INTRODUCTION

Pantoea ananatis, formerly classified as *Erwinia ananas* (syn. *Erwinia uredovora*) (Mergaert *et al.*, 1993), belongs to the family *Enterobacteriaceae* and is a plant-associated pathogen. It is a Gram-negative, rod-shaped, yellow-pigmented bacterium. It is motile by means of a peritrichous flagellum and is non-spore-forming and facultatively anaerobic. Strains can grow at temperatures of 4-41 ºC, but no strain grows at 44 ºC. The type strain of *Pantoea ananatis* was isolated from pineapple in Brazil (Mergaert *et al.*, 1993). The name of this species has its origin in the generic name of pineapple, *Ananas comosus*, the source from which it was first isolated.

Pantoea ananatis has been identified as the causal agent of bacterial blight of *Eucalyptus* (Coutinho *et al.*, 2002); leaf spot of maize (Paccola-Meirelles *et al.*, 2001); necrotic leaf blotch disease of sudangrass (Azad *et al.*, 2000); leaf blight, seed stalk rot and bulb decay of onion (Gitaitis and Gay, 1997); postharvest disease of cantaloupe fruit (Bruton *et al.*, 1991) and brown spot of honeydew melons (Wells *et al.*, 1987). The pathogen has also been detected in naturally infested onion seed (Walcott *et al.*, 2002). The first reports of *P. ananatis* date back to 1928 where it was discovered to cause brown rot of pineapple fruitlets (Serrano, 1928).

In Georgia, U.S.A., *P. ananatis* has been recovered from 25 asymptomatic weed species, including crabgrass, Texas millet and tall verbena (Gitaitis *et al.*, 2002). Additionally the bacterium was also found on crop plants such as Bermuda grass, cowpea and soybean. The epiphytic presence of *P. ananatis* on weeds in Georgia implicates them as a possible source of inoculum. In the case of onion and sudangrass, it is known that the pathogen is both seedborne and seed-transmitted (Walcott *et al.*, 2002; Azad *et al.*, 2000)

The first report of *Pantoea ananatis* causing a disease on *Eucalyptus*, occurred in KwaZulu/Natal in 1998 (Coutinho *et al.*, 2002). Bacterial blight and dieback of the trees appeared in a single nursery on an *E. grandis* x *E. nitens* hybrid and spread to other nurseries and plantations, infecting a number of different *Eucalyptus* species and clones. Symptoms of bacterial blight include tip dieback and leaf spots on young leaves. *P. ananatis* appears to spread from the leaf petiole to the main vein and from there, infects the adjacent tissue. Initially, the leaf spots are

water-soaked and they often coalesce to form larger lesions. Leaf petioles become necrotic, resulting in premature abscission of the leaves. In the advanced stages of the disease, trees assume a scorched appearance and following repeated infection, become stunted (Coutinho *et al.*, 2002).

In South Africa, *Eucalyptus* species, hybrids and clones are grown commercially and account for more than 50 % of all newly afforested areas (Anonymous, 1996). Besides being used in the production of solid timber products, these trees form the basis of an internationally important pulp and paper industry. *P. ananatis* has the ability to infect a number of *Eucalyptus* species, hybrids and clones, which are crucial planting stock for the forestry industry, therefore this pathogen is of considerable concern. *Eucalyptus* trees suffering from bacterial blight in South Africa are more prevalent in areas where the temperature is low (20-25 °C) and the relative humidity is high. Strains have been recovered from Eucalypts showing similar symptoms in Uganda and three South American countries. The identity of these isolates has yet to be confirmed.

It is currently unknown as to how *P. ananatis* enters into its Eucalypt host. However, it has been established that the pathogen can enter other hosts through flowers, mechanical injury (Serrano, 1928), wounds created by feeding insects (Gitaitis *et al.*, 2003; Wells *et al.*, 1987) and plant-to-plant contact during high winds (Azad *et al.*, 2000). *P. ananatis* has been reported as a gut inhabitant of brown planthoppers (Watanabe *et al.*, 1996) and mulberry pyralids (Watanabe and Sato, 1999), however, neither of these insects act as a vector for the pathogen. Recently it was discovered that tobacco thrips do act as vectors for the transmission of *P. ananatis* to onions, causing the disease centre rot (Gitaitis *et al.*, 2003)

Serious economic losses can result from crops infected with *P. ananatis*. Blighted foliage can reach 50 % or more of the total leaf area of sudangrass infected with leaf blotch disease (Azad *et al.*, 2000). Leaf spot disease of maize results in a decrease in grain size and weight, and a decrease of 63,2 % in grain yield has been detected (Pinto, 1995). In the case of centre rot of onions, yield losses from 20-25 % to 80-100 % have been experienced (Gitaitis and Gay, 1997). *P. ananatis* is not only a primary plant pathogen in the field, but is also responsible for

postharvest losses of cantaloupe fruit (Bruton *et al.*, 1991), honeydew melons (Wells *et al.*, 1987) and onions (Gitaitis *et al.*, 2003).

Pantoea ananatis is one of only a few species of bacteria which includes ice-nucleating strains (Sigee, 1993). Bacteria exhibiting this trait can initiate the formation of ice that results in frost injury (Lindow, 1983). Frost sensitive plants, such as *Eucalyptus*, poplar and stone fruit tress, become more susceptible to freezing damage by the epiphytic presence of ice-nucleating *P. ananatis* strains. This ice-nucleation activity has been exploited by the food industry and by biological control specialists. Extracellular ice-nucleators from *P. ananatis* have been tested and applied in the freezing of foods in order to obtain the desired texture (Zasypkin and Lee, 1999) and in freeze-drying of foods (Watanabe and Arai, 1987). Ice-nucleating strains of *P. ananatis* have also been found to markedly reduce the cold hardiness of the mulberry pyralid larvae (Watanabe and Sato, 1999) and thus these stains have the potential as biological control agents of insect pests.

The epidemiology of plant diseases caused by *P. ananatis* is unclear. In South Africa, *P. ananatis* has only been isolated from *Eucalyptus*, onion and maize. It is not known if these outbreaks were caused by the same strain or how the bacterium could have spread from onions to *Eucalyptus* to maize. The recent appearances of bacterial blight and dieback on *Eucalyptus* in countries that have purchased seed from South Africa suggests seed transmission (unpublished data). There are also unconfirmed reports of the occurrence of *P. ananatis* on Eucalypts in South America and Uganda.

Identification of *P. ananatis*, based on phenotypic analysis, is difficult due to similarities in phenotypic characteristics between *Pantoea* species and related *Enterobacteriaceae* (Mergaert *et al.*, 1984; Verdonck *et al.*, 1987). A rapid, accurate molecular identification technique has yet to be developed for *P. ananatis*. It is important to understand the diversity and relationships, within and among related species as this leads to meaningful taxonomic classification, accurate identification, pathogen detection and epidemiology studies (Avrova *et al.*, 2002). Several molecular techniques, including DNA-DNA hybridisation (Mergaert *et al.*, 1993), 16S rRNA gene analysis (Coutinho *et al.*, 2002) and rep-PCR (unpublished results) have been performed on

isolates of *P. ananatis* and related species. However, a taxonomic study of the genus *Pantoea*, using a genomic fingerprinting technique such as AFLP or PFGE, has not yet been undertaken.

The genetic relatedness of *P. ananatis* strains must be examined and compared to other *Pantoea* species, in an attempt to trace the origin and spread of this pathogen. Whole genome analysis of *Pantoea* species may also allow the characterisation of the unidentified South American and Ugandan isolates.

Aim:

To examine the taxonomy and relatedness of *Pantoea* strains recovered from Eucalypts in South Africa, South America and Uganda

Objectives:

- To develop a rapid, reliable molecular-based identification technique for *Pantoea ananatis*
- To evaluate a genomic fingerprinting technique for the genus *Pantoea* to examine the genetic relatedness between species and strains

MOLECULAR TYPING METHODS FOR USE IN THE CLASSIFICATION OF *PANTOEA* SPECIES

2.1) Introduction:

The family *Enterobacteriaceae* was formed to include all facultatively anaerobic, Gram-negative rods. The organisms described in this family are typically straight, rod-shaped bacteria and most are motile by means of a peritrichous flagellum. They grow well at 37 ºC and are oxidase negative and catalase positive, with few exceptions (Holt *et al.*, 1994). The majority of organisms placed in the *Enterobacteriaceae* are associated with the digestive tract and human disease, for example *Escherichia*, *Salmonella* and *Shigella*. However, phytopathogenic bacteria, including *Erwinia*, *Brenneria* and *Pectobacterium* are also found in this family. A recent addition to the family *Enterobacteriaceae* is the genus *Pantoea*, proposed by Gavini *et al.* in 1989.

In 1920, Winslow and co-workers first described the genus *Erwinia* (named after the phytobacteriologist Erwin F. Smith), under which all the Gram-negative, fermentative, non-sporulating peritrichous flagellated plant pathogenic bacteria were placed. In 1972, it was proposed that the "herbicola group" of *Erwinia* be placed under the name *Enterobacter agglomerans.* This included several *Erwinia* species, including *E. herbicola, E. milletiae, E. ananas, E. uredovora* and *E. stewartii* (Ewing and Fife, 1972).

Synonymy of the species names *Enterobacter agglomerans*, *Erwinia herbicola* and *Erwinia milletiae* was suspected, based on a high degree of similarity of protein electropherograms and phenotypic data (Verdonck *et al.*, 1987). The type strains of the *Erwinia herbicola-Enterobacter agglomerans* complex belonged to the same DNA hybridisation group, which confirmed the suspected synonymy (Beji *et al.*, 1988). It was proposed that a new genus, *Pantoea*, should be established consisting of two species: *Pantoea agglomerans* as the type species, including the type strains of

Enterobacter agglomerans, *Erwinia herbicola* and *Erwinia milletiae*, and a new species named *Pantoea dispersa* (Gavini *et al.*, 1989).

In Japan in 1992, bacterial strains isolated from fruit and soils samples shared the general characteristics of the family *Enterobacteriaceae*. Following DNA hybridisation, three new species were described and classified in the genus *Pantoea* as *P. citrea, P. punctata and P. terrea* (Kageyama *et al.*, 1992). A year later it was proposed to transfer *Erwinia ananas*, *Erwinia uredovora* and *Erwinia stewartii* to the genus *Pantoea* following DNA relatedness studies and fatty acid analysis. *E. ananas* and *E. uredovora* were united as a single species which was classified as *Pantoea ananatis*. Two separate subspecies were created within the species *Pantoea stewartii* (formerly *Erwinia stewartii*), *P. stewartii* subsp. *stewartii* and *P. stewartii* subsp. *indologenes* (Mergaert *et al.*, 1993). Currently, seven species and two subspecies have been described under the genus *Pantoea*.

Pantoea stewartii subsp. *stewartii* (Mergaert *et al.*, 1993) is the causal agent of Stewart's vascular wilt of sweet corn and maize, which can result in huge crop losses. The potential risk of seed transmission of *P. stewartii* subsp. *stewartii* is considered so important in the international shipment of corn seed, that over 50 countries prohibit its import unless it is certified to be free of the pathogen (Coplin *et al.*, 2002). In many countries *P. stewartii* subsp. *stewartii* is classified as a quarantine pathogen.

Pantoea ananatis is the next most devastating species in the genus, causing a number of diseases on a wide variety of plant hosts. This pathogen causes bacterial blight and dieback of Eucalypts (Coutinho *et al.*, 2002); leaf spot of maize (Paccola-Meirelles *et al.*, 2001); necrotic leaf blotch disease of sudangrass (Azad *et al.*, 2000); leaf blight, seed stalk rot and bulb decay of onion (Gitaitis and Gay, 1997); postharvest disease of cantaloupe fruit (Bruton *et al.*, 1991) and brown spot of honeydew melons (Wells *et al.*, 1987). The first reports of *P. ananatis* date back to 1928 where it was discovered to cause brown rot of pineapple fruitlets (Serrano, 1928). Up to 100 % crop losses have been experienced for the majority of the cases listed above. There is also a recent report of *P. ananatis* causing a bacteremic infection in an elderly patient (DeBaere *et al.*, 2004).

Pantoea agglomerans causes palea browning of rice (Xie, 2001), crown and root gall of the ornamental plant, *Gypsophila paniculata* (Cooksey, 1986) and is associated with millet, apples and beans. *P. citrea* causes pink disease of pineapple and has found to be associated with citrus fruit (Pujol and Kado, 2000). Although *P. terrea* and *P. punctata* have been isolated from citrus fruit, they have yet to be reported to cause disease on these plant hosts (Kageyama *et al.*, 1992).

Little is known concerning the specificity and genetic relatedness of *P. ananatis* strains from different hosts and geographical regions. Additionally the epidemiology of the diseases caused by this pathogen is unclear. It is known that the pathogen is seedborne and seed-transmitted in several hosts, and is also present as an epiphyte on some crops and weeds. It is possible that *P. ananatis* exists as an epiphyte on many plant hosts and only causes disease under certain conditions.

As *Pantoea* is a recently described genus, there is still much confusion regarding taxonomic nomenclature. *Erwinia ananas* and *Erwinia uredovora* were synonymized based on a high level of genetic relatedness in 1993 (Mergaert *et al.*, 1993). However, the synonymy of these two species is not widely accepted, and the name *E. uredovora* is still commonly used in literature (Beyer *et al*., 2002). Similarly, the species names *Enterobacter agglomerans* and *Erwinia herbicola* are still used (Muniruzzaman *et al.*, 1994; Nizan-Koren *et al.*, 2003), despite the inclusion of these names under *Pantoea agglomerans* (Gavini *et al.*, 1989). In the most recent edition of Bergey's Manual of Determinative Bacteriology, only two species of *Pantoea* are listed, *Pantoea agglomerans* and *Pantoea dispersa* (Holt *et al.*, 1994). There is a need for the reference material to be updated to set a standard for the use of the correct nomenclature for *Pantoea* and its species.

Pantoea may only be a small genus, with seven known species, but its range of hosts is extremely varied. Species of *Pantoea*, especially *Pantoea ananatis*, are isolated from, or found to cause disease on, new hosts almost every year. Despite these regular isolations, there is little sequence data available for specifically *P. ananatis, P. dispersa, P. citrea, P. punctata* and *P. terrea*. This needs to be remedied if the identification and characterisation of these pathogens is to be rapid and accurate. The purpose of this review is to consider the various molecular typing methods previously applied to bacteria of the family *Enterobacteriaceae*; and to determine which of these methods can be utilised for a taxonomic study of the genus *Pantoea*.

2.2) Microbial Typing:

The term "typing" is used to indicate the differentiation of strains at the subspecies level or below and can be either identifying or comparative. Identifying, or determinative typing is utilised in the allocation of organisms to a previously described type in an existing classification scheme. With comparative typing, microbial isolates of a defined set are compared to each other for similarity, with reference to existing classification schemes (Dijkshoorn and Towner, 2001).

A high degree of similarity exists, not only among the species of the phytopathogenic enterobacteria, but between all organisms within the family *Enterobacteriaceae*. This is illustrated by close resemblance of their phenotypic characteristics, chemical composition and conservative 16S rRNA sequences (Stackebrandt *et al.,*1999). In the past phenotypic methods such as fatty acid composition and substrate utilisation profiles were used for the identification and classification of bacteria. However, phenotypic methods have low reproducibility and may not reflect the genetic relatedness of bacterial isolates. Additionally, phenotypic characteristics of microorganisms are susceptible to change depending on growth conditions, growth phase and spontaneous mutation (Wilson and Spencer, 1999). In contrast, DNA-based methods, specifically those that are PCR-based, are more stable, can define genetic relatedness clearly and are therefore better suited for microbial typing (Louws *et al.*, 1999).

2.2.1) Detection and Identification:

Detection involves establishing the presence of a particular target organism within a sample, whilst diagnosis refers to the identification of the nature and cause of a disease problem (Shurtleff and Averre, 1997). Techniques for both detection and diagnosis require certain levels of specificity, sensitivity and speed (Louws *et al.*, 1999). In the case of PCR-based methods, specificity is determined by primer selection and amplification conditions, while sensitivity depends on the nature of the PCR protocol and sample. Identification is defined as the assignment of an unknown organism into a known taxonomic group, based on selected characteristics (Shurtleff and Averre, 1997). PCR-based identification techniques, for detection and diagnosis of pathogens, involve the use of specific primers to amplify a diagnostic fragment or universal primers that provide a diagnostic genomic fingerprint (Louws *et al.*, 1999).

2.2.2) Diversity and Phylogeny:

Diversity refers to the degree of genetic variation within bacterial populations and is related to bacterial systematics at multiple taxonomic and phylogenetic levels (Louws *et al.*, 1999). Taxonomy can be divided into three parts: classification, nomenclature and identification of unknown organisms (Vandamme *et al.*, 1996). Diversity exists even at the species level, meaning that organisms isolated from different sources at different times and in different geographical regions can be classified into subtypes or strains (Olive and Bean, 1999). It is necessary to determine the genetic diversity of bacterial populations to establish a stable taxonomy, which leads to better classification schemes. Phylogenetics is the process of reconstructing possible evolutionary relationships and uses nucleotide sequences from conserved genes that act as molecular markers (Owen, 2004). The use of small-subunit RNA genes in the construction of branching trees, representing the distance of divergence from a common ancestor, has made a vast impact on microbial phylogenetics.

However, it is DNA-DNA hybridisation that is considered the "gold standard" for the delineation of bacterial species (Stackebrandt and Goebel, 1994). A species is defined as a group of strains, including the type strain, sharing 70 % or greater DNA-DNA relatedness with 5 ºC or less melting temperature of the hybrid (Wayne *et al.*, 1987). The percent DNA binding is an indirect parameter of the sequence similarity between two entire genomes. DNA hybridisation techniques have severe disadvantages as they depend on physiochemical parameters, are not cumulative, are cumbersome and require large quantities of DNA (Stackebrand and Goebel, 1994). Despite these disadvantages, DNA-DNA hybridisation values are considered part of the standard description of bacterial taxa, along with the moles percent guanosine plus cytosine (Vandamme *et al.*, 1996).

2.2.3) Molecular Typing Techniques:

There are several criteria which molecular typing methods must meet in order to be broadly useful. Firstly, a molecular typing method should require no previous knowledge of DNA sequence. Secondly, all organisms within a species must be typeable by the method used. Thirdly, any typing method must have a high discrimination power and lastly, a typing method has to provide reproducible results which can be easily interpreted (Olive and Bean, 1999). The choice of typing method depends on technical difficulty, cost, laboratory resources and length of time needed to obtain a result.

Molecular typing techniques can be characterised into two categories: specific gene variation and random whole-genome analysis. Specific gene variation includes single-locus and multi-locus typing. Highly variable genes, such as pathogenicity islands, can be targeted by the single-locus approach as they have been directly implicated in causing disease (Gürtler and Mayall, 2001). Pathogenicity islands have been found in a number of bacterial species, and contain virulence genes and mobilisation elements. However, the high variability of single loci chosen for molecular typing often obscures true relationships between isolates (Gürtler and Mayall, 2001). Multilocus sequence typing (MLST) and analysis of multi-gene families, such as the RNA operons and tRNA genes, are examples of the multi-locus approach to molecular typing. MLST involves

sequencing of several housekeeping alleles where the unique sequence of each gene fragment is considered a unique allele (Gürtler and Mayall, 2001). The rRNA operon contains three rRNA genes (16S, 23S and 5S) and three spacer regions which may include a tRNA gene. 16S rRNA sequencing is considered commonplace for phylogenetic analyses, whilst the 23S rRNA gene and tRNA genes have only been used to a limited extent. However, it is the 16S-23S internally transcribed spacer (ITS) region which is being used more frequently for molecular typing and microbial diversity studies.

Random whole-genome analysis techniques recognise random sites on the genome, that cannot be predicted without the whole genome sequence (Gürtler and Mayall, 2001). These random sites include repetitive elements and restriction enzyme sites. Repetitive extragenic palindromic-PCR (Rep-PCR) targets repetitive elements which are scattered throughout the genome, whilst pulsed-field gel electrophoresis (PFGE) and amplified fragment length polymorphism (AFLP) analysis are used to separate DNA fragments produced by digestion with restriction enzymes. Rep-PCR, PFGE and AFLP all generate DNA banding patterns, or genomic fingerprints, based on the genomic sequence of a microorganism. A genomic fingerprint can be described as a display of a set of DNA fragments from a specific DNA sample. When the entire genome is screened by random whole-genome analysis, variations in banding patterns are a direct reflection of the genetic relationship between the bacterial strains examined. Genomic fingerprints allow numerical analysis for comparative typing and identification purposes (Janssen *et al.*, 1996). In bacteriology, discrimination to the species level is most often referred to as identification, whilst typing denotes differentiation to the strain level (Savelkoul *et al.*, 1999). AFLP analysis can be used for both identification and typing of microorganisms.

From the two molecular typing categories, amplification of the 16S-23S ITS region and AFLP- or PFGE-analysis meet the most requirements stated previously. Neither require previous DNA sequence knowledge; both are reproducible and yield results that can be easily interpreted. These techniques will be examined in more depth in this review

2.3) Specific Gene Variation:

2.3.1) The 16S-23S Internally Transcribed Spacer (ITS) Region:

The rRNA operon is of great evolutionary interest since it is found in both eukaryotic and prokaryotic organisms. The rRNA operon, unlike metabolic operons (eg. *lac* operon and *trp* operon), does not encode for any proteins (Adhya, 1999). In prokaryotes, the rRNA operon consists of three conserved genes: the small subunit 16S rRNA gene, the large subunit 23S rRNA genes and the 5S rRNA gene. These functionally conserved genes are separated by variable spacer regions. The prokaryotic 16S rRNA gene has been used extensively in taxonomy studies at and above the species level. Because this gene has highly conserved regions interspersed with variable and hypervariable sequences, primers can be easily designed to amplify these sequences. Additionally a large database of 16S rRNA sequences is available for comparisons.

One of the disadvantages of targeting the 16S rRNA gene in diversity studies is that the length of the gene is considerably constant. Therefore amplified 16S rRNA genes from different isolates cannot be separated easily by electrophoresis. And, despite containing variable and hypervariable regions, they are often not divergent enough to distinguish between strains within the same species and sometimes even between species of the same genus (García-Martínez *et al.*, 1999). The length of the 23S rRNA gene sequence is twice as long as that of the 16S gene and relatively few sequences have been submitted to the ribosomal database. Consequently neither the 23S gene nor the 5S gene have been widely used in phylogenetic studies (Gürtler and Stanisich, 1996).

The spacer regions that separate the rRNA genes, do however exhibit a large degree of sequence and length variation at the genus and species level. The size of the spacer may vary among different species and even among different operons within a single cell, in the case of multiple operons (Condon *et al.*, 1995). In bacterial genomes, up to 11 copies of the rRNA operon can be present within the chromosome. After transcription of the rRNA operon, one pre-rRNA transcript is present which contains the following components:

5' – 16S gene – spacer – tRNA – spacer – 23S gene – spacer – 5S gene – 3'

This pre-rRNA transcript is then cleaved into separate rRNA and tRNA molecules (Watson *et al.*, 1987). The spacer regions found between the rRNA genes within the operon are more diverse due to the variation in number and type of tRNA sequences located within these spacers. The majority of spacer regions in Gram negative bacteria contain both a $tRNA^{ala}$ and a $tRNA^{ile}$ gene, whereas the spacers in Gram positive bacteria rarely contain any tRNA genes (Gürtler and Stanisich, 1996). Other functional units within the spacer region include the recognition sequence for the enzyme ribonuclease III, which is involved in splicing to produce the mature ribosome (Bram *et al.*, 1980).

The internally transcribed spacer (ITS) region, situated between the 16S and 23S ribosomal genes has been used extensively for both identifying and comparative typing of microbes. This region consists of a series of conserved, alignable sequences, found in all strains of a single species, alternating with stretches of hypervariable sequences that are of equal length or that contain insertions or deletions (García-Martínez *et al.*, 2001). Precise species identification can be achieved using the alignable sequences of the ITS region, whilst the hypervariable sequences often allow for characterization below the species level.

Techniques:

The ITS region is ideally situated for easy PCR amplification and sequencing with conserved regions of the adjacent 16S and 23S genes acting as primer-binding sites. There are several techniques that can be applied to the PCR amplification product to reveal the polymorphisms present in the ITS region.

ITS-PCR:

When universal primers are utilised, identification of isolates can be based on the number and length of ITS regions amplified (Jensen *et al.*, 1993). The occurrence of more than one band generated from a single isolate can result from varying lengths of several copies

of the rRNA operon. The variation among operons within a multi-rRNA operon genome may be as great as the variation of operons between closely related strains (Nagpal *et al.*, 1998).

Primer design, for the amplification of the ITS gene, is made simple when the sequences of the adjacent 16S and 23S rRNA genes are known. Also, PCR conditions can be optimised to increase efficiency of amplification. ITS-PCR has been successfully used in the detection of *P. ananatis* from weeds and crops (Gitaitis *et al.*, 2002) and for a phylogenetic study of phytopathogenic *Enterobacteriaceae* (Hauben *et al.*, 1998).

PCR-RFLP:

PCR-restriction fragment length polymorphism (PCR-RFLP) analysis combines specific PCR-based amplification, usually of a gene or part of a gene, with restriction digestion of that PCR product which is separated electrophoretically. If the restriction enzyme recognition site is uniquely situated, the resultant fragment size pattern can be used to differentiate between species and strains of organisms (Jensen *et al.*, 1993). PCR-RFLP analysis of the 16S-23S spacer region was used in subspecies differentiation of *Salmonella choleraesuis* (Shah and Romick, 1997). However, this technique relies on the presence or absence of a restriction enzyme site and as most mutations do not result in the creation or destruction of restriction enzyme sites, PCR-RFLP is not always a viable option for typing of isolates.

Applications:

Detection and Identification:

In order to facilitate genus and species level identification of bacteria, Jensen and co-workers developed a set of primers and PCR conditions to amplify the 16S-23S ITS region. Strains of *Salmonella, Citrobacter, Yersinia, Enterobacter, Escherichia* and *Proteus* from the family *Enterobacteriaceae* were screened using universal primers (Jensen *et al.*, 1993). The banding patterns produced were distinct for each different

species of the genera. Following comparison of the pattern profiles with a database, it was found that common profiles exist for each species, and it is possible to distinguish between the bacterial isolates screened, based on these profiles.

Seven different rice seed pathogens of the genera *Erwinia*, *Pseudomonas* and *Xanthomonas* were detected based solely on size polymorphisms of the primary and secondary ITS-PCR products generated by the use of universal primers complementary to the 16S and 23S rRNA genes (Kim and Song, 1996). A new species, *Erwinia pyrifoliae*, was proposed after isolates from Asian pears were analysed by sequencing the 16S rRNA gene and the ITS region (Kim *et al.*, 1999). The 16S rRNA gene is almost identical to that of *Erwinia amylovora*, however the ITS region sequence revealed that the Asian pear pathogen is only closely related to *E. amylovora*. DNA hybridisation results and the G + C content were significantly different to *E. amylovora* and a new species was suggested.

An alternative to universal primers, is the design of species-specific primers which were used in the detection of *Pantoea ananatis* as the causal agent of centre rot of onion (Walcott *et al.*, 2002). The 16S-23S ITS regions, of isolates from infected onions, were amplified using species-specific primers, designed based on nucleotide sequences of the ITS region of a known *P. ananatis* isolate. The presence of a single 400 bp PCR product is a positive indication of *P. ananatis* or *P. stewartii*, as the primers amplify the ITS regions of both of these species.

Bacterial isolates from blackleg-infected potatoes in Brazil were identified as pectolytic *Erwinia* species following amplification and sequencing of the 16S-23S ITS region (Duarte *et al.*, 2004). RFLP analysis of the ITS region differentiated the Brazilian strains from the subspecies of *E. carotvora* and from *E. chrysanthemi*. This confirmed that blackleg disease of potato is caused by a different strain of pectolytic *Erwinia* in Brazil than in temperate potato-growing regions.

A PCR assay with universal primers, specific for the 16S-23S ITS region, was evaluated for the detection and identification of waterborne pathogens such as *Escherichia coli*, *Salmonella enterica*, *Shigella dysenteriae* and *Yersinia enterolitica* (Pérez-Luz *et al.*, 2004). Specific profiles for each of the species were obtained, but the technique could not simultaneously detect microbes present in mixed population samples. However, a possible rapid identification method was found for waterborne bacteria.

Diversity and Phylogeny:

Up to eleven copies of the rRNA operon can be present on a bacterial genome. *Escherichia coli* K-12 has seven copies, located near to the origin of replication. PCR-RFLP analysis of the 16S-23S ITS region of *E. coli* strains revealed two different clusters present in the species, α and β (García-Martínez *et al.*, 1996[a]). To detect the sequence divergence between the two clusters, the 16S-23S ITS regions were sequenced and four phylogenetically informative variable sites were identified. Comparison of the sequences confirmed that two major phylogenetic branches exist within the species *E. coli* (García-Martínez *et al.*, 1996[b]). Comparison of sequences from different copies of rRNA operons from the same strain, revealed a degree of intercistron heterogeneity, meaning that identical rRNA operons are rarely found on the same genome.

To investigate the diversity of *Salmonella* strains isolated from different natural aquatic systems, strains were screened using the ribosomal spacer-heteroduplex polymorphism (RS-HP) method (Baudart *et al.*, 2000). Because seven copies of the rRNA operon exist in the *Salmonella* genome, amplification of the 16S–23S ITS region results in multiple amplification products. When multiple amplification products, containing homologous sequences flanking heterologous intervening sequences are generated, the products can cross-hybridize to form heteroduplex DNA structures (Jensen and Hubner, 1996). It was found that all serotypes of *Salmonella* can be easily differentiated using the RS-HP profiles and that a large diversity of serotypes exist within the natural aquatic environment.

Denaturing gradient gel electrophoresis (DGGE) was applied to the 16S-23S ITS region as a means to evaluate strain level differences in *E. coli* (Buchan *et al.*, 2001). DGGE relies on polyacrylamide gels that have an increasing linear gradient of denaturants (formamide and urea) to separate double-stranded DNA fragments on the basis of their sequence, not size (Fischer and Lerman, 1979). The sensitivity of DGGE to any slight sequence differences is high, meaning that single base changes are often revealed. The application of DGGE to the internally transcribed spacer region proved to be an effective method for diversity studies on field-collected samples, as unique banding patterns were obtained for *E. coli* isolates from each different source.

Sequence analysis of the 16S-23S ITS region was evaluated for phylogenetic studies of *E. amylovora* and *E. pyrifoliae* (McGhee *et al.*, 2002). It was found that the use of the ITS region for phylogenetic studies is questionable due to the high degree of variability within the spacer regions. Sequence analysis of the spacer regions of *E. amylovora* and *E. pyrifoliae* implied that recombination occurred between copies of the spacers, from different operons on the same chromosome. Therefore, misleading results can be produced when applying phylogenetic methods, meant for single nucleotide mutational substitutions, to polymorphisms due to recombination events (McGuire *et al.*, 1997).

2.4) Random Whole-Genome Analysis:

2.4.1) Repetitive Extragenic Palindromic-PCR (Rep-PCR):

Technique:

Rep-PCR is based on amplification from the sites of repetitive extragenic palindromic (REP) elements, enterobacterial repetitive intergenic consensus (ERIC) sequences and BOX elements found at different positions on the bacterial genome. The term "rep-PCR" includes amplification of any of the repetitive elements. REP elements are 38 bp long sequences consisting of six degenerate positions and a five bp variable loop between each side of a conserved palindromic stem (Stern *et al.*, 1984).

If two REP elements are located close enough, they can serve as primer binding sites to amplify the region of DNA between them. These introns are called inter-repeat fragments and, depending on the primers used, the number of fragments amplified will vary between strains (Towner and Grundmann, 2001).

ERIC sequences are 126 bp long elements which contain a highly conserved central inverted repeat and are located in extragenic regions of genomes of the *Enterobacteriaceae* (Hulton *et al.*, 1991). The third set of repetitive elements are the BOX elements. These sequences are less commonly used for microbial typing than REP and ERIC sequences. BOX elements are located within intergenic regions and are mosaic repetitive elements composed of various combinations of three subunit sequences referred to as boxA, boxB and boxC (Martin *et al.*, 1992).

Amplification of rep elements can be performed with a single primer, a single set of primers or multiple sets of primers. High annealing temperatures are used in the PCR cycle to produce distinct DNA fingerprints at a relatively high stringency (Towner and Grundmann, 2001). The fingerprints generated by REP-PCR are more complex than those by ERIC-PCR, however both are discriminative at the strain level.

Rep-PCR is one of the most widely used methods for microbial typing. The technique is simple, rapid and can be applied to any number of isolates. Rep-PCR has considerably better discriminatory power than RFLP analysis of the 16S rRNA gene or the 16S-23S ITS region. However, primer annealing to other homologous DNA, besides rep elements cannot be discounted (Gillings and Holley, 1997) and rep-PCR may not necessarily differentiate between very closely related strains (Louws *et al.*, 1999).

Applications:

Rep-PCR was used to establish the genomic diversity of *E. coli* water isolates collected from southern Brazil (dos Anjos Borges *et al.*, 2003). Following analysis, 28 clusters and 16 single patterns were contained in the dendrogram. Genetic diversity between isolates from different sampling sites could be observed from the results. *Erwinia* soft-rot bacterial isolates, from ornamental plants in Florida were characterized by Rep-PCR (Norman *et al.*, 2003). Strains could be clustered into two heterogeneous populations of *E. chrysanthemi* and *E. carotovora* subsp. *carotovora*, which correlated with fatty acid analysis and biochemical test results.

The identity of the bacterium that causes cucurbit yellow vine disease (CYVD) was confirmed as *Serratia marcescens* by Rep-PCR (Zhang *et al.*, 2003). CYVD strains were compared to strains of *S. marcescens* from different niches and type strains of other *Serratia* species. CYVD strains formed a loosely related group with *S. marcescens* strains from other niches, confirming its identity. Pulsed-field gel electrophoresis (PFGE) and Rep-PCR, using BOX, ERIC and REP primers, were compared for the genomic detection for *Salmonella* isolates (Weigel *et al.*, 2004). Both techniques were highly reproducible and produced approximately the same number of DNA fragments. However the Rep-PCR profiles were more variable and allowed greater discrimination.

2.4.2) Pulsed-Field Gel Electrophoresis (PFGE):

Technique:

Macrorestriction analysis of microbial genomes resolved by pulsed-field gel electrophoresis (PFGE) emerged in the 1990's as the method of choice for epidemiological studies, and is considered the "gold-standard" of molecular typing methods (Struelens *et al.*, 2001). Macrorestriction fingerprinting refers to digestion of chromosomal DNA with infrequent cutting enzymes, whilst PFGE implies gel electrophoresis where the electric field periodically changes in direction and intensity.

This is a high resolution technique which allows separation of DNA fragments as large as 12 Mb.

Intact, whole chromosomal DNA is required for macrorestriction analysis. This is achieved by incorporating bacterial isolates into ultra-pure, nuclease-free, low-melting temperature agarose plugs. The agarose plugs, containing whole bacteria, are then exposed to lytic solutions and infrequently-cutting restriction enzymes. Infrequently-cutting enzymes are selected on the basis of the rarity of occurrence of their recognition sequence in the target genome. Endonucleases which include CTAG in their recognition sequence cleave high G + C content genomes less than once per 1 kb, for example *Xba*I (TCTAGA) and *Spe*I (ACTAGT).

The digested bacterial plugs are inserted into an agarose gel and subjected to pulsed-field gel electrophoresis. The contour-clamped homogenous electric field (CHEF) system is the most popular PFGE system used for bacterial typing. A hexagonal array of electrotrodes periodically alternates uniform fields with an angle of reorientation of 120 ° (Struelens *et al.*, 2001). In response to changes in the orientation of the electric field, large DNA molecules migrate through the gel in a zig-zagging motion.

DNA fragments are visualized after ethidium bromide staining of the gel, followed by digital image capture under UV light. Computer-assisted analysis of PFGE gels allows rapid identification of strains by comparison of patterns to those in a large database. Comparison of patterns can be based either on densiometric curve comparisons of the Pearson correlation coefficient, or on the fragment matching Dice coefficient of similarity. The Dice coefficient, calculated as the number of matching size fragments multiplied by two and divided by the total number of fragments in a pair of patterns, is the preferred coefficient for comparing PFGE patterns (Struelens *et al.*, 2001).

A system for standardising the interpretation of PFGE patterns has been proposed, in which bacterial isolates yielding the same PFGE pattern are considered the same strain. Isolates differing by one to three bands are closely related and isolates differing by four to

six bands are possibly related. Bacterial isolates containing six or more band differences are considered unrelated (Tenover *et al.*, 1995).

Although macrorestriction analysis by PFGE has been used extensively for comparative typing in epidemiological investigations, it is still a laborious technique. The one major drawback of PFGE typing is the time needed to complete the analysis, which can take up to three days (Olive and Bean, 1999). Further automation of the technique would enhance the practicability, throughput and speed of the analysis (Struelens *et al.*, 2001).

Applications:

PFGE was performed on *Pantoea stewartii* subsp. *stewartii* isolates and patterns were compared to those of other *Pantoea* and *Erwinia* species (Coplin *et al.*, 2002). *P. stewartii* isolates could be easily distinguished based on their PFGE fingerprints. Sufficient divergence existed in the profiles to differentiate between *P. stewartii* subsp. *stewartii* isolates from different geographical regions. *Yersinia pestis* strains and isolates were screened by PFGE in an epidemiological study and all produced similar banding patterns (Huang *et al.*, 2002). Minor differences in the PFGE patterns occurred if the isolates came from different parent strains. PFGE offers an increased ability to discriminate between *Y. pestis* strains, thereby improving epidemiological studies.

Two different species of *Shigella*, *S. sonnei* and *S. flexneri*, were analyzed by PFGE typing. The genetic diversity of *Shigella sonnei* was examined by PFGE, ribotyping and plasmid profiling (DeLappe *et al.*, 2003). Two major clusters were identified by PFGE and a level of correlation existed between the different molecular typing methods. To trace the source of infection of a *Shigella flexneri* outbreak, PFGE was applied to all isolates recovered (Chiou *et al.*, 2001). Analysis of the PFGE patterns proved that *Shigella* infections were endemic to that area. PFGE analysis was used in epidemiological studies of outbreaks of *Klebsiella pneumoniae* (Ben-Hamouda *et al.*, 2002) and *Enterobacter cloacae* (Talon *et al.*, 2004) in two different neonatal wards.

However, in the case of the *E. cloacae* outbreak, the cause of the emergence of the pathogen could not be identified.

2.4.3) Amplified Fragment Length Polymorphism (AFLP) Analysis:

Technique:

The AFLP technique (Vos *et al.*, 1995), named for it's resemblance to restriction fragment length polymorphism (RFLP), combines the power of RFLP with the flexibility of PCR. AFLP is based on the detection of genomic restriction fragments by PCR amplification and can be applied to DNA from any origin. The technique consists of three steps: digestion of genomic DNA with two restriction endonucleases and ligation of double-stranded adaptors to the resulting restriction fragments; pre-amplification and selective amplification of restriction fragments with two sets of primers, complimentary to the ligated adaptors, and lastly, electrophoretic separation of the amplified products.

The two restriction endonucleases utilised in the AFLP technique include an infrequent (6 bp) cutter, and a frequent (4 bp) cutter. This ensures that a limited number of fragments are obtained which are of a relatively small size, simplifying the amplification process. For a microbial genome with a low G + C content (< 40 mol %), the restriction enzymes *Eco*RI and *Mse*I are recommended whilst *Apa*I and *Taq*I are suited for genomes with a high G + C content (> 55 mol %). For genomes with an intermediate G + C content (40 – 50 mol %), *Hin*dIII and *Taq*I could be used for restriction enzyme digestion (Janssen *et al.*, 1996).

Double-stranded adaptors, with sticky ends complimentary to those generated by the restriction enzyme digestion, are ligated to the restriction fragments in the presence of T4 DNA ligase and ATP. The adaptors are designed in such a way that the original restriction site is not restored after their ligation to the restriction fragments (Zabeau and Vos, 1993). This allows restriction enzymes to be present during ligation to prevent fragment-to-fragment ligation. As adaptors are unphosphorylated at their 5' ends,

adaptor-to-adaptor ligation is also prevented. The adjacent adaptor and restriction site sequences serve as primer binding sites for the subsequent amplification reactions.

AFLP primers consist of a 5' sequence corresponding to the adaptor, the restriction site sequence and a 3' selective extension. Therefore the design of the primers depends on the adaptors used during ligation. A pre-amplification reaction of the ligation mixture is performed, using primers with no selective nucleotides attached at their 3' ends. This pre-amplification step reduces the complexity of the ligation mixture and prevents background smears from appearing on the gel. Only restriction fragments with an adaptor ligated to each end will be amplified exponentially during pre-amplification.

The pre-amplification reaction is diluted and serves as a template for selective amplification. One to three selective nucleotides are included at the 3'-ends of the PCR primers, amplifying only a subset of the pre-amplified restriction fragments. Only restriction fragments, in which the nucleotides flanking the restriction sites are complementary to the selective nucleotides on the primers, will be amplified. As the number of selective nucleotides added to the primers increases, the complexity of the DNA fingerprint decreases. The number of amplified fragments is reduced four-fold with each additional selective nucleotide (Vos *et al*., 1995). At least three selective nucleotides are required at the 3'-ends of both primers to obtain suitable AFLP fingerprints for highly complex genomes, like those of plants or animals. For smaller genomes, like those of bacteria, usually only one selective nucleotide at the 3'-end of each primer is necessary (Vos *et al.*, 1995).

Mismatching between the primers and the template is not expected as AFLP primers are complementary to their target site of adaptor and restriction site sequences. The only possible mismatch that may occur is between the selective nucleotides at the 3'-end of the primer and the corresponding nucleotides flanking the restriction site (Kwok *et al.*, 1990). However, due to high-stringency PCR conditions, fragments with mispaired 3'-ends primers will not be amplified. The AFLP selective amplification begins with a high annealing temperature, which is lowered by 1 ºC per cycle until the optimal annealing

temperature is reached. This step is referred to as a "touchdown" and ensures the dissociation of mismatched primers and fragments, improving the specificity of the amplification.

Following selective amplification, reaction products are mixed with formamide loading dye and separated electrophoretically. In the past, selective primers were labelled radioactively and the resulting amplification products analysed on denaturing polyacrylamide gels. Autoradiography or phosphoimage analysis was used to visualise banding patterns (Vos *et al.*, 1995). However, as many laboratories prefer to limit the use of radioactivity, alternative detection methods have been evaluated, including silver-staining (Chalhoub *et al.*, 1997). Although silver-staining produces a similar resolution to that of autoradiography, silver-staining detects both strands of the AFLP products giving rise to double-band patterns which are difficult to analyse.

A more reliable approach involves the use of fluorescently-labelled selective primers. Fluorescent amplification products are detected as they migrate through the gel and pass through an on-line laser. Detection of AFLP fragments using an automated sequencer has several advantages. The cost of fluorescently-labelled primers is less than radioactively-labelled primers, images can be obtained within several hours instead of several days, and digital gel images can be rapidly analysed using specialised computer software.

Applications:

To investigate the applicability of AFLP for epidemiological studies of *Salmonella*, different serotypes were screened (Aarts *et al.*, 1998). Reproducible, unique profiles were obtained for each of the 62 serotypes genetically identified by AFLP analysis. Different strains, which were previously identified as being identical, could be distinguished proving that AFLP analysis is suitable for bacterial epidemiology and identification of *Salmonella*. Since this study, AFLP analysis has been used extensively for genotyping different *Salmonella* species and serovars, including *S. typhi* (Nair *et al.*, 2000), *S. enterica* serovar Typhimurium (Lan *et al.*, 2003) and *S. enterica* serovar Sofia (Ross *et*

al., 2003). AFLP analysis was compared with Pulsed-Field Gel Electrophoresis (PFGE) for epidemiological studies of *Salmonella enterica* (Lindstedt *et al.*, 2000). The AFLP technique proved to be fast and reproducible with a discriminatory power equal to that of PFGE.

AFLP was evaluated as an epidemiological typing method for *Eschericia coli* (Arnold *et al.*, 1999), and was applied to the isolates from the EcoR collection. Well-defined and reproducible results were obtained, which correlated closely with the MLEE results of the same isolates. Single Enzyme-AFLP (SE-AFLP) was evaluated as a rapid species- and strain-level identification technique (Velappan *et al.*, 2001). *E. coli, Yersina pestis* and several other pathogenic bacteria were screened. SE-AFLP utilises only the infrequent (6 bp) cutter restriction enzyme and there is no labelled primer in the amplification of the restriction fragments. SE-AFLP generates a lesser number of large fragments, allowing direct visual analysis of fingerprint patterns. Species- and strain-level identification of pathogenic bacteria was achieved by SE-AFLP and this technique is highly reproducible and sensitive.

To determine the taxonomic relationship between the soft rot bacteria *Erwinia carotovora* and *Erwinia chrysanthemi* at the subspecies level, AFLP analysis was performed (Avrova *et al.*, 2002). Subspecies-specific banding profiles were generated which allowed four clusters to be delineated. In addition, the technique identified several unknown isolates and discriminated between closely related strains for epidemiological investigations. AFLP analysis was evaluated for epidemiological and phylogenetic analyses of *Klebsiella* isolates (Jonas *et al.*, 2004). Results allowed the discrimination of *K. pneumoniae* and *K. oxytoca* into two and three phylogenetic groups, respectively. A preliminary cut-off value was also provided, based on AFLP data, for distinguishing epidemiologically non-related *Klebsiella* isolates.

2.5) Conclusions:

Bacterial typing based on variations of the 16S-23S ITS region is simple, rapid and cost-effective. However, this technique may not yield the most reproducible and discriminative results. The most reliable results will be those obtained from amplification of the ITS region of species with multiple rRNA operons (Gürtler and Stanisich, 1996).

Alternatively, AFLP is robust, reproducible and highly discriminative; even more so than rep-PCR which is prone to variations in amplification efficiency. AFLP and rep-PCR were compared with DNA-DNA homology studies, using *Xanthomonas* as a model system (Rademaker *et al.*, 2000). A high correlation was observed from the results, suggesting that AFLP and rep-PCR truly reflect the genotypic and phylogenetic relationships of organisms. Although the initial setup cost for AFLP is expensive, the technique is moderately easy and results can be obtained in as little as two days. A standardised scheme does not yet exist for AFLP as the choice of restriction enzymes, gel systems and visualisation technologies differ from laboratory to laboratory.

Whereas AFLP is initially expensive, PFGE is less so. Studies have proven that the discriminatory power of AFLP is equal to that of PFGE (Lindstedt *et al.*, 2000). However, PFGE is limited in its resolving power which contributes to difficulties with reproducibility and it can take up to four days to obtain analysis results. Nonetheless PFGE is still a viable option for microbial typing of *Pantoea* species, especially for genome mapping.

A RAPID MOLECULAR IDENTIFICATION TECHNIQUE FOR *PANTOEA ANANATIS*

3.1) Introduction:

Pantoea ananatis was isolated for the first time in 1924, from pineapple fruitlets exhibiting brown rot (Serrano, 1928). Since then, *P. ananatis* has been found to cause a wide variety of diseases on a range of agricultural crops including honeydew melon (Wells *et al.*, 1987); cantaloupe fruit (Bruton *et al.*, 1991); onion (Gitaitis and Gay, 1997); sudangrass (Azad *et al.*, 2000) and maize (Paccola-Meirelles *et al.*, 2001). In each of the above studies, the bacterium was characterised based on colony morphology, physiological tests and, in some cases fatty acid analyses. Identification of the causal agent based solely on these methods can prove to be unreliable, as a high similarity in phenotypic characteristics exists, not only between *Pantoea* species, but also among related *Enterobacteriaceae* (Verdonck *et al.*, 1987). In the past, identification based only on phenotypic tests has occasionally led to the incorrect identification of isolates. It is only with the characterisation of the most recent isolates of *P. ananatis*, from onion seed (Walcott *et al.*, 2002), weeds and crops (Gitaitis and Gay, 1997) and *Eucalyptus* (Coutinho *et al.*, 2002), that identification of the bacterium was based on genotypic-, as well as phenotypic- characteristics.

The increasingly frequent occurrences of *P. ananatis* from different hosts has raised many questions concerning the spread of the pathogen, its genetic relatedness to other *Pantoea* species and its presence as an epiphyte on many plants. But most importantly, the regular isolation of *P. ananatis* from these hosts has highlighted the need for a rapid, molecular identification technique for the pathogen. The 16S-23S internally transcribed spacer (ITS) region is variable, flanked and spaced by highly conserved sequences of coding ribosomal genes. There are also multiple copies of these genes in the genome, which facilitates PCR amplification. The 16S-23S ITS region is also more conserved within species than between species, making it a suitable marker for species identification.

Pantoea ananatis was identified from asymptomatic weeds and crops in Georgia, U.S.A, by means of a PCR assay targeted at the 16S-23S internally transcribed spacer (ITS) region (Gitaitis *et al.*, 2002). Species-specific primers (PanITS1 and EC5) were designed, based on the nucleotide sequence of the 16S-23S ITS region of a known *P. ananatis* isolate. An amplicon of 398 bp produced by amplification of DNA with the primers, was representative of a positive identification of *P. ananatis*. The same primers were also used in the detection of *P. ananatis* in naturally infested onion seed (Walcott *et al.*, 2002). In both studies, other *Pantoea* species were tested for specificity and sensitivity towards the designed primers. In both studies, the only other *Pantoea* species to produce a 398 bp amplicon was *P. stewartii* subsp. *stewartii* (Gitaitis *et al.*, 2002, Walcott *et al.*, 2002). In an attempt to prevent the amplification of *P. stewartii* subsp. *stewartii*, Gitaitis *et al.* modified the primers, PanITS1 (5' GTC TGA TAG AAA GAT AAA GAC 3') and EC5 (5' CGG TGG ATG CCC TGG CA 3'), which were used in their original study to PanITS sn2b and PanITS as2b (Personal communication). The modified primers, PanITS sn2b and PanITS as2b, were used in this research project.

The aim of this study, therefore, was to evaluate a PCR-based technique for the rapid identification of *Pantoea ananatis* using primers PanITS sn2b and PanITS as2b, targeting the 16S-23S ITS region.

3.2) Materials and Methods:

3.2.1) 16S-23S ITS-PCR Assay:

Authentic *P. ananatis* strains were received from the U.S.A, Hawaii, Japan, Brazil and South Africa from different hosts. Unidentified isolates, from infected Eucalypts exhibiting *Pantoea*-like symptoms were received from Uganda and three South American countries, hereafter referred to as South America 1, South America 2 and South America 3. Representative strains of *P. agglomerans, P. stewartii* subsp. *stewartii, P. stewartii* subsp. *indologenes, P. dispersa, P. terrea, P. punctata* and *P. citrea* were included in the study together with the type strains of all seven *Pantoea* species.

Nutrient broth (Oxoid) was inoculated with each of the isolates and incubated overnight at 37 ºC. Genomic DNA was extracted using the DNeasy™ tissue kit (QIAGEN) and stored at -20 ºC. DNA was amplified with the species-specific primers PanITS sn2b (5' GTC TGA TAG AAA GAA AAG AAG 3') and PanITS as2b (5' TTC ATA TCA CCT TAC CGG CGC 3') (Inqaba Biotechnologies), targeting the 16S-23S ITS region. Each 50 µL PCR reaction contained 1 x Reaction buffer, 1.5 mM $MgCl_2$, 250 µM of each nucleotide (dATP, dCTP, dGTP, dTTP), 25 pmol of each primer (forward and reverse), 1 U Taq DNA polymerase (Southern Cross Biotechnologies) and 50-100 ng genomic DNA. Amplification was carried out in a GeneAmp 2700 Thermal Cycler (Applied Biosystems). PCR conditions included denaturation at 95 °C for 5 minutes, 30 cycles of denaturation at 95 °C for 1 minute, annealing of primers at 56 °C for 1 minute, primer extension at 72 °C for 1 minute, and final chain elongation at 72 °C for 5 minutes. A negative water control was included with the PCR reactions to monitor contamination. Amplification products were run on a 1.5 % agarose gel (Promega) containing 10 mg/mL ethidium bromide, in 1 x TAE buffer at 100 V for 30 minutes and visualised and photographed under UV light. A 1 000 bp DNA marker (Roche) was run alongside the PCR products.

3.2.2) DNA Sequencing:

The 398 bp products from a small selection of the positive PCR reactions were excised from the agarose gels and purified. Three volumes of Binding buffer (Roche) and one volume of isopropanol were added to the gel slices. After heating at 55 ºC for 10 minutes, the mixture was added to a spin column. The QIAquick™ Purification Kit (QIAGEN) was used to complete the purification.

The purified 16S-23S ITS genes were sequenced with the forward primer, PanITS sn2b. The 10 µL sequencing reaction contained 2 µL Big Dye Sequencing Reaction Mix (ABI Prism), 3.2 pmol forward primer and 150 ng purified template DNA. Sequencing PCR conditions included denaturation at 96 ºC for 5 seconds and 25 cycles of denaturation at 96 ºC for 10 seconds, annealing of primer at 50 ºC for 5 seconds and elongation at 60 ºC

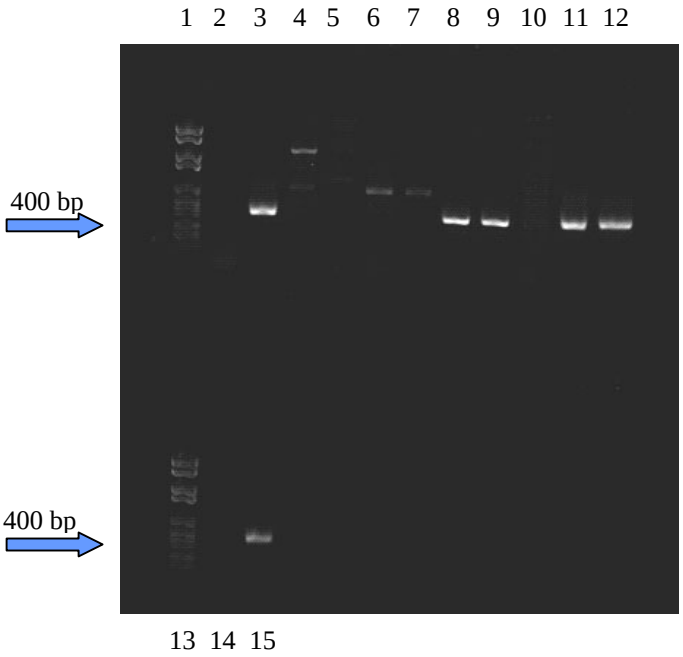
for 4 minutes. PCR products were sequenced on an ABI Prism DNA Automated Sequencer (Perkin-Elmer). Nucleotide sequences were edited and aligned using BioEdit Sequence Alignment Editor v 5.0.9. Heuristic searches with maximum parsimony, using stepwise addition produced a phylogenetic tree in PAUP*4.0b3 (Swofford, 2000). Bootstrap values were obtained from 1 000 bootstrap replicates, using step-wise addition.

3.3) Results:

3.3.1) 16S-23S ITS-PCR Assay:

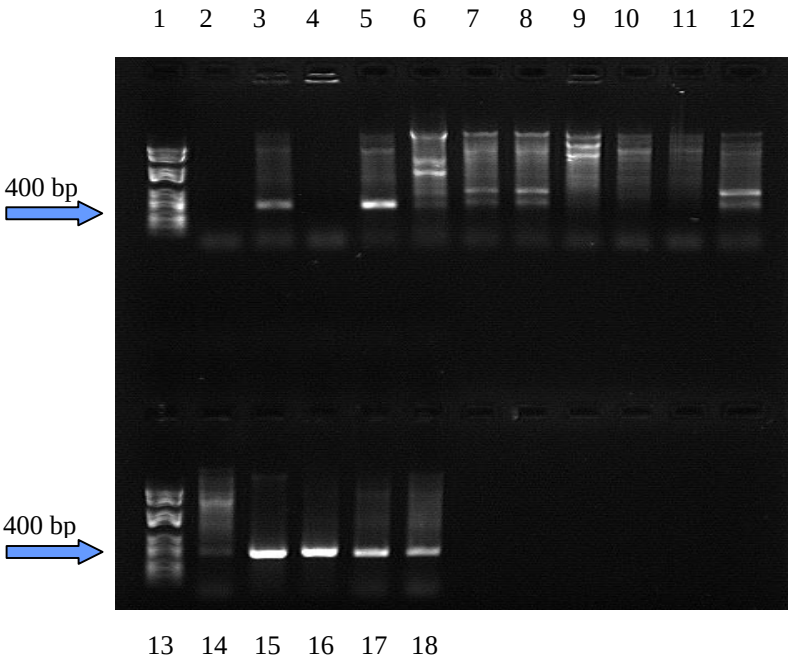
The presence of a 398 bp PCR product was a positive indication of *P. ananatis*, and was visible for all authentic strains (Fig. 3.1). Of the ten unidentified isolates from South America 1, only one (BCC007) yielded a 398 bp product following amplification. The remaining isolates produced different multiple banding patterns or large PCR products of between 600 bp to 900 bp (Fig. 3.1). None of the unidentified isolates from Uganda, South America 2 or 3 produced the correct sized amplicon. Multiple band patterns or smears were visible for the majority of these isolates (Fig. 3.2).

PCR products were not visible for the type strains of *P. agglomerans* (LMG1286^T); *P. stewartii* subsp. *stewartii* (LMG2715^T); *P. dispersa* (LMG2603^T); *P. punctata* (LMG22050^T); *P. terrea* (LMG22051^T) or *P. citrea* (LMG22049^T) (Fig. 3.3). A PCR product of the correct size was visible for the type strain of *P. stewartii* subsp. *indologenes* (LMG2632^T) (Fig. 3.3) and for all representative isolates of this subspecies. A 398 bp band was also present for two isolates received as *P. agglomerans*, from onion bulb (BCC146) (Fig. 3.2) and a *Sorbus* species (LMG2570) (Fig. 3.3), respectively, and possibly for a *P. terrea* isolate (LMG22099) (Fig. 3.3) from persimmon, although the latter PCR product was considerably fainter.



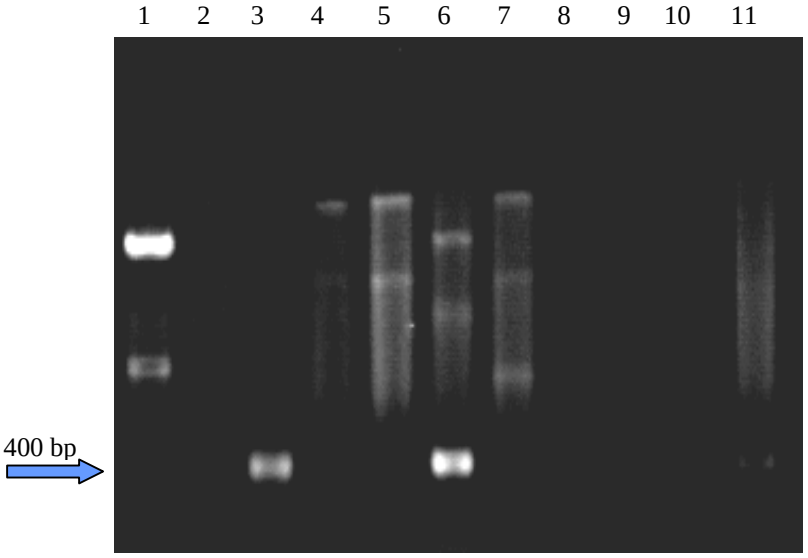
Lane 1: DNA marker VI
Lane 2: Negative control
Lane 3: LMG20103
Lane 4: BCC002
Lane 5: BCC005
Lane 6: BCC004
Lane 7: BCC015
Lane 8: LMG2676
Lane 9: LMG2671
Lane 10: LMG1286^{T}
Lane 11: LMG2665^{T}
Lane 12: LMG2666
Lane 13: DNA marker VI
Lane 14: BCC117
Lane 15: BCC112

Figure 3.1: Agarose gel electrophoresis of the 16S-23S ITS-PCR assay of authentic *Pantoea ananatis* (LMG2665^{T}, LMG20103, LMG2676, LMG2666, BCC112), *Pantoea stewartii* subsp. *indologenes* (LMG2671) and *Pantoea agglomerans* (LMG1286^{T}) strains and possible *P. ananatis* isolates (BCC002, BCC005, BCC004, BCC015, BCC117) from Uganda and South Africa.



Lane 1: DNA marker VI
Lane 2: Negative control
Lane 3: BCC095
Lane 4: BCC077
Lane 5: LMG20104
Lane 6: BCC109
Lane 7: BCC105
Lane 8: BCC107
Lane 9: BCC008
Lane 10: BCC011
Lane 11: BCC003
Lane 12: BCC006
Lane 13: DNA marker VI
Lane 14: BCC145
Lane 15: BCC146
Lane 16: BCC150
Lane 17: BCC154
Lane 18: BCC157

Figure 3.2: Agarose gel electrophoresis of the 16S-23S ITS-PCR assay of authentic *Pantoea ananatis* strains (LMG20104, BCC095, BCC145, BCC146, BCC150, BCC154, BCC157) and possible *P. ananatis* isolates (BCC077, BCC109, BCC105, BCC107, BCC008, BCC011, BCC003, BCC006) from South America and Uganda.



Lane 1: DNA marker XIV
Lane 2: Negative control
Lane 3: LMG2632^{T}
Lane 4: LMG2715^{T}
Lane 5: LMG22049^{T}
Lane 6: LMG2570
Lane 7: LMG22050^{T}
Lane 8: LMG22051^{T}
Lane 9: LMG2603^{T}
Lane 10: LMG22099

Figure 3.3: Agarose gel electrophoresis of the 16S-23S ITS-PCR assay of authentic *Pantoea stewartii* subsp. *indologenes* (LMG2632^{T}), *Pantoea stewartii* subsp. *stewartii* (LMG2715^{T}), *Pantoea agglomerans* (LMG2570), *Pantoea citrea* (LMG22049^{T}), *Pantoea punctata* (LMG22050^{T}), *Pantoea terrea* (LMG22051^{T}, LMG22099) and *Pantoea dispersa* (LMG2603^{T}) strains.

The results from the 16S-23S ITS-PCR assay are summarized in Table 3.1.

Table 3.1: Results from the 16S-23S ITS-PCR assay of authentic *Pantoea* species and possible *P. ananatis* isolates from South America and Uganda. Presence of a 398 bp product is denoted by +, whilst absence of the PCR product is denoted by -.

Code	Species	Location	Host	398 bp
LMG 2665^T	*P. ananatis*	Brazil	Pineapple	+
LMG 2666	*P. ananatis*	Hawaii	Pineapple	+
LMG 2668	*P. ananatis*	Hawaii	Pineapple	+
LMG 20103	*P. ananatis*	Piet Retief	*Eucalyptus*	+
LMG 20104	*P. ananatis*	Harding	*Eucalyptus*	+
LMG 20106	*P. ananatis*	Tzaneen	*Eucalyptus*	+
BCC095	*P. ananatis*	Vryheid	*Eucalyptus*	+
BCC118	*P. ananatis*	KZN	*Eucalyptus*	+
BCC117	?	KZN	*Eucalyptus*	-
BCC112	*P. ananatis*	KZN	*Eucalyptus*	+
BCC114	*P. ananatis*	KZN	*Eucalyptus*	+
BCC154	*P. ananatis*	Limpopo	Onion	+
BCC157	*P. ananatis*	Limpopo	Onion	+
BCC087	*P. ananatis*	USA	Onion	+
BCC083	*P. ananatis*	USA	Onion	+
BCC098	*P. ananatis*	USA	Sudangrass	+
BCC103	*P. ananatis*	USA	Sudangrass	+
BCC099	*P. ananatis*	USA	Sudangrass	+
LMG 2676	*P. ananatis*	USA	Wheat	+
BCC150	*P. ananatis*	USA	Honeydew melon	+
CTB1061	*P. ananatis*	Japan	Rice	+
CTB1135	*P. ananatis*	Japan	Rice	+
LMG 5342	*P. ananatis*	USA	Clinical strain	+
LMG 2678	*P. ananatis*	Zimbabwe	*Puccinia graminis*	+
LMG 2628	*P. ananatis*		Banana	+
LMG 22800	*P. ananatis*	Belgium	Human	+
LMG 1286^T	*P.agglomerans*	Zimbabwe	Human	-
LMG 2565	*P.agglomerans*	Canada	Cereals	-
LMG 2570	*P.agglomerans*	USA	*Sorbus sp.*	+
LMG 2660	*P.agglomerans*	Japan	*Wisteria floribunda*	-
BCC146	*P.agglomerans*	RSA	Onion bulb	+
BCC145	*P. agglomerans*	RSA	Bean leaf	-
LMG 2603^T	*P. dispersa*	Japan	Soil	-
LMG 2602	*P. dispersa*	India	*Sorghum bicolor*	-

LMG 2604	*P. dispersa*	Netherlands	*Rosa sp.*	-
LMG 2749	*P. dispersa*		Human	-
LMG 2632^{T}	*P. stewarti spp. indologenes*	India	Fox millet	+
LMG2631	*P. stewarti spp. indologenes*	India	Millet	+
LMG 2671	*P. stewarti spp. indologenes*	Hawaii	Pineapple	+
LMG 2630	*P.stewartii ssp. indologenes*		Guar gum powder	+
LMG 2673	*P.stewartii ssp. indologenes*	Hawaii	Pineapple	+
LMG 2674	*P.stewartii ssp. indologenes*	Hawaii	Pineapple	+
LMG 2715^{T}	*P. stewartii ssp.stewartii*	USA	Corn	-
LMG 2718	*P. stewartii ssp.stewartii*	USA	Corn	-
LMG 2713	*P. stewartii ssp.stewartii*	USA	Corn	-
LMG 22050^{T}	*P. punctata*	Japan	Mandarin orange	-
LMG 22097	*P. punctata*	Japan	Mandarin orange	-
LMG 22098	*P. punctata*	Japan	Persimmon	-
LMG 22051^{T}	*P. terrea*	Japan	Soil	-
LMG 22100	*P. terrea*	Japan	Persimmon	-
LMG 22099	*P. terrea*	Japan	Persimmon	+
LMG 22049^{T}	*P. citrea*	Japan	Mandarin orange	-
BCC109	?	Uganda	*Eucalyptus*	-
BCC105	?	Uganda	*Eucalyptus*	-
BCC107	?	Uganda	*Eucalyptus*	-
BCC002	?	South America 1	*Eucalyptus*	-
BCC003	?	South America 1	*Eucalyptus*	-
BCC004	?	South America 1	*Eucalyptus*	-
BCC005	?	South America 1	*Eucalyptus*	-
BCC007	?	South America 1	*Eucalyptus*	+
BCC006	?	South America 1	*Eucalyptus*	-
BCC015	?	South America 1	*Eucalyptus*	-
BCC016	?	South America 1	*Eucalyptus*	-
BCC008	?	South America 2	*Eucalyptus*	-
BCC009	?	South America 2	*Eucalyptus*	-
BCC010	?	South America 2	*Eucalyptus*	-
BCC011	?	South America 2	*Eucalyptus*	-
BCC013	?	South America 2	*Eucalyptus*	-
BCC072	?	South America 2	*Eucalyptus*	-
BCC074	?	South America 2	*Eucalyptus*	-

BCC075	?	South America 2	*Eucalyptus*	-
BCC076	?	South America 2	*Eucalyptus*	-
BCC077	?	South America 2	*Eucalyptus*	-
BCC078	?	South America 2	*Eucalyptus*	-
BCC079	?	South America 2	*Eucalyptus*	-
BCC080	?	South America 2	*Eucalyptus*	-
BCC081	?	South America 2	*Eucalyptus*	-
BCC082	?	South America 2	*Eucalyptus*	-
BCC067	?	South America 3	*Eucalyptus*	-

3.3.2) DNA Sequencing:

The most parsimonious phylogenetic tree was selected from sequence analysis of the partial 16S-23S ITS genes of *P. ananatis* and *P. stewartii* subsp. *indologenes* strains. Of the total number of characters used (334), 47 of the variable characters were parsimony-uninformative and 118 were parsimony-informative. The total tree length was 271 steps. The consistency index was 0.841 and the retention index was 0.882.

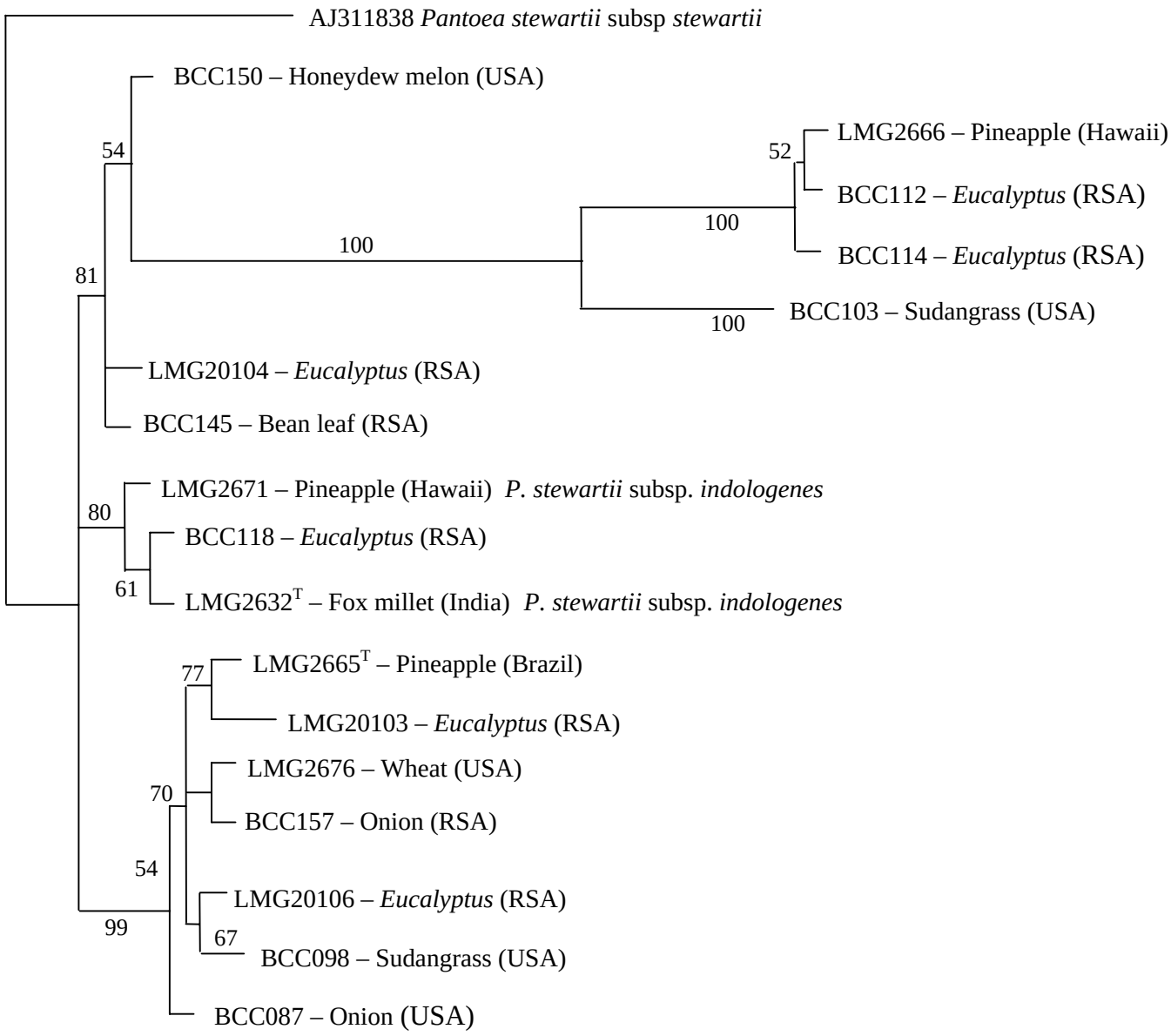


Figure 3.4: Phylogenetic tree of *P. ananatis* and *P. stewartii* subsp. *indologenes* strains based on partial 16S-23S ITS gene sequences (Bootstrap values are indicated on the branches).

3.4) Discussion and Conclusions:

3.4.1) 16S-23S ITS-PCR Assay:

Amplification of a region of the 16S-23S ITS gene with the species-specific primers PanITS sn2b and PanITS as2b has proven to be a rapid, molecular identification technique for *Pantoea ananatis*. However, it was not possible to exclude the non-specific amplification of several isolates previously identified as *P. stewartii* subsp. *indologenes*. The identities of the South American isolates could not be confirmed with the 16S-23S ITS-PCR assay and will be further characterised using alternative methods.

It was expected that authentic strains of *P. ananatis* would produce a PCR product of 398 bp, as the primers were designed based on the ITS sequence of a known *P. ananatis* isolate. The results of the study by Walcott and co-workers state that DNA from *P. stewartii* subsp. *stewartii* also yields a 398 bp product following amplification with their primers (Walcott *et al.*, 2002). Unexpectedly, none of the *P. stewartii* subsp. *stewartii* isolates, including the type strain, produced a 398 bp amplicon following PCR. The type strain and isolates of *P. stewartii* subsp. *indologenes* did, however produce the 398 bp product. Because the original primer sequences were modified it is possible that they now target *P. stewartii* subsp. *indologenes*, instead of *P. stewartii* subsp. *stewartii*, along with *P. ananatis*.

The absence of the correct size PCR product for the Ugandan and South American 2 isolates suggests these isolates are not *P. ananatis*. The fact that only one South American 1 isolate yielded a 398 bp product is unusual. This isolate, BCC007, is assumed to belong to either *P. ananatis* or *P. stewartii* subsp. *indologenes* as these are the only two species detected by the PCR assay. The remaining unidentified isolates from South America 1 may belong to a subspecies of *P. ananatis*, as they produced larger amplicons of 900 bp (BCC002) and 600 bp (BCC004 and BCC015). It is probable that the two isolates received as *P. agglomerans* (BCC146 and LMG2570) were incorrectly

identified, as both yielded the 398 bp amplicon. However, it was not confirmed whether these isolates were *P. ananatis* or *P. stewartii* subsp. *indologenes*.

The presence of a 398 bp PCR product for three out of four unidentified isolates recovered from Eucalypts infected with Coniothyrium canker, suggests that these three isolates (BCC118, BCC112 and BCC114) are *P. ananatis*. It appears that *P. ananatis* can exist in a synergistic relationship with the fungus *Coniothyrium zuluense* on *Eucalyptus* (van Zyl, 1999). Only one isolate recovered from this host did not yield a positive PCR product, BCC117. However, this correlates with previous findings that two *Pantoea* species were originally isolated along with the fungus. 16S rRNA analysis of these two species revealed high similarity to *P. ananatis* and *P. stewartii* subsp. *stewartii*, respectively (van Zyl, 1999).

The three *P. ananatis* isolates from sudangrass (BCC098, BCC099 and BCC103) all produced the correct sized PCR product, as expected. However, isolate BCC103 produced a second, slightly larger band. Despite changes made in annealing temperature and $MgCl_2$ concentration, the larger band was still amplified. This larger band is possibly another copy of the 16S-23S ITS gene containing an insertion, making it of a higher molecular weight.

A rapid, molecular identification technique for *Pantoea ananatis* was evaluated, and found to be species-specific except in the case of *P. stewartii* subsp. *indologenes*. The primer pair PanITS sn2b/PanITS as2b, require further modification and specificity to exclude the amplification of *P. stewartii* subsp. *indologenes* isolates. Further analysis of isolates BCC007, BCC146, LMG2570 and LMG22099 is necessary to determine their correct identification. The identity of the South American isolates remains unknown, but the possibility that they belong to a subspecies of *P. ananatis* cannot be discounted.

3.4.2) DNA Sequencing:

The manner in which the isolates grouped together (Fig. 3.4), provides little information concerning the origin or host specificity of *P. ananatis*. Neither geographical location, nor host plant determined the clustering of the isolates. Therefore at this stage, little can be determined from the phylogenetic tree, except that this pathogen is diverse in its distribution.

The clade containing the type strain of *P. ananatis* (LMG2665^T), and isolates from *Eucalyptus* (LMG20103, LMG20106), wheat (LMG2676), onion (BCC157) and sudangrass (BCC098) reveals that the sequences are closely related. The clustering of two of the unidentified isolates from *Eucalyptus* (BCC112, BCC114) and isolates from pineapple (LMG2666), sudangrass (BCC103) and honeydew melon (BCC150) suggests that these isolates could be a different subspecies of *P. ananatis*. It is possible that three, not two, *Pantoea* species were isolated along with *Coniothyrium zuluense* from *Eucalyptus* in KwaZulu/Natal (van Zyl, 1999). Two of the unidentified Coniothyrium canker isolates (BCC112 and BCC114) group with *P. ananatis* (LMG2666), whilst BCC118 groups with the *P. stewartii* subsp. *indologenes* type strain. The fourth unidentified isolate (BCC117) could belong to *P. stewartii* subsp. *stewartii* as it failed to produce a 398 bp PCR product and therefore was not sequenced.

Due to the fact that only those isolates that yielded a positive PCR product could be sequenced, the resulting dendrogram is not sufficiently informative or significant. The entire 16S-23S ITS region must be amplified and sequenced using universal primers to produce a product that can be examined for polymorphisms between strains and species. In this way, isolates from all seven species of *Pantoea* can be included in the study to possibly determine a pattern in the spread and geographical distribution of plant-pathogenic *Pantoea* isolates.