGAA ATTTTTAGAA
  61    AAATTAACAT AATTTTTGCC ATAATTTCAA TTGGTATTTT AGGTTTTATT GTTTGAGCTC
 121    ATCATATATT TACTATTGGT TTAGATATTG ACACACAATT ATATTTTATA TCAGCTACAA
 181    TAATTATTGC AATCCCTACT AGAATTAAAA TTTTTAGTTG AATAATAACT TTAAATGGTA
 241    AAAAAATTTT AAATTCATCT ATTAATTTTT GATCAATTGG ATTCATTATT ATATTTACTT
 301    TAGGAGGTTT AACTGGTATT ATTTTATCAA ATTCTATTAT TGATATTAAT TTACATGATA
 361    CATACTTTGT TGTAGCTCAC TTTCACTATG TATTATCAAT AGGAGTAATT TTTTCTATTT
 421    TTTCAAGTTT TATTTTTTGA TCTCCATTAT TATTTAATAT TTCTTTAAAT AATAATTGAT
 481    TAAAAATTAA TTTTTTCAAC CTTTTTTTAT CTATTAATTT AACTTTTTTT CCTCAACATT
 541    TTTTAGGTAT AAATGGAATA CCCCGTCGTT ATATTATATA TTCTGATTAT TTTATTTTAT
 601    GAAATAATAT TTCATCAATT GGATCTTCGA TAACAATTAT TTTTACTTTA ATTTTTATTT
 661    ATATTATTAT TGAGTCTTTA ATATGTAAAC GTTTAATTAT ATTTAAAATA AAATTTTTTA
 721    ATATTGAATG ATTAAATAAT TCTCCTAATT TAAATCATAC
```

Nucleotide sequence of *Planococcus citri* from the mid-section of CO I, PCR amplified and sequenced with primers C1-J-1718 and C1-N-2191 (Genbank accession number DQ238219).

```
  ORIGIN
   1    CATTAATATT AATATCATCA GATTTAATTT TTCCCCGATT AAATAATTTT AGATTTTGAT
  61    TATTAATTCC ATCACTTATT TTAATAATAA TAAATATAAT ATTATCTAAT AATATTAATA
 121    CAGGTTGAAC ACTTTACCCT CCTTTAATTA ATCAAAATTT TATTACATTA AATTTTATTA
 181    TTTTTTCTTT ACATTTAAAT GGAATTTCTT CTATTTTTAG ATCAATTAAT TTTATTTCAT
 241    CAATTTTTAT TATCAATAAT AATAATTTTT TTTTAAATAA TATTACTTTA TATATTTGAT
 301    CTATTATTAT TACAACTATT TTATTAATTA TTTCTATTCC AATTTTATCA AGAGCAATTA
 361    CTATAATTAT TTTAGATAAT AATCTTAATA TAAATTTTTT TAATCCATTA GGAAATGGTA
 421    ATCCAATTTT ATATCAACAT TTATTTTGAT TTTTTGGACA TCCAG
```

Nucleotide sequence of *Planococcus citri* from the 3'-end of CO I, PCR amplified and sequenced with primers C1-J-2183 and TL2-N-3014 (Genbank accession number DQ238221).

```
  ORIGIN
   1    TTGGAGCTAT ATCTCAAATT ATAAATCARG AAAGAGGAAA AATARAAATT TTTAGTAAAA
  61    TTAATATAAT TTTTGCTATA ATTTCAATTG GAATTTTAGG TTTTATTGTT TGAGCTCACC
 121    ATATATTTAC TATCGGATTA GATATYGATA CACAATTATA TTTTATATCA GCTACAATAA
 181    TTATTGCTAT TCCTACAAGA ATTAAAATCT TTAGATGAAT AATAACTTTA AATGGAAAAA
 241    AAATTCTTAA TTCATCTATT AACTTTTGAT CAATTGGATT TATTATTATA TTYACATTAG
 301    GAGGATTRAC TGGAATTATT TTATCAAATT CTATTATTGA TATTAATTTA CATGACACAT
 361    ATTTCGTTGT AGCTCATTTT CATTATGTAT TATCTATAGG AGTAATTTTT TCAATTTTTT
 421    CAAGATTTAT TTTTTGATCA CCATTACTAT TTAATGTTTC TTTAAATAAT AATTGATTAA
 481    AAATTAATTT TTTTAATCTA TTTTTATCTA TYAATTTAAC CTTTTTTCCT CAACATTTTT
 541    TAGGAATAAA TGGAATACCT CGTCGTTATA TTATATATTC TGATTATTTT ATTTTATGAA
 601    ATAATATTTC ATCAGTTGGT TCATCTATAA CAATTATTTT TACTYTAATT TTTATTTACA
 661    TTATTATTGA ATCATTTATA TGTAAACGTT TAATTATATT TAAAATAAAA TTTTTTAATA
 721    TTGAATGATT AAATAATTCA CCTAATTTAA ATCATACATT TA
```

Nucleotide sequence of *Pseudococcus longispinus* from the 3'-end of CO I, PCR amplified and sequenced with primers C1-J-2183 and TL2-N-3014 (Genbank accession number DQ238222).

```
  ORIGIN
   1    GGAGCTATAT CACAAATTAT AAATCAAGAA ACAGGAAAAA TAGAAATATT CAGAAAAATT
  61    AATATAATTT TTGCTATAAT TTCAATTGGA ATTTTAGGAT TTATTGTTTG AGCTCATCAT
 121    ATATTTACAA TTGGATTAGA TATTGATACA CAACTTTATT TTATATCAGC AACAATAATT
 181    ATTGCAATTC CAACAAGAAT TAAAATTTTT AGATGAATAA TAACTTTAAA TGGAAAAAAA
 241    ATTTTAAATT CATCCATTTA TCTTTGATCC ACAGGATTTA TTATTATATT TACATTAGGA
 301    GGGTTAACTG GAATTATTCT TTCTAATTCA ATTATTGATA TTAATCTACA TGATACTTAT
 361    TTTGTTGTAG CACATTTTCA TTATGTACTA TCAATAGGAG TAATTTTTTC AATTTTTTCA
 421    AGTTTTATTT TTTGGTCTCC TTTATTAATA AATATTAGAT TAAACAATAA TTGATTAAAA
 481    ATTAATTTTT TTAATCTATT TATTTCCATT AATTTAACAT TTTTTCCCCA ACATTTCTTA
 541    GGAATTAATG GAATGCCACG TCGATATATT ATATATTCAG ATTATTTTAT TTTATGAAAT
 601    AATATTTCTT CAATTGGTTC TTCAATAACT ATTATTTTCA CCATAATATT TATTTTTATC
 661    ATTATTGAAT CTTTAATTTC TAAACGATTA ATTATATTTA AAATAAAATT TTTCAATAGA
 721    GAATGATTAA ATAATTCACC AAATTTAAAT CATACCTTTA ATG
```

Appendix 2: Mealybug amino acid translations

Amino acid translation of *Planococcus ficus* from nucleotide sequence obtained from the mid-region of CO I.

```
  ORIGIN
   1    PRLNNXSFWL LIPSLILMML NMILSNNINT GWTLYPPLIN QNFITLNFII FSLHLNGISS
  61    IFSSINFISS IFIINNNNFF LNNITLYIWS IIITTILLII SIPILSSAIT MIILDXNLNM
 121    NFFNPLGNGN PXLYQHLFWF FGHPEVYILI LPG
```

Amino acid translation of *Planococcus ficus* from nucleotide sequence obtained from the 3'-end of CO I.

```
  ORIGIN
   1    FGAMSQIMNQ ESGKLEIFSK INMIFAMISI GILGFIVWAH HMFTIGLDID TQLYFMSATM
  61    IIAIPTSIKI FSWMMTLNGK KILNSSINFW SIGFIIMFTL GGLTGIILSN SIIDINLHDT
 121    YFVVAHFHYV LSMGVIFSIF SSFIFWSPLL FNISLNNNWL KINFFNLFLS INLTFFPQHF
 181    LGMNGMPRRY IMYSDYFILW NNISSIGSSM TIIFTLIFIY IIIESLMCKR LIMFKMKFFN
 241    IEWLNNSPNL NHT
```

Amino acid translation of *Planococcus citri* from nucleotide sequence obtained from the mid-region of CO I.

```
  ORIGIN
   1    LMLMSSDLIF PRLNNFSFWL LIPSLILMMM NMMLSNNINT GWTLYPPLIN QNFITLNFII
  61    FSLHLNGISS IFSSINFISS IFIINNNNFF LNNITLYIWS IIITTILLII SIPILSSAIT
 121    MIILDNNLNM NFFNPLGNGN PILYQHLFWF FGHP
```

Amino acid translation of *Planococcus citri* from nucleotide sequence obtained from the 3'-end of CO I.

```
  ORIGIN
   1    GAMSQIMNQE SGKMXIFSKI NMIFAMISIG ILGFIVWAHH MFTIGLDIDT QLYFMSATMI
  61    IAIPTSIKIF SWMMTLNGKK ILNSSINFWS IGFIIMFTLG GLTGIILSNS IIDINLHDTY
 121    FVVAHFHYVL SMGVIFSIFS SFIFWSPLLF NVSLNNNWLK INFFNLFLSI NLTFFPQHFL
 181    GMNGMPRRYI MYSDYFILWN NISSVGSSMT IIFTLIFIYI IIESFMCKRL IMFKMKFFNI
 241    EWLNNSPNLN HTF
```

Amino acid translation of *Pseudococcus longispinus* from nucleotide sequence obtained from the 3'-end of CO I.

```
  ORIGIN
   1    GAMSQIMNQE TGKMEMFSKI NMIFAMISIG ILGFIVWAHH MFTIGLDIDT QLYFMSATMI
  61    IAIPTSIKIF SWMMTLNGKK ILNSSIYLWS TGFIIMFTLG GLTGIILSNS IIDINLHDTY
 121    FVVAHFHYVL SMGVIFSIFS SFIFWSPLLM NISLNNNWLK INFFNLFISI NLTFFPQHFL
 181    GINGMPRRYI MYSDYFILWN NISSIGSSMT IIFTMMFIFI IIESLISKRL IMFKMKFFNS
 241    EWLNNSPNLN HTFN
```

Appendix 3: Comparison of mealybug nucleotide sequences

Comparison of the nucleotide sequence obtained from *Planococcus ficus*, *Planococcus citri* and *Pseudococcus longispinus* using primers C1-J-1719, C1-N-2191, C1-J-2183 and TL2-N3014. Only the region of overlap is shown. The nucleotide sequence of *P. ficus* is given in full, others are given only as they differ from that of *P. ficus*. Numbers refer to the position of the bases relative to the entire amplified sequence.

P. ficus	1	CCACGATTAA	ATAATTTSAG	ATTTTGATTA	TTAATTCCAT	CTCTTATTTT	AATAATATTA	AATATAATTT	TATCTAATAA	TATTAATACT	GGTTGAACTT
P. citri	33	··C·····	·····T··	·····	·····	·A·····	·····A··	·····A·	·····	·····A	·····AC
P. ficus	101	TATATCCCCC	TTTAATTAAT	CAAAATTTTA	TTACATTAAA	TTTTATCATT	TTTTCTTTAC	ATTTAAATGG	TATTTCATCA	ATTTTTAGAT	CAATCAATTT
P. citri	133	·T··C··T··	·····	·····	·····	·····T···	·····	·····	A·····T··T	·····	····T·····
P. ficus	201	TATTTCTTCA	ATTTTTATTA	TTAATAATAA	TAATTTTTTT	TTAAATAATA	TTACTCTTTA	TATTTGATCT	ATTATTATCA	CAACTATTTT	ATTAATTATT
P. citri	233	·····A···	·····	·C·····	·····	·····	·····T·A··	·····	·····T·	·····	·····
P. ficus	301	TCTATTCCTA	TTTTATCTAG	AGCTATTACT	ATAATTATTT	TAGATAAWAA	TCTTAATATA	AATTTTTTTA	ATCCCTTAGG	AAATGGTAAC	CCAWTTTTAT
P. citri	333	·····A·	·····A··	···A·····	·····	·····T··	·····	·····	····A·····	·····T	···A·····
P. ficus	401/1	ACCAACATTT	ATTTTGATTT	TTTGGGCATC	CAGAAGTTTA	TATTTTAATT	TTACCGGGAT	TTGGTGCTAT	ATCTCAAATT	ATAAATCAAG	AAAGGGGAAA
P. citri	433/1	·T·····	·····	·····A····	···???????	??????????	??????????	····A·····	·····	·····R·	····A·····
P. longispinus	1							??··A·····	···A·····	·····	···CA·····
P. ficus	44	ACTAGAAATT	TTTAGAAAAA	TTAACATAAT	TTTTGCCATA	ATTTCAATTG	GTATTTTAGG	TTTTATTGTT	TGAGCTCATC	ATATATTTAC	TATTGGTTTA
P. citri	41	·A··R·····	·····T····	····T·····	·····T···	·····	·A·····	·····	·····C·	·····	···C··A···
P. longispinus	39	·A·····A	··C·····	····T·····	·····T···	·····	·A·····	A·····	·····	·····	A·····A···
P. ficus	144	GATATTGACA	CACAATTATA	TTTTATATCA	GCTACAATAA	TTATTGCAAT	CCCTACTAGA	ATTAAAATTT	TTAGTTGAAT	AATAACTTTA	AATGGTAAAA
P. citri	141	·····Y··T·	·····	·····	·····	·····T··	T·····A···	·····C·	····A·····	·····	·····A····
P. longispinus	139	·····T·	·····C·T··	·····	··A·····	·····	T··A··A···	·····	····A·····	·····	·····A····
P. ficus	244	AAATTTTAAA	TTCATCTATT	AATTTTTGAT	CAATTGGATT	CATTATTATA	TTTACTTTAG	GAGGTTTAAC	TGGTATTATT	TTATCAAATT	CTATTATTGA
P. citri	241	·····C·T··	·····	··C·····	·····	T·····	··Y··A····	····A··R··	···A·····	·····	·····
P. longispinus	239	·····	·····C···	T··C·····	·C·CA·····	T·····	·····A····	····G·····	···A·····	C·T··T····	·A·····

P. ficus	344	TATTAATTTA	CATGATACAT	ACTTTGTTGT	AGCTCACTTT	CACTATGTAT	TATCAATAGG	AGTAATTTTT	TCTATTTTTT	CAAGTTTTAT	TTTTTGATCT
P. citri	341	·····C····	·T··C·····	·····T···	··T·····	····T·····	·····A·····	····A·····	·····A·····	····A·····	·····A
P. longispinus	339	·····C··	·····T·	·T·····	···A··T···	··T·····C	·····A·····	····A·····	·····G···		
P. ficus	444	CCATTATTAT	TTAATATTTC	TTTAAATAAT	AATTGATTAA	AAATTAATTT	TTTCAACCTT	TTTTTATCTA	TTAATTTAAC	TTTTTTTCCT	CAACATTTTT
P. citri	441	·····C···	·····G····	·····	·····	·····	···T··T··A	·····	·Y·····	C·····	·····
P. longispinus	439	··T·····A	·A·····AG	A·····C···	·····	·····	···T··T··A	···A·T··C·	·····	A·····C	·····C·
P. ficus	544	TAGGTATAAA	TGGAATACCC	CGTCGTTATA	TTATATATTC	TGATTATTTT	ATTTTATGAA	ATAATATTTC	ATCAATTGGA	TCTTCGATAA	CAATTATTTT
P. citri	541	····A·····	·····T	·····	·····	·····	·····	·····	····G····T	··A··T····	·····
P. longispinus	539	····A··T··	·····G··A	·····A····	·····	A·····	·····	·····	·····T	·····A····	·T·····
P. ficus	644	TACTTTAATT	TTTATTTATA	TTATTATTGA	GTCTTTAATA	TGTAAACGTT	TAATTATATT	TAAAATAAAA	TTTTTTAATA	TTGAATGATT	AAATAATTCT
P. citri	641	···Y·····	·····C·	·····	A··A··T···	·····	·····	·····	·····	·····	·····A
P. longispinus	639	C·CA·····A	·····T··	·C·····	A·····T	·C·····A·	·····	·····	·····C····	GA·····	·····A
P. ficus	744	CCTAATTTAA	ATCATAC								
P. citri	741	·····	·····								
P. longispinus	739	··A·····	·····								

Appendix 4: Comparison of mealybug amino acid sequences

Comparison of the amino acid sequence obtained from *Planococcus ficus*, *Planococcus citri* and *Pseudococcus longispinus* from amplified nucleotide sequence from the mid- and 3'-regions of CO I. The amino acid sequence of *P. ficus* is given in full, others are given only as they differ from that of *P. ficus*.

P. ficus	1		PRLNNXSFWL	LIPSLILMML	NMILSNNINT	GWTLYPPLIN	QNFITLNFII	FSLHLNGISS	IFSSINFISS	IFIINNNNFF	LNNITLYIWS
P. citri	1	LMLMSSDLIF	·····F····	·········M	··M·····	··	··	··	··	··	··
P. ficus	91	IIITTILLII	SIPILSSAIT	MIILDXNLNM	NFFNPLGNGN	PXLYQHLFWF	FGHPEVYILI	LPG			
P. citri	101	··	··	·····N····	··	·I·····	....				
P. ficus	1	F	GAMSQIMNQE	SGKLEIFSKI	NMIFAMISIG	ILGFIVWAHH	MFTIGLDIDT	QLYFMSATMI	IAIPTSIKIF	SWMMTLNGKK	ILNSSINFWS
P. citri	1		··	····X·····	··	··	··	··	··	··	··
P. longispinus	1		··	T··M·M····	··	··	··	··	··	··	·····YL··
P. ficus	92	IGFIIMFTLG	GLTGIILSNS	IIDINLHDTY	FVVAHFHYVL	SMGVIFSIFS	SFIFWSPLLF	NISLNNNWLK	INFFNLFLSI	NLTFFPQHFL	GMNGMPRRYI
P. citri	91	··	··	··	··	··	··	·V·····	··	··	··
P. longispinus	91	T·····	··	··	··	··	·····M	··	·····I··	··	·I·····
P. ficus	192	MYSDYFILWN	NISSIGSSMT	IIFTLIFIYI	IIESLMCKRL	IMFKMKFFNI	EWLNNSPNLN	HT			
P. citri	191	··	····V·····	··	····F·····	··	··	··F			
P. longispinus	191	··	··	····MM··F·	·····IS···	·····S	··	··FN			

Appendix 5: BLAST search results for mealybug sequences

BLAST searches performed on GenBank on 13/10/05 with mitochondrial CO I mealybug sequences obtained from *Planococcus ficus*, *Planococcus citri* and *Pseudococcus longispinus*. Only the first 5 results are shown. Percent identity refers only to the region of overlap.

Results of BLAST of nucleotide sequence from *P. ficus* from mid-CO I with primers C1-J-1718 and C1-N-2191 (458 bp).

Organism	Accession number	Genetic region	Alignment length	% identity	Reference
No significant similarity found					

Results of BLAST of nucleotide sequence from *P. citri* from mid-CO I with primers C1-J-1718 and C1-N-2191 (465 bp).

Organism		Accession number	Genetic region	Alignment length	% identity	Reference
honey bee	*Apis dorsata* (htype 6)	AY012727	mtDNA, CO I	131	83.97	Tanaka *et al.* 2001a
honey bee	*Apis dorsata* (htype 5)	AY012726	mtDNA, CO I	131	83.97	Tanaka *et al.* 2001a
honey bee	*Apis dorsata* (htype 4)	AY012725	mtDNA, CO I	131	83.97	Tanaka *et al.* 2001a
honey bee	*Apis dorsata* (htype 3)	AY012724	mtDNA, CO I	131	83.97	Tanaka *et al.* 2001a
honey bee	*Apis dorsata* (htype 2)	AF153113	mtDNA, CO I	131	83.97	Tanaka *et al.* 2001b

Results of BLAST of nucleotide sequence from *P. ficus* from 3'-CO I with primers C1-J-2183 and TL2-N-3014 (760 bp).

Organism		Accession number	Genetic region	Alignment length	% identity	Reference
mealybug	*Dysmicoccus brevis*	AF483204	mtDNA, CO I & II	407	92.14	Thao *et al.* 2002
mealybug	*Melanococcus albizziae*	AF483205	mtDNA, CO I & II	398	91.46	Thao *et al.* 2002
mealybug	*Maconellicoccus hirsutus*	AF483207	mtDNA, CO I & II	408	90.69	Thao *et al.* 2002
mealybug	*Planococcus citri*	AY179431	mtDNA, CO I	353	92.35	Gullan *et al.* 2003
mealybug	*Ferrisia virgata* (FL2)	AY179444	mtDNA, CO I	352	92.33	Gullan *et al.* 2003

Results of BLAST of nucleotide sequence from *P. citri* from 3'-CO I with primers C1-J-2183 and TL2-N-3014 (762 bp).

Organism		Accession number	Genetic region	Alignment length	% identity	Reference
mealybug	*Planococcus citri*	AY179431	mtDNA, CO I	178	98.88	Gullan *et al.* 2003
mealybug	*Dysmicoccus brevis*	AF483204	mtDNA, CO I & II	178	99.44	Thao *et al.* 2002
mealybug	*Dysmicoccus brevis*	AF483204	mtDNA, CO I & II	65	100.0	Thao *et al.* 2002
mealybug	*Ferrisia malvastra*	AY179432	mtDNA, CO I	174	93.10	Gullan *et al.* 2003
mealybug	*Melanococcus albizziae*	AF483205	mtDNA, CO I & II	139	95.68	Thao *et al.* 2002

Results of BLAST of nucleotide sequence from *P. longispinus* from 3'-CO I with primers C1-J-2183 and TL2-N-3014 (763 bp).

Organism		Accession number	Genetic region	Alignment length	% identity	Reference
mealybug	*Pseudococcus longispinus*	AY179439	mtDNA, CO I	348	98.56	Gullan *et al.* 2003
mealybug	*Melanococcus albizziae*	AF483205	mtDNA, CO I & II	399	91.23	Thao *et al.* 2002
mealybug	*Dysmicoccus brevis*	AF483204	mtDNA, CO I & II	406	90.64	Thao *et al.* 2002
mealybug	*Maconellicoccus hirsutus*	AF483207	mtDNA, CO I & II	398	90.70	Thao *et al.* 2002
mealybug	*Planococcus citri*	AF483206	mtDNA, CO I & II	406	89.66	Thao *et al.* 2002