EVALUATION OF THE ENTIRE 16S-23S ITS REGION FOR MOLECULAR TYPING OF THE GENUS *PANTOEA*

4.1) Introduction:

The genus *Pantoea* was proposed in 1989 to include *Enterobacter agglomerans*, renamed as *Pantoea agglomerans*, and a newly described species, *Pantoea dispersa* (Gavini *et al.*, 1989). Three new species, isolated from soil and fruit samples in Japan, were added to the genus in 1992 as *P. citrea*, *P. punctata* and *P. terrea* (Kageyama *et al.*, 1992). The most recent additions to the genus took place in 1993, when *Erwinia ananas* (synonym *E. uredovora*) and *Erwinia stewartii* were renamed as *Pantoea ananatis* and *Pantoea stewartii* subsp. *stewartii*, respectively; and a new subspecies, *Pantoea stewartii* subsp. *indologenes* was described (Mergaert *et al.*, 1993). The proposal of each of the seven *Pantoea* species was based primarily on DNA hybridisation data, phenotypic tests and electrophoretic protein patterns or quinine composition. To date, seven species and two subspecies of *Pantoea* have been described, yet no standard molecular technique exists for differentiating between *Pantoea* species.

Sequencing of the 16S rRNA gene is widely used for differentiating between bacterial species. Along with DNA-DNA hybridisation, it is considered to be part of the standard description of bacterial taxa. However, 16S rRNA sequences are highly conserved, particularly among the family *Enterobacteriaceae* (Stackebrandt *et al.*, 1999). Many bacteria belonging to this family have genomes containing multiple copies of the rRNA operon. Only minor sequence differences exist between the multiple operons within a bacterium and amongst operons of closely related species (Stackebrandt and Goebel, 1994). Differentiation of *Pantoea* species, based only on 16S rRNA sequence data would not be possible due to this high conservation.

In contrast, the 16S-23S internally transcribed spacer (ITS) region is under less selective pressure than the rRNA structural genes and has exhibited greater sequence variability than the 16S rRNA gene. The size of the ITS region may vary among different species and even among different operons within a single cell, in the case of multiple operons (Condon *et al.*, 1995). Use of the 16S-23S ITS region sequence variability is becoming an important supplement to 16S rRNA sequencing as the standard for differentiating bacterial species (Nagpal *et al.*, 1998).

Differentiation of closely related bacterial species, based on the 16S-23S ITS region, has been achieved by either sequencing of the gene or analysis of band patterns, in the case of polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) or multiple copies of the rRNA operon. Bacterial typing based on spacer-length variation is supposedly rapid, cost-effective and favoured over sequencing (Gürtler and Stanisich, 1996). Sequence analysis of the ITS region has been used to successfully characterise species of the genus *Erwinia*, to which several *Pantoea* species formerly belonged (Duarte *et al.*, 2004; Kim *et al.*, 1999). *P. agglomerans* has been included in studies aimed at identification and detection of Gram-negative bacteria, utilising the 16S-23S ITS region (Jeng *et al.*, 2001; Jensen *et al.*, 1993). Based on the success of these studies, the ITS region seems an appropriate choice for differentiation of *Pantoea* species. Because a partial region of the 16S-23S ITS gene has been examined and deemed too limited for detection of any significant polymorphisms (Chapter 3), it is necessary to examine the entire 16S-23S ITS gene of *Pantoea* species.

The aim of this study was to evaluate a typing technique for the genus *Pantoea*, based on amplification of the entire 16S-23S ITS region using universal primers.

4.2) Materials and Methods:

4.2.1) Amplification of the Entire 16S-23S ITS Region:

Authentic *P. ananatis* strains were received from the U.S.A, Hawaii, Japan, Brazil and South Africa from different plant 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.

The 16S-23S ITS gene, from each of the isolates, was amplified with the universal primers FGPS 1490 (5' TGC GGC TGG ATC ACC TCC TT 3') and FGPL 132 (5' CCG GGT TTC CCC ATT CGG 3') (Laguerre *et al.*, 1996) (Inqaba Biotechnologies). Each 50 µL reaction contained 1 x Reaction buffer, 1.5 mM $MgCl_2$, 250 µM of each nucleotide (dATP, dCTP, dGTP, dTTP), 12 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). The PCR conditions included denaturation at 95 ºC for 3 minutes, 35 cycles of denaturation at 94 ºC for 30 seconds, annealing of primers at 55 ºC for 30 seconds and primer extension at 72 ºC for 1 minute, and final chain elongation at 72 ºC for 5 minutes. A negative water control was included with the PCR reactions to monitor contamination.

4.2.2) Product Separation:

Agarose Gels:

Amplification products were run on 1.5 % and 3 % agarose gels (Promega) containing 10 mg/mL ethidium bromide, in 1 x TAE buffer at 100 V for 30 minutes. A 1 000 bp DNA marker (Roche) was run alongside the PCR products. The gels were visualised and photographed under UV light.

5 % Polyacrylamide (PAGE) Gels:

Gels were prepared by mixing 5 % acrylamide (BDH Laboratory Supplies), 0.25 % bis-acrylamide (BDH Laboratory Supplies), as a crosslinker and 7.5 % urea in a 50 mM Tris/50 mM Boric acid/1 mM EDTA solution. 10 % ammonium persulphate and 50 µL TEMED were added to the gel solution. Gels were cast using the Bio-Rad gel apparatus and left to polymerise overnight. A 20 minute pre-run was performed at 1 W to equilibrate the ions in the gels and running buffer. PCR samples from the entire 16S-23S ITS-PCR assay, were mixed with formamide loading dye, heated for 5 minutes at 95 ºC and then cooled on ice for 10 minutes. A 2 000 bp marker (Roche) was run alongside the amplification products, at each end of the gel. Electrophoresis was performed at 60 W for 60 minutes, using a 1 x TBE (100 mM Tris/100 mM Boric acid/2 mM EDTA) running buffer. Following electrophoresis, gels were stained in a 1 x SYBR® Gold nucleic acid staining solution (Molecular Probes) for 40 minutes and then visualised and photographed under UV light.

Sequencing Gels:

The forward primer FGPS 1490 was fluorescently labelled with IRD-700 (Biolegio) and amplification was carried out on all isolates as described in section 4.2.1. Sequencing gels were prepared using 20 µL Long Ranger gel stock solution (8 % Long Ranger gel solution (LI-COR Biosciences), 7 M urea, 10 x TBE buffer) and 150 µL 10 % ammonium

persulphate and 15 µL TEMED for polymerisation. Gels were poured using the LI-COR gel casting apparatus and left to polymerise for 60 minutes. A 30 minute pre-run was performed at 1 500 V and 35 W to equilibrate the ions in the gel and running buffer. PCR products were diluted 1:100 with nuclease-free water (Promega), to minimise fluorescent smears, and mixed with an equal volume of formamide loading buffer (95 % formamide, 20mM EDTA, bromophenol blue). The mixture was heated at 90 ºC for 3 minutes and then cooled on ice for 10 minutes. 0.8 µL of each sample was loaded onto the LI-COR gels, along with an IRD-700 labelled sizing standard at each end of the gel. Gels were run on a LI-COR IR^2 automated sequencer (LI-COR Biosciences) for 4 hours at 1 500 V and 42 W with a 0.8 x TBE running buffer. The resulting band patterns were analysed with GelCompar (Applied Maths). Gels were normalised by aligning the 700 bp sizing standards at each end of the gels. Following analysis, a UPGMA dendrogram was constructed using the Dice similarity coefficient. An optimisation setting of 0.09 % and a tolerance setting of 0.48 % was applied to the analysis.

4.3) Results:

4.3.1) Amplification of the Entire 16S-23S ITS Region:

The 16S-23S ITS gene for all authentic strains of *Pantoea* and all unidentified South American and Ugandan isolates was amplified. However, more than one amplification product was visible for each of the isolates included in the study. The multiple amplification products were run on several different types of gels to try and improve the separation and resolution of the bands.

4.3.2) Product Separation:

Agarose Gels:

Amplification products were between 500 and 1 000 bp in size and most were bright and clearly defined when separated on 1.5 % agarose gels (Fig. 4.1). Fainter bands were observed between the brighter bands in some cases (Fig. 4.1, Lane 8). Larger, faint bands of ± 1 000 bp were also present for the majority of isolates. The type strain of *P. ananatis* (LMG2665^T) yielded three amplification products, whilst two PCR products were visible for the type strains of *P. agglomerans* (LMG1286^T) and *P. stewartii* subsp. *indologenes* (LMG2632^T).

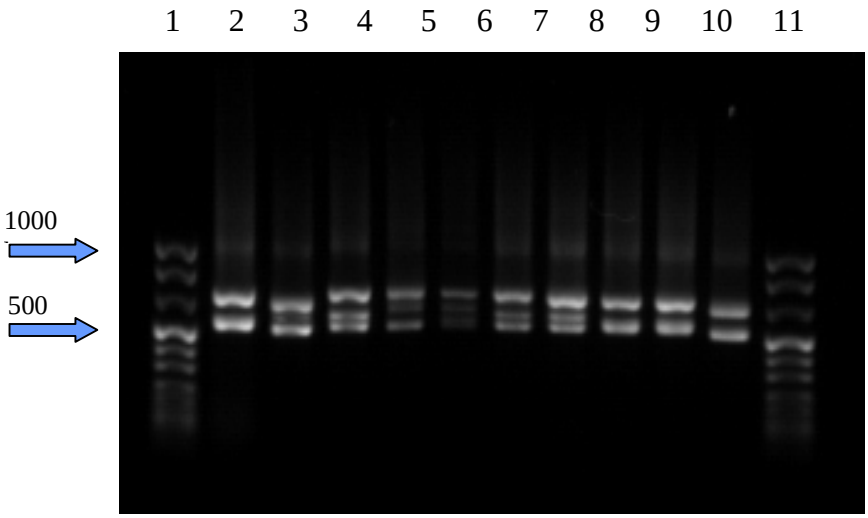


Figure 4.1: Electrophoresis of the entire 16S-23S ITS-PCR assay of authentic *Pantoea ananatis* (BCC095, LMG20106, BCC087, BCC083, LMG2665^T, BCC099), *Pantoea stewartii* subsp. *indologenes* (LMG2632^T) and *Pantoea agglomerans* (LMG1286^T) strains and possible *P. ananatis* isolates (BCC109, BCC003) from South America 1 on a 1.5 % agarose gel

Adjustments to $MgCl_2$ concentration and annealing temperature were made in an attempt to obtain single amplicons. When this failed, the PCR products were run on 3 % agarose gels to improve the resolution. Once separated on a higher percentage gel, there appeared to be between two and five bands for each amplified isolate (Fig. 4.2).

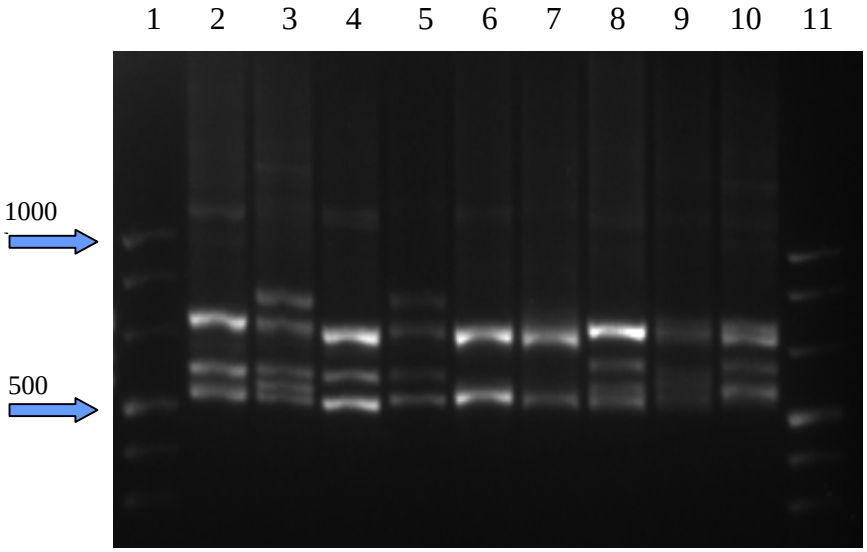


Figure 4.2: Electrophoresis of the entire 16S-23S ITS-PCR of authentic *Pantoea ananatis* (LMG20106, LMG20104, LMG2676, LMG2665^T, LMG20103, LMG2666), *Pantoea stewartii* subsp. *indologenes* (LMG2671, LMG2632^T) and *Pantoea agglomerans* (LMG1286^T) strains on a 3 % agarose gel

5 % Polyacrylamide (PAGE) Gels:

The same PCR products which were run on agarose gels, were separated further on the PAGE gels. The bands ranged from 500 to 2 600 bp in size, but were not clearly defined and most appeared as smears on the gel. The type strain of *P. ananatis* (LMG2665^T), which appeared to produce three to four bands when run on the agarose gels, yielded between six and seven bands when run on a 5 % PAGE gel (Fig. 4.3). The same was true for the type strains of *P. agglomerans* (LMG1286^T) and *P. stewartii* subsp. *stewartii* (LMG2632^T).

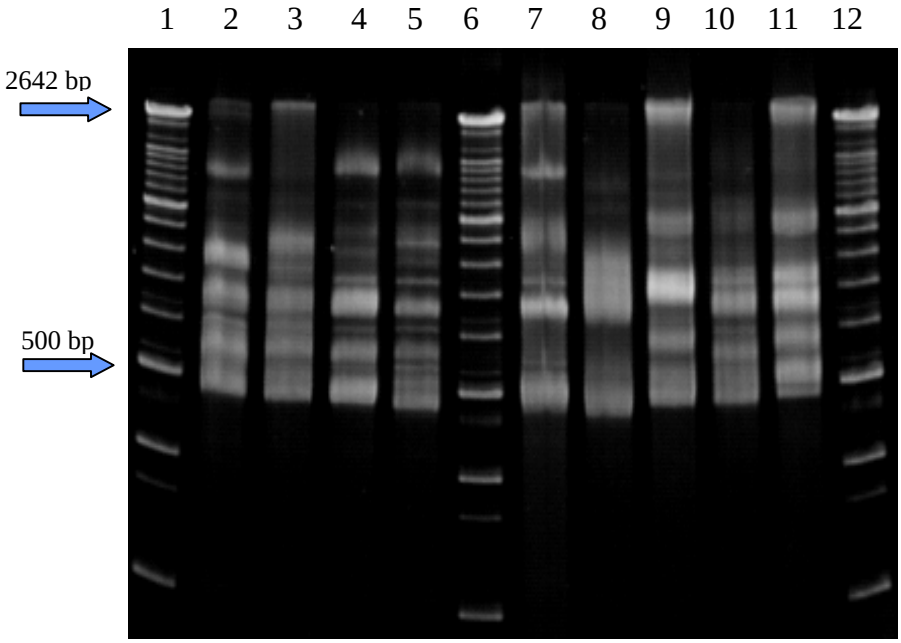


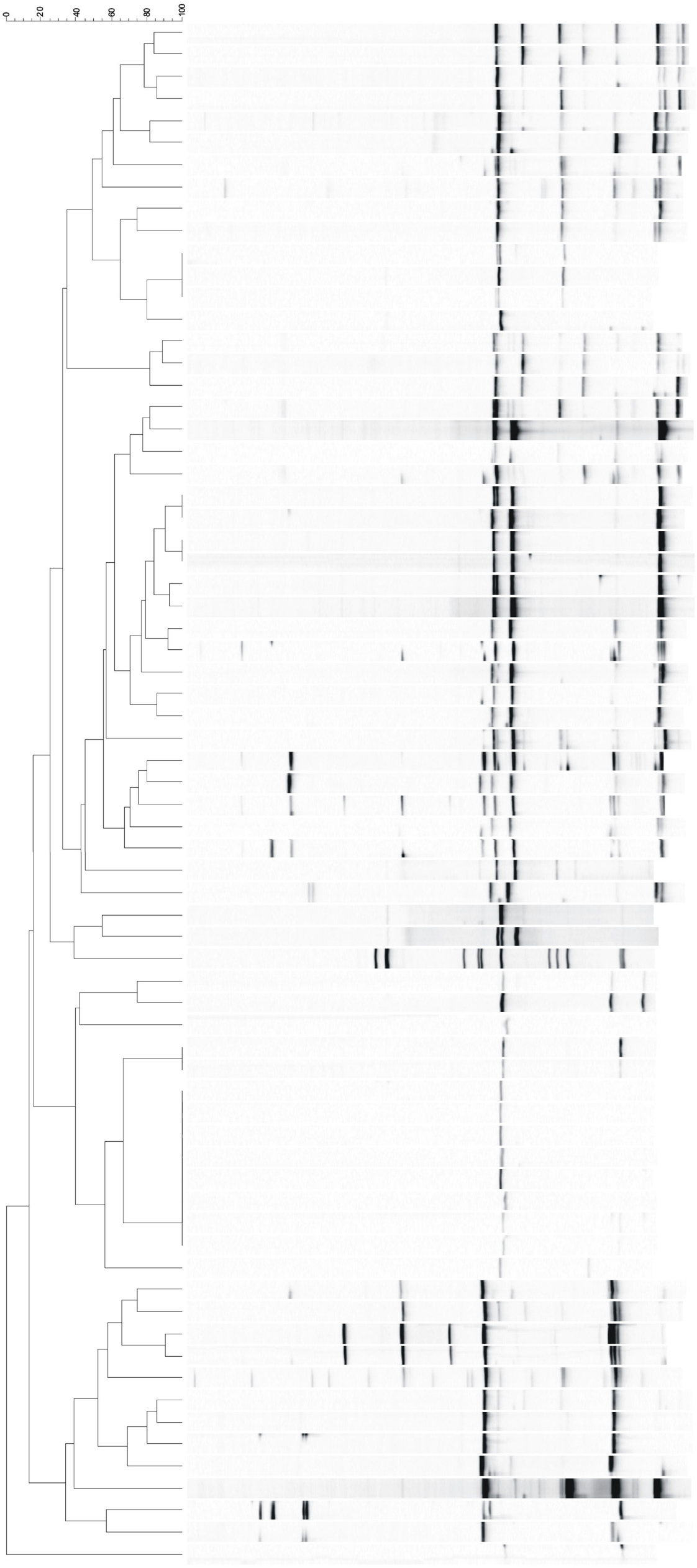
Figure 4.3: Electrophoresis of the entire 16S-23S ITS-PCR of authentic *Pantoea ananatis* (LMG20106, LMG20104, LMG2676, LMG2665^T, LMG20103, LMG2666), *Pantoea stewartii* subsp. *indologenes* (LMG2671, LMG2632^T) and *Pantoea agglomerans* (LMG1286^T) strains on a 5 % PAGE gel

Sequencing Gels:

Clearly defined, multiple amplification products, ranging from 350 to 700+ bp, were visible for all isolates following amplification with a fluorescent primer and separation on an automated sequencing gel. There was a minimum of two bands for each isolate, but not one common-sized band among the isolates. A dendrogram was constructed based on analysis of the band patterns.

Figure 4.4 (on folded insert): UPGMA dendrogram based on band patterns from amplification of the entire 16S-23S ITS regions of species of *Pantoea* and unidentified South American and Ugandan isolates with universal primers FGPS 1490 and FGPL 132. The levels of similarity, representing the Dice similarity co-efficient, are expressed as percentages. The banding patterns adjacent each branch are normalised and background subtracted digitised gel strips processed in GelCompar.

4.4)



Strain	Species			Location	Host
LMG20103	Pantoea	ananatis		Piet Retief	Eucalyptus
LMG2666	Pantoea	ananatis		Hawaii	Pineapple
CTB1061	Pantoea	ananatis		Japan	Rice
CTB1135	Pantoea	ananatis		Japan	Rice
BCC118	Pantoea	ananatis		KZN	Eucalyptus + Con
LMG2632T	Pantoea	stewartii	indologenes	India	Fox millet
LMG20106	Pantoea	ananatis		Tzaneen	Eucalyptus
BCC103	Pantoea	ananatis		USA	Sudangrass
BCC099	Pantoea	ananatis		USA	Sudangrass
LMG2671	Pantoea	stewartii	indologenes	Hawaii	Pineapple
LMG2570	Pantoea	agglomerans		USA	Sorbus sp
LMG2628	Pantoea	ananatis			Banana
LMG2678	Pantoea	ananatis		Zimbabwe	Puccinia graminis
LMG22049T	Pantoea	citrea		Japan	Mandarin orange
BCC083	Pantoea	ananatis		USA	Onion
BCC114	Pantoea	ananatis		KZN	Eucalyptus + Con
LMG2665T	Pantoea	ananatis		Brazil	Pineapple
BCC098	Pantoea	ananatis		USA	Sudangrass
BCC080	?	?		South America 2	Eucalyptus
LMG2676	Pantoea	ananatis		USA	Puccinia graminis
BCC087	Pantoea	ananatis		USA	Onion
BCC077	?	?		South America 2	Eucalyptus
BCC107	?	?		Uganda	Eucalyptus
BCC076	?	?		South America 2	Eucalyptus
BCC082	?	?		South America 2	Eucalyptus
BCC078	?	?		South America 2	Eucalyptus
BCC081	?	?		South America 2	Eucalyptus
BCC015	?	?		South America 1	Eucalyptus
BCC067	?	?		South America 3	Eucalyptus
LMG1286T	Pantoea	agglomerans		Zimbabwe	Human
BCC075	?	?		South America 2	Eucalyptus
BCC105	?	?		Uganda	Eucalyptus
BCC145	Pantoea	agglomerans		RSA	Bean leaf
BCC004	?	?		South America 1	Eucalyptus
BCC008	?	?		South America 2	Eucalyptus
BCC007	?	?		South America 1	Eucalyptus
BCC002	?	?		South America 1	Eucalyptus
BCC013	?	?		South America 2	Eucalyptus
LMG22051T	Pantoea	terrea		Japan	Soil
BCC109	?	?		Uganda	Eucalyptus
LMG22098	Pantoea	punctata		Japan	Persimmon
LMG2660	Pantoea	agglomerans		Japan	Wisteria floribunda
LMG22097	Pantoea	punctata		Japan	Mandarin orange
LMG2604	Pantoea	dispersa		Netherlands	Rosa sp
LMG2718	Pantoea	stewartii	stewartii	USA	Corn
LMG2631	Pantoea	stewartii	indologenes	India	Millet
LMG2674	Pantoea	stewartii	indologenes	Hawaii	Pineapple
LMG2713	Pantoea	stewartii	stewartii	USA	Corn
LMG22050T	Pantoea	punctata		Japan	Mandarin orange
LMG22099	Pantoea	terrea		Japan	Persimmon
LMG22100	Pantoea	terrea		Japan	Persimmon
LMG2602	Pantoea	dispersa		India	Sorghum bicolor
LMG2603T	Pantoea	dispersa		Japan	Soil
LMG2630	Pantoea	stewartii	indologenes		Guar gum powder
LMG2715T	Pantoea	stewartii	stewartii	USA	Corn
LMG2749	Pantoea	dispersa			Human
LMG5342	Pantoea	ananatis		USA	Human
BCC011	?	?		South America 2	Eucalyptus
LMG20104	Pantoea	ananatis		Harding	Eucalyptus
BCC009	?	?		South America 2	Eucalyptus
BCC010	?	?		South America 2	Eucalyptus
BCC095	Pantoea	ananatis		Vryheid	Eucalyptus
BCC150	Pantoea	ananatis		USA	Honeydew melon
BCC154	Pantoea	ananatis		Limpopo	Onion
BCC157	Pantoea	ananatis		Limpopo	Onion
BCC005	?	?		South America 1	Eucalyptus
BCC146	Pantoea	agglomerans		RSA	Onion bulb
BCC072	?	?		South America 2	Eucalyptus
BCC074	?	?		South America 2	Eucalyptus
LMG2673	Pantoea	stewartii	indologenes	Hawaii	Pineapple

4.4) Discussion and Conclusions:

4.4.1) Amplification of the entire 16S-23S ITS Region:

To differentiate among *Pantoea* species, as well as possible *Pantoea ananatis* isolates from South America and Uganda, the entire 16S-23S ITS gene was examined for polymorphisms. The 16S-23S ITS region was amplified for all authentic *Pantoea* strains and unidentified isolates. However, multiple amplification products were observed for all isolates included in the study. These results suggest that the genome of *Pantoea* contains multiple copies of the rRNA operon and therefore, multiple copies of the 16S-23S ITS region.

As universal primers were used to amplify the 16S-23S ITS gene, it was expected that each isolate would yield an amplification product. The presence of multiple amplification products suggests that more than one copy of the 16S-23S ITS region was amplified. This provides a reasonable explanation as *Pantoea* belongs to the family *Enterobacteriaceae* along with *Escherichia coli* and *Salmonella enterica*, both of which contain seven copies of the rRNA operon (Jensen *et al.*, 1993). This was further supported when changes in $MgCl_2$ concentration and an increase in the annealing temperature of the PCR cycle failed to remove the multiple amplicons.

4.4.2) Product Separation:

Agarose Gels:

PCR products were run on a 3 % agarose gel in an attempt to further separate the bands, so a common band could be identified among the isolates. Following electrophoresis, it became clear that a common band among the isolates, notably the type strains of *Pantoea* species, did not exist. It was also not certain if additional amplification products were present for many of the isolates, which would not be visible on a low resolution agarose gel. Poor resolution of band patterns on agarose gels was also observed for

Clostridium difficile (Gürtler, 1993), which was resolved by separating the fragments on long denaturing polyacrylamide gels.

Amplification of the 16S-23S ITS region using universal primers was successful, although more than one amplicon was visible for each isolate. Sequencing or PCR-RFLP analysis of the entire 16S-23S ITS gene was ruled out, as one common band could not be identified among the multiple PCR products. If the resolution of the PCR products could be improved by separation on polyacrylamide gels, it was decided that the band patterns would be treated as DNA fingerprints.

5 % Polyacrylamide (PAGE) Gels:

It was thought that the multiple amplification products would separate further from each other on polyacrylamide gels, as they typically provide a higher resolution than agarose gels. When the polyacrylamide gels were run and stained successfully, the resolution of the PCR products was poor. The 2 000 bp marker separated very well and each size standard was clearly defined. However, the amplification products were smeared and it was impossible to distinguish a clear band pattern. Six to seven bands were visible for the type strain of *P. ananatis* (LMG2665^{T}) and for other *P. ananatis* isolates (LMG20106, LMG20104, LMG2676 and LMG2666) (Fig. 4.2). However the bands did not form a definite pattern when compared to the type strain. Isolate LMG2666 from Hawaii (Lane 11), yielded the band pattern that was most similar to that of the type strain of *P. ananatis* (Lane 9). Although this was not conclusive, as the amplification products were not clearly defined. The resolution of the gels could possibly be improved by using a higher percentage of acrylamide or adjusting the running conditions. However, due to the time-consuming nature of the preparation of polyacrylamide gels, and problems encountered with the staining of the gels, it was decided to discontinue attempts to separate the multiple PCR products on these gels.

Sequencing Gels:

The LI-COR IR^2 automated sequencer provided a rapid and efficient means of separating the multiple amplification products from the entire 16S-23S ITS-PCR assay. Although the cost per gel is relatively high, the technique is simple and the results can be easily interpreted. The amplification products separated well, were clearly defined and very little smearing was visible. Therefore, the multiple PCR products could be treated as DNA fingerprints and were analysed as such.

The UPGMA dendrogram that was constructed, following analysis of the band patterns, revealed little concerning the genetic relatedness of *Pantoea* species and the unidentified isolates. The majority of *P. ananatis* isolates are contained within one large cluster. However, interspersed among the *P. ananatis* isolates are the type strains of *P. stewartii* subsp. *indologenes* (LMG2632^T), *P. citrea* (LMG22049^T) and a *P. agglomerans* isolate (LMG2570). The *P. agglomerans* isolate (LMG2570) produced a positive PCR product when amplified with *P. ananatis* species-specific primers (Chapter 3). The fact that it now groups with *P. ananatis* isolates in the dendrogram with a similarity value of 100 % confirms that LMG2570 was misidentified and is *P. ananatis*, not *P. agglomerans*. The type strain of *P. citrea* (LMG22049^T) was the only isolate of its species to be included in the assay and was expected to cluster with *P. punctata*, as both were isolated from mandarin orange. Instead, LMG22049^T groups with *P. ananatis* isolates with a homology of 80 %.

Several different banding patterns are visible for the *P. ananatis* isolates, although the majority do share at least one common-sized fragment. This intraspecies variability has also been observed for strains of *Staphylococcus aureus*, *Salmonella enterica* and *Escherichia coli* (Pérez-Luz *et al*., 2004). These heterogeneous fragment profiles were correlated with species containing a high copy number of ribosomal operons.

It is unusual that the type strain of *P. stewartii* subsp. *indologenes* (LMG2632^T) falls within the *P. ananatis* cluster with a similarity value of 80 %, especially as the remaining *P. stewartii* subsp. *indologenes* and *P. stewartii* subsp. *stewartii* isolates group together in a cluster with *P. punctata*, *P. terrea* and *P. dispersa*. This suggests that the multiple copies of the rRNA operon are highly similar between certain strains of *Pantoea* species and that the 16S-23S ITS region is not representative of the genotypic differences among *Pantoea* species. This is further illustrated by the large group of *P. punctata*, *P. terrea* and *P. dispersa* isolates which cluster with *P. stewartii* subsp. *indologenes* (LMG2630) and the type strain of *P. stewartii* subsp. *stewartii* (LMG2715^T) with a similarity value of 100 %. All of these isolates seem to possess only one copy of the rRNA operon. Similar behaviour has been observed for *Shigella dysenteriae* and *Shigella sonnei*, both of which contain only two copies of the rRNA operon, where the electrophoretic profiles of the 16S-23S ITS regions of both these species were identical (Pérez-Luz *et al.*, 2004).

It is not possible to distinguish *Pantoea* isolates from each other at or below the species level, based on DNA fingerprints of the 16S-23S ITS region. This is illustrated by the clustering of *P. stewartii* subsp. *stewartii* and *P. stewartii* subsp. *indologenes* with a similarity value of 100 %. The *P. agglomerans* isolates are scattered throughout the dendrogram, as are the unidentified isolates from Uganda and South American countries 1, 2, and 3. A number of the unidentified isolates from South America 1 group with the type strain of *P. agglomerans* (LMG1286^T) with a homology of 80 %, whilst others from South America 1 and 2 cluster with the type strain of *P. terrea* (LMG22051^T) or with a smaller cluster of *P. ananatis* isolates (LMG20104, BCC095, BCC150, BCC154, BCC157).

Although LI-COR automated sequencing gels provided excellent separation of the multiple amplification products, typing using the 16S-23S ITS gene products did not present an accurate portrayal of the genetic relatedness of the genus *Pantoea*. The identity of the Ugandan and South American isolates is still uncertain, as no conclusions can be drawn from the manner in which they clustered in the dendrogram. A high degree of similarity appears to exist among the rRNA operons of species of the genus *Pantoea*,

while heterogeneity was visible between strains of *Pantoea ananatis*. The same behaviour was observed for *Staphylococcus aureus*, where the high level of intraspecies variation was not observed for other *Staphylococcus* species, indicating that it is possible for a single species to produce a variety of spacer amplification products (Jensen *et al.*, 1993). If the genetic relatedness between *Pantoea* species and the unidentified isolates is to be examined successfully, the typing technique will have to be based on screening of the entire genome.