The Epidemiology and Virulence Factors of *Pneumocystis*

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Helen Elizabeth Ambrose

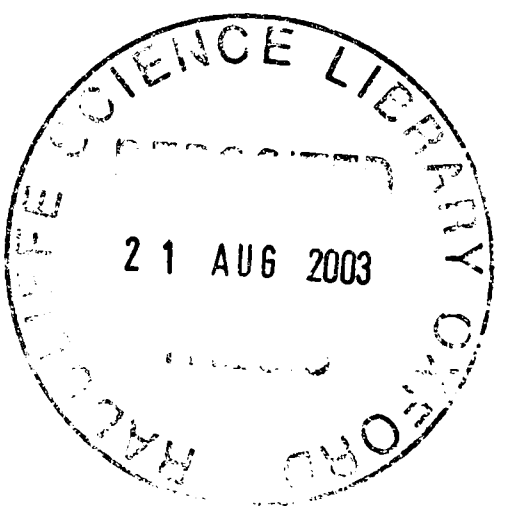
Department of Paediatrics

and

Wolfson College

University of Oxford

Hilary Term, 2003



Abstract **Helen Elizabeth Ambrose**

The Epidemiology and Virulence Factors of *Pneumocystis*

Presented for the degree of Doctor of Philosophy, Hilary Term, 2003
Department of Paediatrics and Wolfson College, University of Oxford

Pneumocystis is a diverse group of fungi found in the lungs of mammals, and can cause a fatal pneumonia in immunosuppressed humans. The project focussed on two main areas of *Pneumocystis* biology. The first part investigated the transmission of *P. jiroveci* between humans and comprised three studies. The first study demonstrated direct transmission between a mother and her infant who had *Pneumocystis* pneumonia, PCP, contemporaneously. The second investigated transmission of *P. jiroveci* between health care workers and patients with PCP. *P. jiroveci* was identified in HCW samples but the genotyping data did not demonstrate direct transmission between the HCW and the patients. The third study examined a representative distribution of *P. jiroveci* genotypes within a whole human lung and identified two foci of infection.

The second main area of research investigated *PRT1*, a multigene family, encoding a protein with homology to KEX-like proteases. The first part of this study attempted to identify the original single-copy *PRT1* gene that encodes the progenitor kexin-like protease. Inconclusive sequence data was obtained. The second part of this study examined regulation of *PRT1* gene expression, initially using an assay that characterised the length polymorphisms within the proline rich region of the *PRT1* gene. In the complex populations of rat-derived *Pneumocystis* used it was difficult to distinguish reliable differences between the cDNA (expressed) and genomic DNA proline rich region profiles.

Subsequent investigations used simple rat-derived *Pneumocystis* populations that were derived from inocula of 10 *Pneumocystis* organisms intratracheally injected into nude rats. The catalytic domain of *PRT1* was sequenced from cDNA and genomic copies from each of the five low-dose populations. Two majority sequence types were identified, one genomic and the other cDNA. Three of the populations contained a cDNA majority sequence type, implying that some form of regulation of expression was occurring. This hypothesis was further investigated using a second set of simple populations, hybridisation techniques and 5' RACE.

This thesis is dedicated to my supervisor

Ann E. Wakefield

(1948 – 2002)

whose endless encouragement and enthusiasm

helped me through all of this D.Phil.

Acknowledgements

I am indebted to both Derek Hood and Rob Miller for picking up the pieces of my D.Phil in the past year, Derek for warning me of the perils of the nature of the beast and Rob for refusing to let me turn into a Jeremiah. I would never have reached this point without them. I am also indebted to Mary D for knowing the answer to every non-*Pneumocystis* question and to Ann Harris for her help in untangling the administration of this D.Phil.

From the Oxford *Pneumocystis* group, I must thank Rob Palmer for being so patient in my first year and for answering all my basic lab questions and a huge thank you to Austin Lindley for keeping the second and third year light-hearted and generally for being a good coffee and beer partner. In the world of *Pneumocystis*, Sergio Vargas and Carolina Ponce made my time in Chile very welcoming and fun. I can't wait to return and I am grateful to Jim Stringer for helping me publish the *PRT1* data.

Both Rebekah Price and Ed Mee in their own ways have been two outstanding friends, even though I now know far more about Liverpool Football Club than I ever wanted or needed to. A huge thanks also goes to all the members of the Department of Paediatrics, both past and present for making the WIMM such a fun and enjoyable place to work in.

Outside the lab, I am very grateful to Pam Rowling, the mother inferior, Mike Steward and Sam Powell for encouraging me to start my D.Phil and persuading me that I could do it. The Stratfield Road Brigade, Phil Davis, Liliana Costa, Kristie Johnson, Rebecca Holmes and especially Sarah Dutton were all good friends (but pathetic washers up) and were always on hand for a cup of tea or a swift pint. And Dave Eyre for his friendship and for giving good advice.

Big thanks go to my family whose numbers have fluctuated somewhat in the past three years. They have continually supported and encouraged me and never once queried the attraction of studying a fungus for three years. And finally the biggest thanks go to my boyfriend Paul for being such a good friend and for understanding when a gin and hug were needed.

INDEX OF CONTENTS

1 Introduction 1

1.1. What is Pneumocystis? 1

1.1.1. The History of *Pneumocystis* 1

1.1.2. *Pneumocystis*: Fungus or Protozoa? 2

1.1.3. Host Specificity of *Pneumocystis* 2

1.1.4. Levels of Genetic Heterogeneity 4

1.1.5. Links to Nomenclature 6

1.1.6. Life cycle of *Pneumocystis* 6

1.1.7. Primary Infection of *Pneumocystis* 8

1.1.8. *Pneumocystis jiroveci* causes *Pneumocystis* Pneumonia 8

1.2. The Study of *Pneumocystis* 9

1.2.1. Culture System 9

1.2.2. Animal Models 10

1.2.3. Molecular Techniques to Study *Pneumocystis* 11

1.3. Epidemiology of *Pneumocystis* 12

1.3.1. Genotyping of *Pneumocystis jiroveci* 12

1.3.2. Latency versus Reinfection 15

1.3.3. Transmission of *P. jiroveci* 17

1.3.4. Colonisation and Carriage of *P. jiroveci* 18

1.4. Multigene Families in *Pneumocystis* 20

1.4.1. Multigene Families within the Genome of *Pneumocystis* 20

1.4.2. Major Surface Glycoprotein 22

1.4.3. Protease1 26

1.5. Aims of Project 32

2 Materials and Methods 34

2.1. Units, Abbreviations and Acronyms 34

2.2. Samples and Reagents 36

2.2.1. Chemicals 37

2.2.2. Enzymes 38

2.2.3. Standard Solutions 38

2.2.4. Bacterial Media 39

2.3. Clinical Samples 39

2.3.1. Ethical Guidelines 39

2.3.2. Mother and Infant Samples 39

2.3.3. Health Care Worker Samples 41

2.3.4. Patient Samples 41

2.3.5. Post-mortem whole lung tissue sample 42

2.4. Techniques to Obtain Clinical Samples 43

2.4.1. Bronchoalveolar lavage (BAL) samples 43

2.4.2. Induced Sputum (IS) samples 43

2.4.3. Oropharyngeal (OP) samples 43

2.4.4. Nasal Rinse Samples 43

2.5. Extraction of *Pneumocystis jiroveci* DNA from Human Samples 44

2.6. The rat model of *Pneumocystis* Infection 45

2.6.1. Source of Rat-derived Simple *Pneumocystis* Populations 45

2.6.2. Other Rat-derived *Pneumocystis* Samples 46

2.6.3. Rat-derived *Pneumocystis* DNA Libraries 46

2.6.4. Rat-derived *Pneumocystis* Parasite Purification from Rat Lung Tissue 46

2.6.5. DNA Extraction from Rat-derived *Pneumocystis* Organisms 47

2.7. Amplification of *Pneumocystis* DNA Sequences 48

2.7.1. PCR Amplification of the *Pneumocystis jiroveci* Genotyping Loci 49

2.8.. Cloning of Amplified DNA into Plasmid Vectors, PCR Screen and Plasmid Purification 53

2.9. DNA Sequencing protocols 54

2.9.1. Genescanning 54

2.9.2. Sequencing by the Dideoxy Chain Termination Method 54

2.9.3. Sequence Analysis 55
2.9.4. Synonymous versus Non-synonymous Nucleotide Changes 56
2.10. BAL-31 Nuclease digests of rat-derived *Pneumocystis* Genomic DNA 56
2.11. Southern Blotting and Membrane Hybridisation 57
2.11.1. Transfer of DNA and Binding of Probes 57
2.11.2. Stripping of Southern Blots 59
2.11.3. Dot Blots 59
2.12. RNA Protocols 59
2.12.1. RNA Extraction from Rat-derived *Pneumocystis* Parasites 60
2.12.2. DNase Treatment of RNA 60
2.12.3. Reverse Transcriptase Polymerase Chain Reaction, RT-PCR 61
2.12.4. 5' RACE Amplification of cDNA Ends 61

3 Epidemiological Studies to Investigate the Transmission of *Pneumocystis jiroveci* 63
3.1.1. Probable Mother-to-Infant Transmission of *Pneumocystis jiroveci* Infection 63
3.1.2. *Pneumocystis jiroveci* DNA in Immunocompetent Health Care Workers in Contact with Patients with *Pneumocystis* Pneumonia 72
3.1.3. Distribution of *Pneumocystis jiroveci* Types in the Lung of a Child Dying of *Pneumocystis* Pneumonia 77
3.2. Discussion 79
3.2.1. Probable Mother-to-Infant Transmission of *Pneumocystis jiroveci* Infection 79
3.2.2. *Pneumocystis jiroveci* DNA in Immunocompetent Health Care Workers in Contact with Patients with *Pneumocystis* Pneumonia. 83
3.2.3. Distribution of *Pneumocystis jiroveci* Types in the Lung of a Child Dying of *Pneumocystis* Pneumonia 87
3.2.4. Overall Discussion 89

4 Analysis of Telomeric and Non-telomeric Copies of *PRT1* 93
4.1.1. Southern Hybridisations to Identify Non-telomeric Copies of *PRT1* 93
4.1.2. Further Investigation of Non-telomeric Copies of *PRT1* 96
4.2. Discussion 103

5 Characterisation of the Complex Pattern and the Possible Regulation of *PRT1* Expression within the Context of Different Populations of Rat-Derived *Pneumocystis* 107
5.1.1. Visualisation of PRR Amplification Products 107
5.1.2. Template Choice for the PRR Assay Control Experiments. 108
5.1.3. Primer Design for the PRR Assay 111
5.1.4. Adapting the PRR assay to Characterise cDNA copies of the PRR 113
5.1.5. Genescan and Proline Rich Region PCR Reproducibility. 117
5.1.6. Comparison of PRR DNA Profiles from Rat-Derived *Pneumocystis* Genomic DNA, Extracted from Populations from the Same Rat Colony 118
5.1.7. Comparison of PRR Profiles Amplified from cDNA and DNA Extracted from the Same Rat - Derived *Pneumocystis* Population 123
5.1.8. Cloning and Sequencing of PRR Amplification Products 127
5.1.9. Assessment of PRR Assay 134
5.2. Discussion 137

6 Characterisation of the Complex Pattern and the Possible Regulation of *PRT1* Gene Expression within the Context of Simple Populations of Rat-Derived *Pneumocystis*.....140
6.1.1. Simple Populations of Rat-Derived *Pneumocystis* 140
6.1.2. *PRT1* Assay 141
6.1.3. Clone Summary 142
6.1.4. Phylogenetic Trees for Individual Populations 145
6.1.5. The cDNA majority sequence type..... 148
6.1.6. Published Catalytic Domain Sequences 150
6.1.7. Genomic Majority Sequence Types..... 150
6.1.8. Phylogenetic Analysis of Different Combinations of Sequences from the 209 Clones 150
6.1.9. Amino Acid and Tertiary Structure of the 14 Clades 154
6.1.10. SNAP analysis..... 161
6.2. Discussion.....164
6.2.1. Does Low Dose Inoculation Result in Simple Populations? 164
6.2.2 Phylogenetic Analysis of Sequences of the Catalytic Domain Clones 170
6.2.3 The Majority Sequence Types 171
6.2.4 SNAP Analysis..... 173
6.2.5 Limitations of the *PRT1* Assay 173
6.2.6 Conclusion..... 175

7 Further Characterisation of the Complex Pattern and the Possible Regulation of *PRT1* Gene Expression using Simple Populations and 5' RACE.....176
7.1. Second Passage Populations.....176
7.2. Discussion of Second Passage Populations.....185
7.3. Results of DNA Hybridisation Experiments to Characterise the Majority Sequence Types. 188
7.3.1. Probe Design 188
7.3.2. Dot Blots to Test the Specific Probes 190
7.3.3. Using the Specific Probes as Primers in PCR amplification. 194
7.3.4. Hybridisation of Specific Probes to Complex Rat- Derived *Pneumocystis* DNA. 197
7.4. Discussion of DNA Hybridisation Experiments to Characterise the Majority Sequence Types.....203
7.5. Characterisation of 5'UTR of *PRT1* Genes using 5' RACE 205
7.5.1. Sequence Data for the 5'UTR of *PRT1* Genes..... 205
7.5.2. Hybridisation to Characterise the 5'UTR of *PRT1* Genes 210
7.6. Discussion of Characterisation of 5'UTR of *PRT1* Genes using 5' RACE 215
7.7. Summary of the Further Characterisation of *PRT1* Gene Expression 216

8 Final Discussion 218
8.1. Epidemiology of *Pneumocystis jiroveci* 218
8.2. Protease1 222
8.2.1. *PRT1* Gene Expression is Regulated 224
8.2.2. Co-expression of *PRT1* genes and *MSG* genes..... ... 228
8.2.3. *PRT1* genes may be Antigenically Variable 229

References.....233

INDEX OF FIGURES

Figure 1.1. Proposed life cycle of *Pneumocystis* 7
Figure 1.2. A schematic diagram of the DNA sequence and proposed protein structure for a *P. carinii PRT1* gene. 28
Figure 3.1. Electrophoretic separation of PCR amplification products from the mitochondrial large subunit, mt LSU, using primers pAZ102X and pAZ102Y. 64
Figure 3.2. Electrophoretic separation of PCR amplification products from the dihydropterate synthase gene, DHPS, using primers DHPS3 and DHPS4. 67
Figure 3.3. Electrophoretic separation of PCR amplification products from the internal transcribed spacer (ITS) regions of the nuclear rRNA, using primers ITS2F3 and ITS2R3. 69
Figure 4.1. Southern hybridisation analysis of the sensitivity of rat -derived *Pneumocystis* genomic DNA to BAL-31 digestion. 94
Figure 4.2. Agarose gel of *Pst*1 digested rat-derived *Pneumocystis* genomic DNA. 98
Figure 4.3. Electrophoretic separation of DNA amplification products using the subdivided excised agarose blocks as templates. 98
Figure 4.4. Amino acid alignment of the catalytic domain of different copies of published *PRT1* sequences compared with the catalytic domains of two representative potential progenitor single copy *PRT1* gene sequences 100
Figure 4.5. Amino acid alignment of the 5' domain of different copies of published *PRT1* sequences compared with the 5' domain of a representative potential progenitor single copy *PRT1* gene sequence. 102
Figure 5.1. Electrophoretic separation of products from a PCR amplification using the primers PcProt19 and PcProt57. 109
Figure 5.2. Electrophoretic separation of products from the PCR amplifications using different *PRT1* templates with different combinations of proline rich region specific primers. 112
Figure 5.3. ClustalW alignment of the beginning of the PRR of published *PRT1* sequences with the new PRR primers marked. 114
Figure 5.4 (a). Electrophoretic separation of RT-PCR products, using RNA treated with different DNases. (b). Electrophoretic separation of RT-PCR products using RNA at different dilutions. 116
Figure 5.5. Electrophoretic separation of products from two PCR amplifications using the primers PcProt65 and PcProt57-1, (PCRs 5 and 6). 120
Figure 5.6. Electrophoretic separation of products from the PCR amplification, using different rat-derived *Pneumocystis* genomic DNA preparations and the primers PcProt65 and PcProt57 -1. 124
Figure 5.7. Electrophoretic separation of RT-PCR products. 126
Figure 5.8. Extraction of PRR products from gels afte r the PCR amplification 128
Figure 5.9. ClustalW sequence alignment of the 440 bp product from the PCR amplification of rat -derived *Pneumocystis* genomic DNA with PcProt19 and PcProt57. 130
Figure 5.10. ClustalW sequence alignment of the ≈ 600 bp product from the PCR amplification of rat -derived *Pneumocystis* genomic DNA with PcProt19 and PcProt57 132
Figure 5.11. ClustalW sequence alignment of the ≈ 600 bp product from the PCR amplification of rat -derived *Pneumocystis* genomic DNA with PcProt19 and PcProt57. 135
Figure 6.1. Electrophoretic separation of the products of an RT- PCR amplification using the primers PcProt1 and PcProt3. 143
Figure 6.2. Individual phylogenetic trees for *PRT1* sequences from populations EA15, EA17, EA19 and EA21.1. 146
Figure 6.3. Individual phylogenetic trees for *PRT1* sequences from population EA21.2, EA24, inoculum and 1984 147
Figure 6.4. Phylogenetic trees for the copies of the *PRT1* cDNA majority sequence type. 149
Figure 6.5. Phylogenetic tree comparing *PRT1* sequences from population EA21 151
Figure 6.6. Frequency distribution of *PRT1* catalytic domain sequences. 153
Figure 6.7. Phylogenetic tree from the nucleotide sequence alignment of the 14 clades. 155
Figure 6.8. Phylogenetic tree from the amino acid sequence alignment of the 14 clades. 156
Figure 6.9. Amino acid alignment of different copies of *PRT1* compared with other protease sequences. 157
Figure 6.10. Scatter graphs of results of SNAP analysis for all clones from 14 clades. 163
Figure 6.11. Cumulative behaviour of synonymous and non-synonymous changes codon by codon, across *PRT1* catalytic domain. 165
Figure 7.1. Phylogenetic tree comparing the sequence of cDNA copies from populations 714 and EA21 180
Figure 7.2. Phylogenetic tree comparing the sequence of cDNA copies of populations 723 and 716. 182
Figure 7.3. Phylogenetic trees comparing nucleotide and amino acid sequences of the cDNA majority sequence type with equivalent clones from the second passage populations 714 and 723. 184
Figure 7.4. ClustalW alignment of the first 50 bp of a representative group of 40 *PRT1* catalytic domain sequences 189

Figure 7.5. (a) Dot blot hybridisation demonstrating the different levels of efficiency of hybridisation between PCR products and plasmid DNA. (b) Dot blot hybridisation for *PRT1* catalytic domain clones using specific probes. 191
Figure 7.6. Dot blot hybridisation for 160 *PRT1* catalytic domain clones using cDNA specific probe. 193
Figure 7.7. (a) Electrophoretic separation of the products of PCR amplification using genomic rat-derived *Pneumocystis* DNA and the primers HA1 and PcProt3. (b) Electrophoretic separation of the products of a PCR amplification using bacterial clones of the *PRT1* catalytic domain as templates and the primers HA3 and PcProt3. 195
Figure 7.8. Southern hybridisation analysis of digested rat- derived *Pneumocystis* genomic DNA probed with an oligonucleotide and a PCR amplification product. 199
Figure 7.9. Southern hybridisation analysis of digested rat- derived *Pneumocystis* genomic DNA probed with geno and PcProt3. 200
Figure 7.10. Southern hybridisation analysis of digested rat -derived *Pneumocystis* genomic DNA probed with cDNA and PcProt3. 201
Figure 7.11. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with MID and PcProt3. 202
Figure 7.12. Electrophoretic separation of 5' RACE products. 206
Figure 7.13. ClustalW sequence alignment of 5'UTR of published *PRT1* sequences, 5' RACE products and full length EST sequences. 208
Figure 7.14. Dot blots after hybridisation with probes specific to the *PRT1* 5'UTR. 213
Figure 7.15. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with HA4 and HA5 214

INDEX OF TABLES

Table 1.1. Host species specificity of different types of *Pneumocystis* revealed by cross- infection experiments 3
Table 2.1. Units, abbreviations and acronyms 34
Table 2.2. Samples donated by other researchers 36
Table 2.3. Work done by other researchers or collaborators 37
Table 2.4. Oligonucleotide Primers Used 50
Table 3.1. Results of the genotyping analysis at three loci performed on samples from mother and infant. 66
Table 3.2. ITS1 sequence types of *Pneumocystis jiroveci*. 70
Table 3.3. ITS2 sequence types of *Pneumocystis jiroveci* 71
Table 3.4. Presence of *Pneumocystis jiroveci* DNA in respiratory samples from HCW. 74
Table 3.5. ITS genotypes of *Pneumocystis jiroveci* isolated from asymptomatic HCW and p atients with PCP. 75
Table 3.6. Detection of *Pneumocystis jiroveci* DNA and subsequent results from ITS genotyping in two asymptomatic HCW who performed bronchoscopies on patients with PCP. 78
Table 3.7. Distribution of *Pneumocystis jiroveci* genotypes in the whole lung tissue. 80
Table 5.1. Description of the eight templates for the PRR PCR. 110
Table 5.2. Specificity of primers to different *PRT1* clones. 110
Table 5.3. Numbering of the eight different PCR amplifications with the four primer sets. 118
Table 5.4. Genescan results for the two PCR amplifications using the primers PcProt65 and PcProt57-1. 121
Table 5.5. Numbering of the eight different RT-PCR amplifications with two primer pairs and four RNA preparations from rat-derived *Pneumocystis* populations 125
Table 6.1. Summary of the clones analysed for each population. 144
Table 6.2. Averages for ds, dn and ds/dn values and interquartile ranges for ds, dn and ds/dn for the simple populations. 162
Table 6.3. Number and types of UCS/MSG sequences found in inoculated nude rats 166
Table 7.1. cDNA clones from populations 714, 716 and 723. 179
Table 7.2. Comparison of the clusters found in the two phylogenetic trees. 183
Table 7.3. Average values for ds, dn and ds/dn ratios for the populations 714, 716 and 723 and the corresponding interquartile ranges. 185
Table 7.4. The explanation for binding of the different specific probes to semi-homologous clones. 192
Table 7.5. PCR amplification of plasmid clones amplified by PcProt3 with PcProt1, HA1, HA2 or HA3. 196
Table 7.6. Clones used in the 5'UT R dot blots. ..... .212
Table 8.1. Majority *MSG* and *PRT1* types within each of the simple populations 229

1. Introduction

1.1. What is Pneumocystis?

1.1.1. The History of *Pneumocystis*

Pneumocystis is an opportunistic pathogen that causes *Pneumocystis* Pneumonia (PCP), in the immunosuppressed, for example transplant and HIV patients. *Pneumocystis* was first identified in 1909 by Chagas (Hughes, 1998b) who recognized it as a type of trypanosome. Three years later Delanoe and Delanoe described it as a separate type of microbe. In Europe in the late 1930s and 1940s, an epidemic of interstitial plasma cell pneumonia occurred in premature and malnourished children, particularly those in orphanages. With social and economic improvements after World War II, the epidemic petered out. It was not until 1951 that it was shown to cause human disease (Hughes, 1998b). A new epidemic of *Pneumocystis* was described in the 1960s in children with congenital defects of the immune system and in children and adults with acquired defects secondary to malignancy or its treatment. PCP was also a high risk factor with immunosuppression after organ transplantation. With the introduction of prophylaxis, case rates in these populations fell.

The AIDS pandemic marked the beginning of the current impact of the disease and research into the pathogen increased as PCP is one of the main AIDS defining diseases. Now, with the widespread use of prophylaxis, PCP in HIV-infected patients is seen largely in those unaware of their HIV serostatus at presentation or in those non-compliant with or intolerant of prophylaxis and antiretroviral therapy (Day *et al.*, 1999). In the non-AIDS immunosuppressed patient, *Pneumocystis* remains a significant opportunistic pathogen (Miller, 1999b).

1.1.2. *Pneumocystis*: Fungus or Protozoan?

For many years the organism was considered to be a protozoan for the following reasons,

- strong similarities with protozoan morphology and host pathology
- absence of some phenotypic features typical of fungi
- ineffectiveness of antifungal drugs
- effectiveness of drugs generally used to treat protozoa

Phylogenetic evidence proved that *Pneumocystis* was in fact a fungus rather than a protozoan. The 16S rRNA gene was the first to be sequenced (Stringer *et al.*, 1989, Edman *et al.*, 1988). This was followed by the thymidylate synthase gene (Edman *et al.*, 1989) and 6.8 kb of the mitochondrial DNA (Pixley *et al.*, 1991). In addition, the translation elongation factor 3, EF3, which is only ever found in fungi, was identified in *Pneumocystis* (Ypma-Wong *et al.*, 1992). *Pneumocystis* has been placed within the Class Archiascomycetes in the Phylum Ascomycota (Guarro *et al.*, 1999). Yet, there are still some differences that make *Pneumocystis* distinct from other fungi: the refusal of the organism to grow *in vitro*, fragile cell wall, lack of ergosterol and resistance to antifungal agents. Horizontal gene transfer between protozoa and *Pneumocystis* would explain the similarities between the two organisms (Stringer, 1993b).

1.1.3. Host Specificity of *Pneumocystis*

Pneumocystis is detected in a wide range of mammalian hosts, from the wild small mammals of California (Laakkonen *et al.*, 2001) to Cavalier King Charles Spaniels (English *et al.*, 2001). *Pneumocystis* organisms from different host species differ both phenotypically and genetically. This phenomenon was first identified in a study that showed that the serological response against *Pneumocystis* in rats and humans was different (Frenkel *et al.*, 1966). Most studies have shown that different forms of *Pneumocystis* could not be transmitted from one host species to another. The majority of the studies are summarised in Table 1.1. One study,

Origin of *Pneumocystis* inoculum	Inoculation Route	Recipient hosts	*Pneumocystis* detected	Detection Methods[a]	References
Mouse	Intranasal	SCID mice	Yes	ME, PCR[b]	(Aliouat *et al.*, 1993), (Aliouat *et al.*, 1994)
Mouse	Intratracheal	SCID mice	Yes	ME, PCR, Mabs	(Gigliotti *et al.*, 1993)
Mouse	Intratracheal	Nude rats	No	ME, PCR[b]	(Aliouat *et al.*, 1994)
Rat	Intratracheal	Nude rats	Yes	ME, PCR[b]	(Aliouat *et al.*, 1994)
Rat	Intranasal	SCID mice	No	ME, PCR[b]	(Aliouat *et al.*, 1993), (Aliouat *et al.*, 1994)
Rabbit	Intranasal	SCID mice	No	ME, PCR[b]	(Aliouat *et al.*, 1993), (Aliouat *et al.*, 1994)
Rabbit	Intratracheal	Nude rats	No	ME, PCR[b]	(Aliouat *et al.*, 1994)
Ferret	Intratracheal	SCID mice	No	ME, PCR, Mabs	(Gigliotti *et al.*, 1993)
Rhesus monkey	Intranasal	SCID mice	No	ME	(Furuta *et al.*, 1993)
Rhesus monkey	Intranasal	Nude rats	No	ME	(Furuta *et al.*, 1993)
Human	Intratracheal	SD rats[c]	No	ME, PCR	(Atzori *et al.*, 1999)
Human	Intranasal	SCID mouse	No	ME, PCR	(Durand-Joly *et al.*, 2002)

[a] ME, microscopical examination; Mabs, *Pneumocystis*-specific monoclonal antibodies, PCR was used for confirming the absence of the parasite in uninfected hosts and for identifying *Pneumocystis* developing in positive hosts by using specific primers or hybridisation with specific DNA probes.
[b] Aliouat et al., unpublished data.
[c] In this experiment, recipient animals were already parasitised with *Pneumocystis* before being inoculated with human -derived *Pneumocystis* organisms.

Table 1.1. Host species specificity of different types of *Pneumocystis* revealed by cross-infection experiments. Data from (Durand-Joly *et al.*, 2002).

for example, attempted to infect SCID mice with ferret-derived *Pneumocystis* (Gigliotti *et al.,* 1993). The SCID mice that were intratracheally inoculated with ferret-derived *Pneumocystis* did not develop a *Pneumocystis* infection. As a positive control, mouse-derived *Pneumocystis* was used to infect SCID mice and here *Pneumocystis* infection occurred. Those that did demonstrate cross-species transmission failed to show that the recipient animals were *Pneumocystis* free, prior to exposure or inoculation. In a more recent and rigorous study SCID mice were inoculated intranasally with either human-derived *Pneumocystis* isolates or mouse-derived *Pneumocystis* isolates (Durand-Joly *et al.,* 2002). Only mice inoculated with mouse-derived *Pneumocystis* developed *Pneumocystis* infection as assessed by microscopy and PCR amplification analysis. Overall, these results suggest that *Pneumocystis* is host specific and that humans do not contract pneumocystosis from a known animal reservoir.

1.1.4. Levels of Genetic Heterogeneity

The genetic heterogeneity amongst different *Pneumocystis* genomes can be divided into three levels of divergence (Stringer, 1996a). The lowest level of divergence, termed class I, was seen between isolates of *Pneumocystis* from the same host species, for example different genotypes of human-derived *Pneumocystis*. An intermediate level of divergence, termed class II, was observed between rat isolates (Cushion *et al.,* 1993b). Two types of *Pneumocystis,* defined as special forms, have been detected by karyotype analysis in rat lungs. The highest level of divergence, termed class III, was observed between isolates from different host species, for example between human-derived *Pneumocystis* and rat-derived *Pneumocystis* (Wakefield, 1998).

Class III analysis of the rat-derived and human-derived *Pneumocystis* karyotypes showed them to be distinct (Stringer *et al.,* 1993a). Five mammalian *Pneumocystis* host species, rat,

mouse, ferret, rabbit and human, were distinguished by comparing their antigenic proteins, isoenzymes, chromosomal karyotypes and DNA sequences (Hunter *et al.,* 1996). Nearly 30 *Pneumocystis* organism types from different hosts have been compared at one or more loci, for example, porcine and equine *Pneumocystis* DNA were compared with other types of *Pneumocystis* DNA at the mitochondrial large subunit rRNA, mt LSU rRNA locus and were shown to be distinct (Wakefield *et al.,* 1997, Peters *et al.,* 1994). Forty three parasite isolates from rat, mouse, rabbit and human were analysed at the thymidylate synthase locus (Mazars *et al.,* 1995). The amino acid sequences were 83% identical and the diversity of the thymidylate synthase sequence related to the host species.

The *Pneumocystis* mt LSU rRNA and dihydropterate synthase, DHPS, loci from 18 different primate species confirmed host specificity and suggested coevolution (Demanche *et al.,* 2001). A specific mt LSU rRNA and DHPS sequence could be attributed to each primate species or subspecies. No nonhuman primate harboured human-derived *Pneumocystis.* Genetic divergence in primate-derived *Pneumocystis* organisms varied in terms of the phylogenetic divergence existing among the corresponding host species, suggesting some coevolution between host and pathogen.

A similar study sequenced both the mt LSU rRNA and the mt SSU rRNA loci from *Pneumocystis* derived from 24 different mammalian species (Guillot *et al.,* 2001). The mammals divided into six major groups; primates (12 species), rodents (5 species), carnivores (3 species), bats (1 species), lagomorphs (1 species), marsupials (1 species) and ungulates (1 species). No animal harboured human-derived *Pneumocystis*. Comparison of combined mt LSU rRNA and mt SSU rRNA sequences confirmed coevolution of *Pneumocystis* and corresponding mammalian hosts.

1.1.5. Links to Nomenclature

In recognition of its genetic divergence and specificity for a given host species *Pneumocystis* cannot be considered as a single species but instead is acknowledged as a genus containing multiple species. In 1994, after a *Pneumocystis* workshop, a tripartite nomenclature system that described *Pneumocystis* from each host as a special form was proposed and endorsed (Pneumocystis Workshop, 1994) whereby all forms of *Pneumocystis* were called *P. carinii* f. sp. (*formulae specialis*) followed by the host derivation, for example *hominis* or *muris*. Yet, the system is cumbersome and understates the magnitude of the genetic differences between many *Pneumocystis* species. Therefore, in 2001 at a subsequent *Pneumocystis* workshop, a new nomenclature system was suggested, (Stringer *et al.*, 2002, Stringer *et al.*, 2001a). Firstly, the organism that causes human-derived *Pneumocystis* pneumonia is now named *Pneumocystis jiroveci* (Frenkel, 1999). This does not preclude the use of the acronym PCP, previously the acronym for *P. carinii* pneumonia, because this is now read as '*Pneumocystis* pneumonia'. Secondly, *P. carinii* now refers exclusively to the organism formerly known as *P. carinii* f. sp. *carinii*, one of the two *Pneumocystis* special forms found only in rats.

1.1.6. Life cycle of *Pneumocystis*

Since a culture system for *Pneumocystis* has yet to be obtained, attempts to deduce the life cycle of *Pneumocystis* are limited to observations using techniques such as electron microscopy (Stringer *et al.*, 1996b) and epidemiological studies. Within the lung, two forms are visible; the trophic form which is unicellular, and the cyst, which consists of approximately eight spores (Cushion *et al.*, 1988). It is possible that the trophic form undergoes asexual replication by binary fission, while the cyst form or ascus might be the product of sexual reproduction (Figure 1.1).

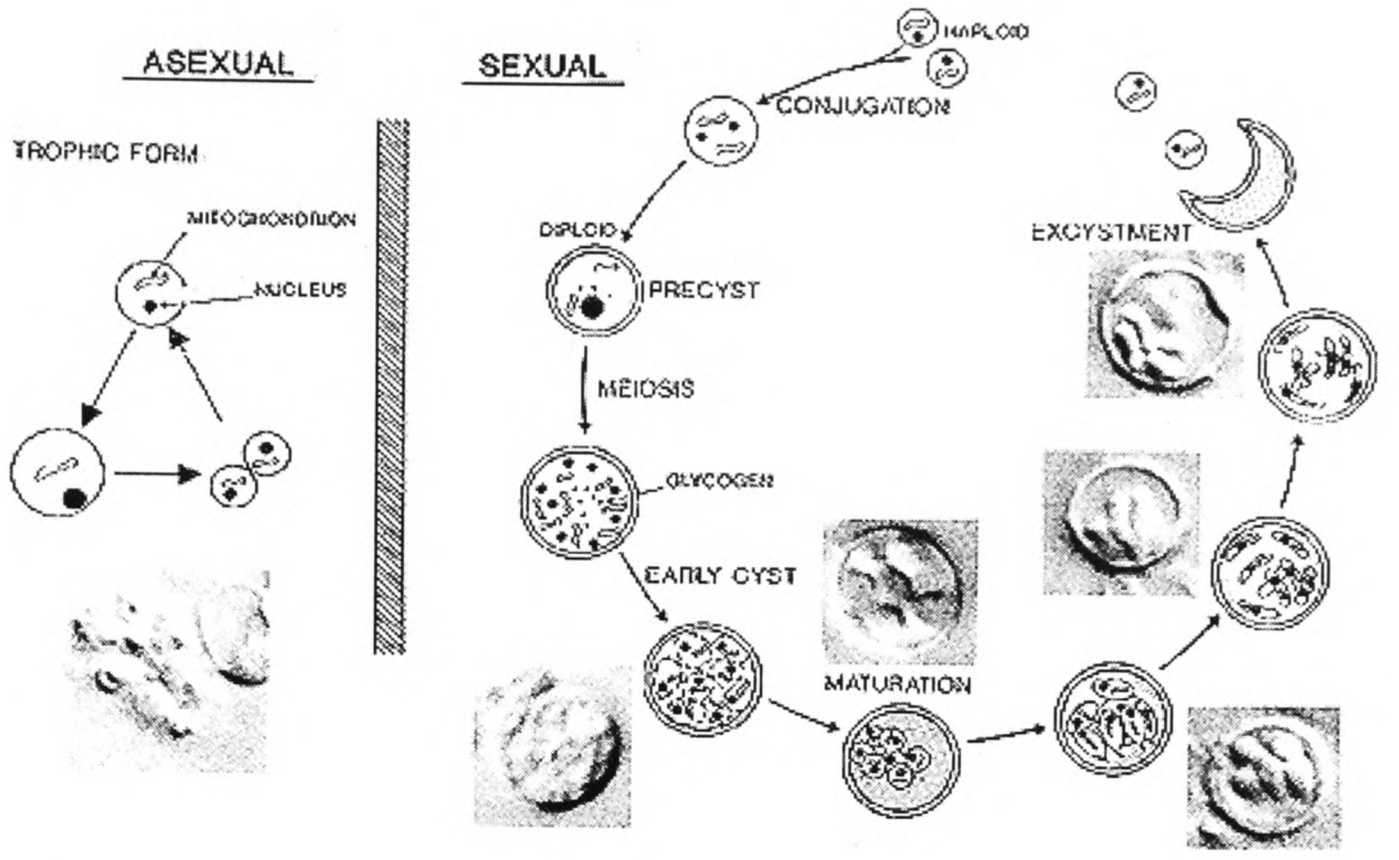


Figure 1.1. Proposed life cycle of *Pneumocystis*

Taken from the *Pneumocystis* Genome Project and resource site, http://biology.uky.edu/Pc/

A sexual stage has been proposed as part of the life cycle of *Pneumocystis*. A *P. carinii* cDNA library revealed the transcription of meiosis-specific genes and other genes related to sexual reproduction (Smulian *et al.,* 2001b). In the same study, the *ste*3 pheromone receptor gene of *P. carinii* was characterised and shown to be surrounded by a cluster of signal transduction genes, which have significant similarity to the mating locus recently described in *Cryptococcus neoformans* (Karos *et al.,* 2000). Furthermore, it has been postulated that the alveolar epithelium is necessary as a substratum for organism conjugation but this can only be implied from photographs (Limper *et al.,* 1997). *Pneumocystis* may have an airborne spore stage in its life cycle, as *Pneumocystis* DNA can be detected from air samples (Wakefield, 1996).

1.1.7. Primary Infection of *Pneumocystis*

There is well-documented serological evidence that up to 94% of normal immunocompetent children have detectable antibody to *P. jiroveci* by 30 months to 4 years of age (Peglow *et al.*, 1990, Pifer *et al.*, 1978) which indicates that *P. jiroveci* is one of the most common infections in humans and that exposure occurs early in life. A prospective cohort study was designed to determine whether primary infection of *P. jiroveci* is associated with clinical illness in the competent host (Vargas *et al.*, 2001). The study indicated that *P. jiroveci* DNA can be frequently detected in healthy infants, and it raises the hypothesis that children may be an infectious reservoir of *P. jiroveci* in the community. More investigation is needed to determine whether *P. jiroveci* directly causes overt respiratory disease in infants. A similar study provided histological evidence that primary *P. jiroveci* infection is associated with sudden infant death syndrome, SIDS (Vargas *et al.*, 1999).

1.1.8. *Pneumocystis jiroveci* causes *Pneumocystis* Pneumonia

P. jiroveci causes pneumonia in the immunosuppressed. It is a common AIDS-defining illness (Katz *et al.*, 1994) and also affects those with other causes of immunosuppression (Miller, 1999b, Arend *et al.*, 1995). Effective treatments are available for PCP. Trimethoprim-sulfamethoxazole (TMP-SMZ) is the drug of choice but many cannot tolerate it because of the substantial toxicity associated with this drug, with adverse events such as gastrointestinal disorders, rash, fever and liver function abnormalities, reported in as many as 65% of cases (Hughes *et al.*, 1995b). Thus many patients are treated with less effective regimens such as atovaquone or pentamidine. HIV protease inhibitors were reported to be active against *P. carinii* in cell culture, so are possibly effective against *P. jiroveci* in HIV patients (Atzori *et al.*, 2000). However, in another study, protease inhibitors at clinical concentrations did not materially reduce the organism count in the treatment of *Pneumocystis* infection in immunosuppressed mice (Walzer *et al.*, 2001). Yet, improved

survival has been shown in HIV-infected patients with severe PCP who are taking highly active anti-retroviral therapy, HAART (Morris *et al.,* 2003). Prophylatic drugs are used for patients who are aware of their immunocompromised status (Schneider *et al.,* 1992).

Diagnosis of PCP is dependent on identifying *P. jiroveci* organisms by histochemical staining, for example with Grocott's methenamine silver, of a respiratory sample, usually a bronchoalveolar lavage, BAL, sample. Previous studies analysing BAL samples suggested there was an uneven distribution of *P. jiroveci* organisms in an infected lung, with an increased organism load in the upper lobes compared to the middle and lower lobes (Grebski *et al.,* 1994). On the chest radiograph, PCP may be more concentrated in the upper lobes (Jules-Elysee *et al.,* 1990). This phenomenon was thought originally to be associated with use of aerosolised pentamidine for treatment, but further studies have shown this to be incorrect (Baughman *et al.,* 1993).

1.2. The Study of *Pneumocystis*

A reliable and reproducible culture system has yet to be developed for *Pneumocystis*. Therefore, key to all research carried out on *Pneumocystis* is that investigations are limited to the use of animal models and molecular biological techniques.

1.2.1. Culture System

Many different attempts have been made to culture *Pneumocystis,* as reviewed by Sloand (Sloand *et al.,* 1993). None have been successful. The stumbling blocks are,

- growth requirements for *Pneumocystis* are undefined
- contamination of cultured organisms by feeder cells

- secondary infection in culture by adventitious infectious agents introduced by *Pneumocystis* inoculum
- technical difficulties in accurate quantification of all forms of *Pneumocystis*

Culturing would allow many things such as, establishment of banks of *Pneumocystis* isolates, exploration of cell biology and life cycle stages, detection of drug resistance in clinical isolates and the ability to clone individual *Pneumocystis* organisms for genetic manipulation, such as the switching of multicopy telomeric genes. A continuous axenic culture has been developed using a culture vessel that allows for frequent medium exchange without disturbance of the organisms (Merali *et al.,* 1999). However, many attempts have been made to reproduce this culture system without success. Short-term culture of up to ten hours has been achieved with A549 lung cells (Limper *et al.,* 1997) and this has aided studies of the cell cycle of *Pneumocystis*.

1.2.2. Animal Models

Pneumocystis can be grown directly within many different immunocompromised mammals, for example, rats, mice, ferrets and rabbits. *Pneumocystis* is host specific yet parallels between rat, mouse and human-derived *Pneumocystis* infection can be drawn so that the models can be compared. Since it was developed in the 1960s, the most important model of infection has been the corticosteroid treated rat (Frenkel *et al.,* 1966), (section 2.6). The rat is immunosuppressed by glucocorticoid steroid treatment. It then naturally acquires rat-derived *Pneumocystis* organisms from several possible sources, either from the environment, other rats in the colony or has latent organisms in the lung that were acquired from birth onwards. After 8-10 weeks, disease should develop. Another rat model utilises a simple, non-invasive intratracheal inoculation of cryopreserved rat-derived *Pneumocystis* organisms and this results in development of severe *Pneumocystis* infection within 5 weeks (Boylan *et al.,*

1992). This method has the advantage that the inoculum source and the concentration of organisms are known.

Intratracheal inoculation of an immunosuppressed rat with as few as 10 *P. carinii* organisms routinely produced an infection resulting in millions of *P. carinii* organisms (Cushion *et al.*, 1999). The complexity of the population resulting from the low dose inoculum was much simpler than that of the population used to prepare the inoculum, as assessed by sequencing and analysing the UCS/MSG junction as described later (Keely *et al.*, 1999). These findings suggested that low dose inoculation could produce a clonally derived *Pneumocystis* population. Unfortunately, further studies were compromised by the common infestation of commercial rat colonies with latent *P. carinii*.

Pathogen-free rats from a commercial colony produced a clonally derived rat *Pneumocystis* population (Keely *et al.*, 2003). Alternatively, inoculation of immunosuppressed nude rats could be another source of animals without latent *P. carinii* organisms. Both pathogen free animals and nude rats could be inoculated with as few as 10 *P. carinii* organisms to give rise to simple populations of rat-derived *Pneumocystis* containing very few types of *Pneumocystis* organisms, ideally only one, with very little risk of latent infection. These potentially clonal organisms would be very useful for studies using molecular techniques.

1.2.3. Molecular Techniques to Study *Pneumocystis*

PCR amplification of *Pneumocystis* DNA allows small quantities of *Pneumocystis* to be detected even in impure samples. As few as one or two organisms may be detected by PCR amplification followed by an oligonucleotide hybridisation (Peters *et al.*, 1992a). PCR amplification can be used as a diagnostic tool and semi-quantitative PCR amplification can be used in order to estimate the organism load of a PCP infection. Nested PCR amplification

is now more commonly used (Tsolaki *et al.,* 1996). This involves two PCR amplifications with two different sets of primers and eliminates the need for oligonucleotide hybridisation. PCR amplification can also be used for epidemiological studies and molecular genetic studies of *P. jiroveci* (Wakefield, 2002). Other techniques have been used, for example single strand conformation polymorphism, SSCP (Hauser *et al.,* 2001), serotyping (Leigh *et al.,* 1993) and fluorescent *in situ* hybridisation (Jensen *et al.,* 2001). However, interpretation of these techniques is very subjective.

Recently, real time quantitative PCR amplification using a light-cycler and RT-PCR amplification has been used for analysis of *P. jiroveci* (Larsen *et al.,* 2002, Palladino *et al.,* 2001). Quantitative PCR amplification has the advantage of greater precision and more objectivity than counting stained *Pneumocystis* organisms on microscope slides, (Kaiser *et al.,* 2001). An RT-PCR amplification assay was developed to test the viability of *P. jiroveci,* using primers specific to the constitutively expressed heat shock protein 70 multigene family, HSP70 (Maher *et al.,* 2000). The RT-PCR assay was capable of detecting as few as 10^2 viable trophozoites as determined by ethidium bromide staining. In contrast, no HSP70 mRNA was detected in the RT-PCR assay from 10^6 nonviable trophozoites killed by high heat, desiccation or UV exposure.

1.3. Epidemiology of *Pneumocystis*

1.3.1. Genotyping of *Pneumocystis jiroveci*

The acquisition and transmission of *Pneumocystis,* especially *P. jiroveci* is not well understood. In order to conduct epidemiological studies of *P. jiroveci* infection without a culture system, it is necessary to be able to distinguish between isolates of *P. jiroveci.* Studies comparing health care worker serum titres of antibodies against *P. jiroveci* have been

used (Leigh *et al.,* 1993) yet, molecular techniques utilising DNA sequence polymorphisms have been more successful for the typing of *P. jiroveci* isolates (Morris *et al.,* 2002).

Variation in *P. jiroveci* DNA sequence has been described at several loci; mitochondrial large subunit rRNA, mt LSU rRNA (Wakefield *et al.,* 1994), dihydropteroate synthase, DHPS (Lane *et al.,* 1997), thymidylate synthase, TS (Mazars *et al.,* 1995), nuclear 5S rRNA (Latouche *et al.,* 1997c), the mitochondrial small subunit rRNA locus, mt SSU rRNA (Hunter and Wakefield, 1996), the arom locus, (Banerji *et al.,* 1993), the superoxide dismutase, SOD (Denis *et al.,* 2000), and the internal transcribed spacer region of the nuclear ribosomal RNA operon, ITS (Lu *et al.,* 1994). The ITS regions contain the most variable nucleotide sequences therefore it is the most discriminatory locus to use to distinguish between clinical samples. It is assumed that the ITS locus is present in one copy in *Pneumocystis*. This is somewhat unusual compared with other fungi but this has been demonstrated to be the case in *P. carinii* (Giuntoli *et al.,* 1994) and *P. jiroveci* (Nahimana *et al.,* 2000, Tang *et al.,* 1998). Twenty-one ITS1 genotypes have been reported along with twenty nine ITS2 genotypes (Nimri *et al.,* 2002).

The reliability and reproducibility of ITS genotyping have been studied (Tsolaki *et al.,* 1999). BAL samples were taken from eleven patients at the beginning and end of an episode of PCP and the same genotypes were seen in both samples from each patient. Therefore, the ITS genotypes do not change during the course of an episode of PCP. Six samples were taken pre 1981, before the AIDS pandemic, from non-HIV patients with PCP (Tsolaki *et al.,* 1998). The same ITS genotypes were identified as those seen in HIV positive patients with PCP. This suggests that the same *P. jiroveci* types have persisted over time. The same ITS genotypes are observed in widely different geographical locations. Samples from Britain, USA, Brazil and Zimbabwe were genotyped and again they were no differences between the

genotypes from the different countries (Wakefield *et al.*, 1994). However, another study has suggested that the ITS locus was evolving faster than other loci (Nimri *et al.*, 2002). Consequently, the sequence information derived from this locus is useful for purposes such as the examination of potential person-to-person transmission and recurrent infections but not for other genotyping applications that rely on more stable genetic loci. It has been suggested that some ITS genotypes are associated with severity of PCP, for example, B_2a_1 is often found in cases of mild PCP (Miller and Wakefield, 1999a). By contrast, another study did not find an association between ITS genotypes and clinical severity of PCP (Helweg-Larsen *et al.*, 2001b).

The *P. jiroveci* dihydropteroate synthase (DHPS) locus is a target for sulphur drugs. Mutations in the gene encoding dihydropteroate synthase resulting from selective pressure from sulphur exposure have been observed in numerous bacterial pathogens and in *Plasmodium falciparum* (Fermer *et al.*, 1995, Brooks *et al.*, 1994) The *P. jiroveci* DHPS locus can contain two common mutations, which could result in a resistance to sulphur drugs for example, sulfamethoxazole and dapsone (Lane *et al.*, 1997). Some studies have shown that mutations in the DHPS locus are associated with a worse prognosis in PCP (Helweg-Larsen *et al.*, 1999, Kazanjian *et al.*, 1998) and are also associated with the duration of sulphur prophylaxis exposure in AIDS patients (Kazanjian *et al.*, 2000). Another study contradicts these associations by demonstrating that patients with and without mutations in the DHPS locus have similar outcomes (Navin *et al.*, 2001). Genetic analysis of multiple loci suggests that mutations in the *P. jiroveci* DHPS locus arose independently in multiple strains (Ma *et al.*, 2001). Other loci that might be subject to drug selection have been analysed. The dihydrofolate reductase, DHFR, gene is the target for trimethoprim; however, no mutations have been identified at this locus (Ma *et al.*, 1999). Cytochrome b is inhibited by atovaquone

and mutations have been identified in the cytochrome b gene. However, atovaquone resistance cannot be explained by these mutations alone (Kaneshiro, 2001).

When used to describe differences between isolates of an organism, multilocus typing using several of the characterised loci together is a more powerful tool than genotyping at a single locus (Wakefield *et al.,* 2003, Feavers *et al.,* 1999, Maiden *et al.,* 1998). There is a greater sensitivity for detecting differences when several loci are examined concurrently.

Mixed infections of *P. jiroveci* are observed (Tanabe *et al.,* 1988). This is due to co-infection of two or more different *P. jiroveci* genotypes. One study observed the presence of more than one genotype of *P. jiroveci* in 60% of the respiratory samples examined (Nimri *et al.,* 2002). Mixed infections may complicate epidemiological studies (Helweg-Larsen *et al.,* 1998). Respiratory samples are assumed to represent the *P. jiroveci* genotypes that are uniformly distributed throughout the human lung. Yet, the distribution of *P. jiroveci* genotypes in the human lung needs to be characterised to validate this assumption

1.3.2. Latency versus Reinfection

Exposure to *P. jiroveci* is frequent and reinfection of organisms rather than the proliferation of latent organisms is thought to be the more likely model of infection leading to the development of disease (Latouche *et al.,* 1997b, Tsolaki *et al.,* 1996, Keely *et al.,* 1995). Exposure to *P. jiroveci* occurs in infants before the age of three and antibodies are produced against *P. jiroveci* (Peglow *et al.,* 1990, Pifer *et al.,* 1978). The latency model of infection postulates that after natural exposure a few organisms remain latent within the lung, as is the case for tuberculosis. Then, if the patient became immunocompromised the organisms proliferate and PCP would result. Fifteen post-mortem lungs from immunocompetent

individuals were analysed for the presence of *P. jiroveci* by DNA amplification (Peters *et al.*, 1992b), which can detect one to two organisms; none were found.

Dexamethasone immunocompromised mice were used in another study to investigate the latency model (Vargas *et al.*, 1995). After eight to twelve weeks, the mice contracted *Pneumocystis* infection. Then the immunosuppression was withdrawn and the animals were moved to a high-efficiency particulate air-filtered isolator to prevent further exposure to environmental *Pneumocystis*. At intervals thereafter, sample groups were transferred to a second isolator and reimmunosuppressed with dexamethasone to provoke *Pneumocystis* infection if *Pneumocystis* organisms were still present. Infection returned if there was less than six months between the first infection episode and the second round of immunosuppression but if this time period was more than a year, infection did not return. This implies that the mice are capable of clearing the *Pneumocystis* organisms and persistence of latent organisms is limited.

BAL samples were taken from ten patients who suffered from two episodes of PCP at least three months apart (Keely *et al.*, 1995). The samples were typed at the mt LSU rRNA locus and five patients had a different genotype at the second episode of PCP. Reinfection rather than reactivation seems to have occurred. Allele specific PCR, (ASP) amplification of the mt LSU rRNA locus was then used on the samples from the five patients with a different genotype at the second episode of PCP, (Keely *et al.*, 1996). In four of the five patients, ASP did not detect the second-episode genotype in the first-episode sample. In another study, samples from twelve patients were typed at the ITS locus, (Tsolaki *et al.*, 1996). Mixed infection and reinfection were observed. Also, genotype frequency distribution patterns vary by place of diagnosis rather than place of birth (Beard *et al.*, 2000) suggesting recent acquisition of infection.

1.3.3. Transmission of *P. jiroveci*

In the rat model, airborne transmission of *P. carinii* has been shown (Hughes *et al.,* 1983) and in the mouse (Walzer, 1984). The existence of airborne *P. jiroveci* organisms has been supported by the detection of *P. jiroveci* DNA in samples of airborne spores from English rural environments (Wakefield, 1996), from animal facilities housing immunosuppressed rats with *Pneumocystis* infection, and in hospital rooms used for treating patients with PCP (Bartlett *et al.,* 1997, Latouche *et al.,* 1997a, Olsson *et al.,* 1996, Bartlett *et al.,* 1994). This suggests that there might be airborne transmission of *P. jiroveci.* Epidemics of *P. jiroveci* were recorded after the Second World War and clusters of infection are common in immunosuppressed children with malignancy and transplant patients (Hughes, 1998a). Geographic clustering of PCP in patients with HIV infection has been observed in Cincinnati (Dohn *et al.,* 2000), and San Francisco (Morris *et al.,* 2000).

Direct transmission of the organism between immunocompromised adults has yet to be conclusively shown. Genotyping at the ITS locus was performed on isolates of *P. jiroveci* from three observed clusters of PCP on hospital wards (Helweg-Larsen *et al.,* 1998). A common ITS genotype was observed in 2 of the patients who shared a hospital room and in 2 of the patients who had prolonged close contact on the ward. In contrast, different ITS genotypes were detected in samples from an HIV-infected homosexual couple who shared the same household. These data suggest that some form of person-to-person transmission may occur but does not account for the majority of the cases within the clusters.

DHPS mutations occur at a high frequency among persons never prescribed sulphur prophylaxis, including those newly diagnosed with HIV infection (Huang *et al.,* 2000). This provides indirect evidence that these mutations are transmitted from person-to-person either directly or through a common source. The mutations would have developed in an individual

who was taking some form of sulphur prophylaxis. The *P. jiroveci* organism with the mutations in the DHPS locus would then have to be transmitted to be present in an immunocompromised patient who was not taking sulphur prophylaxis (Beard *et al.*, 2000).

Transmission may occur between immunocompromised patients to healthy immunocompetent people. *P. jiroveci* DNA was detected in a respiratory sample from a child with PCP and in respiratory samples from his mother and two contact health care workers (HCW) (Vargas *et al.*, 2000). In contrast, no *P. jiroveci* DNA was detected in respiratory samples from 30 hospital staff members, none of whom entered the patient's room. In another study, serum titres of *P. jiroveci* antibodies from immunocompetent HCW caring for patients with AIDS were raised when compared with the serum titres of *P. jiroveci* antibodies from HCW exclusively treating elderly patients (Leigh *et al.*, 1993). These results contrast with those from another study, which compared hospital staff treating patients with PCP with hospital staff not treating patients with PCP (Lundgren *et al.*, 1997). There was no difference in their serological responses to *P. jiroveci* and no detectable *P. jiroveci* DNA was detected in oropharyngeal washings from any of the staff, implying that hospital staff do not act as carriers of *P. jiroveci*. However, only a single round of PCR amplification was performed rather than a more sensitive nested PCR amplification, possibly explaining the lack of detectable *P. jiroveci* DNA.

1.3.4. Colonisation and Carriage of *P. jiroveci*

Asymptomatic carriage of *P. jiroveci* has been postulated as another stage in the life cycle of the organism. Low numbers of *P. jiroveci* organisms have been detected in immunocompromised patients without PCP, both HIV infected and patients that are mildly immunosuppressed, (Helweg-Larsen *et al.*, 2002, Hauser *et al.*, 2000a, Nevez *et al.*, 1999b,

Sing *et al.*, 1999, Calderon *et al.*, 1996). It is not known if these organisms are able to cause disease.

In an animal study on the transmission of *Pneumocystis,* using the mouse model of infection, it was shown that immunocompetent BALB/c mice that were carrying subclinical levels of *Pneumocystis* were able to transmit the infection by the airborne route to highly susceptible, uninfected SCID mice (Dumoulin *et al.,* 2000). Whether this scenario is true for *P. jiroveci* remains to be seen.

Asymptomatic carriage was investigated in multiple BAL fluid samples taken from 16 HIV infected patients (Wakefield *et al.,* 2003). Multilocus genotyping was done at the mt LSU rRNA, mt SSU rRNA, DHPS and SOD loci. Seven patients had asymptomatic carriage of *P. jiroveci* over periods of nine months or more, following an episode of PCP. During carriage, in all seven patients, a change in genotype from that associated with the original episode of PCP was observed. Two patients had asymptomatic carriage for up to nine months or more prior to an episode of PCP with no change in genotype and five patients were asymptomatic carriers for up to two years without developing PCP. Asymptomatic carriage clearly plays some role in the life cycle of *P. jiroveci.*

P. jiroveci ITS genotypes were determined in 11 PCP patients and 6 HIV-negative patients colonised with *P. jiroveci.* A BAL sample was taken from the 6 colonised patients in order to investigate pulmonary symptoms or fever (Nevez *et al.,* 2003). In total, 14 different genotypes were identified, of which 1 was found only in colonised patients, 10 were found only in patients with PCP, and 3 were found in both patient populations. New ITS1 and ITS2 types were identified in this study as well as new combinations of ITS1 and ITS2 types. The results show that similar genotypes can be responsible for PCP as well as pulmonary

colonisation. There were mixed infections and a high diversity of genotypes in both groups of patients. Asymptomatic carriage in HIV infected and immunocompromised patients may play a role in transmission of *P. jiroveci* and may even be a reservoir for future infections.

1.4. Multigene Families in *Pneumocystis*

1.4.1. Multigene Families within the Genome of *Pneumocystis*

From karyotype analysis, the *P. jiroveci* genome consists of a haploid form of 16 chromosomes, ranging from 200 to 1000 Mb in size. Respiratory samples containing only a few *P. jiroveci* organisms are the main source of *P. jiroveci* DNA for analysis. The small number of organisms limits the amount of genomic analysis that can be done. Animal models can be utilised to grow large numbers of organisms, therefore, more is known about the genome structure of *P. carinii.* There are two special forms of rat-derived *Pneumocystis* that can be distinguished by a simple PCR amplification (Palmer *et al.,* 1999). Three other special forms have been identified in Danish wild rats (Palmer *et al.,* 2000b). The karyotype analysis shows that rat-derived *Pneumocystis* consists of 16 chromosomes, ranging from 300 to 700 kb, giving a total genome size of 7 to 8 Mb (Cushion *et al.,* 1993a). The *P.carinii* genome is being sequenced, http://www.gene.genetics.uga.edu and http://www.biology.uky.edu/Pc/.

Three multigene families encoding cell surface proteins have been identified in the subtelomeres of *P. carinii* chromosomes; major surface glycoprotein, the *MSG* gene, major surface related glycoprotein, the *MSR* gene and protease1, the *PRT1* gene. The telomeric location for genes encoding variable surface antigens is very advantageous for the organism. A silent gene can be moved to an expression site via a reciprocal chromosomal exchange event without translocating major segments of chromosomes. Batteries of linked genes can

be moved and new genes can be generated via unequal crossing over between existing genes without disrupting the chromosomal structure.

Many subtelomeric multigene families have been identified in other parasites, for example, the variant surface glycoprotein (*VSG*) genes of *Trypanosoma brucei* (Rudenko, 1999) and the variant surface antigen, *var,* genes of *Plasmodium falciparum* (Newbold, 1999). This implies that either multigene families arose once early in evolutionary history, or alternatively, telomeric multigene families arose more than once in different organisms (Stringer *et al.,* 2001b). Many of these multigene families are necessary for host adaptation, in many cases switching between these variable proteins results in antigenic variation. *T. brucei* has 20 similar telomeric-expression sites for *VSG* genes (Rudenko, 1999). Expression appears to be controlled at the level of transcription initiation, resulting in only one site being active at any time. Switching between expression sites to transcribe different isoforms of VSG occurs at a low rate (Chaves *et al.,* 1999). The switching of VSG generates antigenic variation. With the *P. falciparum var* genes all promoters are capable of activation and there is no necessity for duplication or translocation of *var* genes to a single expression site (Scherf *et al.,* 1998). The mechanism underlying switching in *P. falciparum* remains elusive (Kyes *et al.,* 2001).

It is possible that the multigene families of *Pneumocystis* promote antigenic variation to allow the organism to evade the immune system of the host. Both MSG and PRT1 are immunodominant proteins. MSG is recognised by serum antibodies and T cells from the host (Theus *et al.,* 1997, Gigliotti *et al.,* 1996) while proteins with the same size range as PRT1, 42 to 52 kDa, have frequently been documented as being immunodominant, showing strong reactivity to hyperimmune and recovery serum, in both the rat and human infection (Hong *et al.,* 1995, Walzer *et al.,* 1987). Expressed *MSG* genes vary between rat-derived populations

of *Pneumocystis* (Sunkin *et al.,* 1996) and many variant cDNA copies of *PRT1* have been identified (Lugli *et al.,* 1997). Therefore, there is evidence supporting antigenic variation in *Pneumocystis*. This might allow the microbe to colonise immunocompetent hosts often and long enough to allow replication and transmission to new hosts.

1.4.2. Major Surface Glycoprotein

There are approximately 100 copies of the *MSG* gene, found in clusters at the chromosome ends of *Pneumocystis*. The encoded proteins have a molecular weight of 95 to 140 kDa. Ten percent of the molecular weight is carbohydrate, mainly mannose. Most of the research on MSG has been carried out using the rat model. MSG is expressed on the cell surface presumably via a GPI anchor as the carboxyl terminus of the *MSG* gene encodes a functional GPI signal sequence (Guadiz *et al.,* 1998). MSG has been postulated to be involved in the binding of the parasite to the epithelium and can bind to a variety of host structures (Benfield *et al.,* 1998). MSG is important in host T-cell recognition of *Pneumocystis* and the immune response to *Pneumocystis* through cytokine production and the recruitment of inflammatory cells (Theus *et al.,* 1993). MSG may also undergo antigenic variation.

Initially the study of MSG was via biochemical and immunological approaches (Gigliotti, 1992). When *MSG* genes became accessible more was learnt about the composition, organisation and expression of MSG. *MSG* cDNA from ferret-derived *Pneumocystis* was the first to be cloned (Haidaris *et al.,* 1992). Immunoscreening identified *MSG*-specific cDNA from *P. carinii* (Wada *et al.,* 1993). There was a high level of expression of many different *P. carinii MSG* genes with 70% amino acid identity between them, generating potential variation at the cell surface. At least 26 different *MSG* genes are expressed in a *P. carinii* population (Sunkin *et al.,* 1997). Also, when over 10 kb of genomic chromosomal DNA was cloned, it contained two copies of *MSG* genes and a third unidentified open reading frame,

ORF-3 (Wada *et al.*, 1994). This study showed that the *MSG* gene is clustered in tandem repeats.

An upstream conserved sequence, UCS, is required for the transcription of an *MSG* gene. It was identified by analysis of the 5' upstream regions of *MSG* cDNAs by 5' Rapid Amplification of cDNA Ends, 5' RACE (Wada *et al.*, 1995). 5' RACE is a procedure for the amplification of nucleic acid sequences from a messenger RNA template between a defined internal site and unknown sequences at the 5' end of the mRNA. The UCS was detected at the 5' ends of *MSG* mRNAs (Wada *et al.*, 1995). The UCS element maps to one band in an electrophoretic karyotype suggesting the UCS locus to be unique. This hypothesis was substantiated by quantitative hybridisation which found that the hybridisation signal from a UCS probe was comparable to the hybridisation signal from an equivalent probe specific to known single copy genes, p55 and α-tubulin (Sunkin and Stringer, 1996).

There were many different cDNA copies of the *MSG* gene attached to the UCS in each *P. carinii* population, which suggests that individual organisms may transcribe different variants of the *MSG* gene (Wada *et al.*, 1995). This prediction was supported by indirect immunofluorescence observations. Only some *P. carinii* organisms were brightly stained with a mouse monoclonal antibody specific for a limited number of MSG isoforms (Sunkin and Stringer, 1996). In the same study, comparison of three different populations of *P. carinii* showed that each contained a different set of *MSG* genes linked to the UCS, suggesting that UCS/MSG junctions are switched by recombination during population growth.

Both the UCS and most *MSG* gene copies were shown to be subtelomeric by BAL-31 digests (Sunkin and Stringer, 1996, Underwood *et al.*, 1996). BAL-31 exonuclease degrades both

the 3' and 5' termini of duplex DNA without generating internal scissions, so is a good technique for determining if a gene is subtelomeric or not. The subtelomeric location of the *MSG* gene copies implies that they might recombine by reciprocal exchange. Not all *MSG* gene copies were subtelomeric. Although high levels of recombination at the telomeres could lead to homogenisation of *MSG* gene sequences, divergence may have occurred once copies of *MSG* genes were moved to a more internal chromosomal location, and may have lead to the observed heterogeneity in the *MSG* gene family (Underwood *et al.,* 1996).

Attachment to the UCS is necessary for transcription of an *MSG* gene (Sunkin and Stringer, 1997). The *MSG* gene is not spliced out of a polycistronic message to the UCS. *MSG* genes often contain length polymorphisms of up to 30 bp in the region adjacent to the UCS and the length polymorphisms always differ by 3 bp increments, which conserves the reading frame encoding MSG. A PCR amplification technique was developed for displaying variation at the UCS/MSG junction and took advantage of this length variation. The different sized products were separated on sequencing gels and a different profile of PCR products was generated from the genomic DNA from each rat-derived *Pneumocystis* population investigated. The PCR amplification products from each population generally comprised of 12 differently sized UCS/MSG bands, which consisted of at least 26 different *MSG* genes. The UCS/MSG junctions in the mRNA matched those in the genome, as would be expected if splicing did not contribute to the attachment of the UCS to the 5' ends of the *MSG* mRNAs.

The PCR assay described above could be used to determine if other multigene families have corresponding genomic and mRNA profiles which would give information about the regulation of these multigene families within a population. Two factors contribute to the complexity of these multigene families, the heterogeneity of the population and the

switching of the genes to expression sites. When investigating multigene families within the context of a population, knowing the relative contribution of each factor facilitates the interpretation of the results.

The UCS locus is transcribed and is presumed to be translated. To explore a possible functional role of the UCS in the trafficking of MSG, the nucleotide sequence encoding the UCS peptide was ligated to the 5' end of an *MSG* gene and incorporated into a recombinant baculovirus (Sunkin *et al.,* 1998). Insect cells, *Spodoptera frugiperda*, infected with the UCS-MSG hybrid gene, expressed a 160 kDa protein, which was N-glycosylated. By contrast, insect cells infected with a baculovirus carrying an *MSG* gene lacking the UCS, expressed a non-glycosylated 130 kDa protein. The data suggests that in *P. carinii* translation begins in the UCS to produce a pre-MSG protein, which is subsequently directed to the endoplasmic reticulum and undergoes proteolytic cleavage of the UCS to be processed to the mature form.

The *MSG* gene from *P. jiroveci* has been cloned and characterised (Garbe *et al.,* 1994). Recombinant MSG proteins were generated which were recognised by sera from healthy humans and from patients. The *P. jiroveci* UCS had been identified and has been shown to be single copy, thus *P. jiroveci* regulates *MSG* gene expression in a manner similar to *P. carinii* (Kutty *et al.,* 2001).

Major Surface Related Glycoprotein, *MSR,* gene, is another multigene subtelomeric family that is related to the *MSG* gene (Huang *et al.,* 1999b). To date, *MSR* has only been identified in *P. carinii*. *MSR* genes lack a 1 kb region present in all *MSG* genes. While most regions of any two MSG isoforms are more than 70% identical, only the carboxy terminus of the MSR protein was more than 70% identical to MSG (Schaffzin *et al.,* 1999). Furthermore, MSR

had several regions that were less than 20% identical to MSG. Such regions of low identity are unknown among MSG isoforms. With CHEF analysis, *MSR* specific probes hybridised to at least 13 rat-derived *Pneumocystis* chromosomes. Also, a Southern blot of digested genomic rat-derived *Pneumocystis* DNA detected 16 bands ranging in size from 4 to 50 kb, showing that a multigene family encodes MSR proteins. Multiple *MSR* genes are expressed in a *P. carinii* population and two different size classes of MSR mRNA, 2.4 and 3.5 kb, have been detected (Schaffzin *et al.,* 1999). Both sizes of *MSR* cDNA lacked an upstream UCS sequence suggesting that expression of *MSR* genes does not require linkage to the UCS locus.

The UCS/MSG locus was examined in the second special form of rat-derived *Pneumocystis* (Schaffzin *et al.,* 2000). The first 380 nucleotides of the second special form of rat-derived *Pneumocystis* were 59% identical to the UCS of *P. carinii*. The DNA encoding the UCS of the rat second special form was not unique in the genome. Instead, only the first 97 nucleotides came from a unique locus and the remainder from over 7 different loci as determined by Southern blots and CHEF analysis. These data suggest that the two different types of *MSG* multigene family may use different mechanisms to generate surface variation. This appears to be a theme throughout subtelomeric multigene families, that each has its own means of regulation of expression.

1.4.3. Protease1

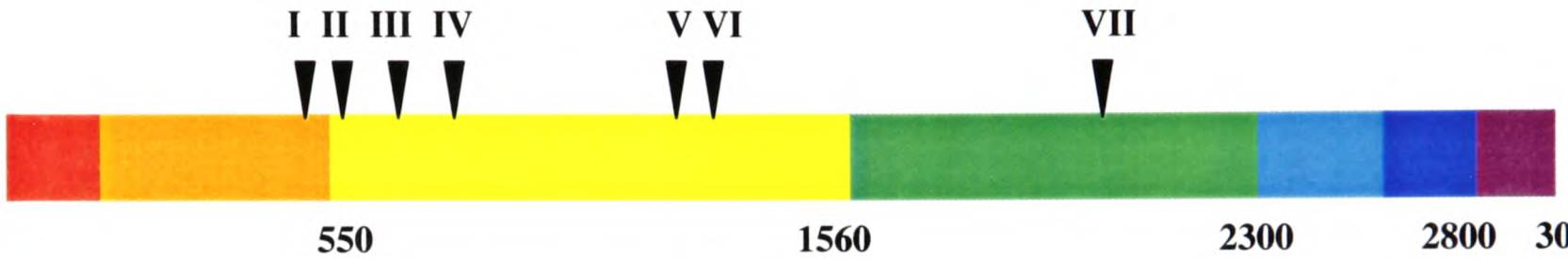
Another multigene family has been identified in the subtelomeric regions of *P. carinii*. It encodes a surface expressed subtilisin-like serine protease, protease1, PRT1. PRT1 has highest homology with the kexin family, which encode single copy intracellular subtilisin-like serine proteases (Fuller *et al.,* 1989). Kexin is a prohormone processing enzyme, which cleaves the precursors of the α-mating-factor and the killer toxin. It is encoded by the *KEX2*

gene of *Saccharomyces cerevisiae*. The human protein furin is a similar subtilisin-like protease and has emerged as one of the major processing enzymes of the constitutive secretory pathway (Steiner, 1998). Other subtilisin-like proteases are the prohormone convertases PC1 and PC3. They are involved in the processing of a wide variety of hormone and neuropeptide precursors (Steiner, 1998).

A copy of the *PRT1* gene was first isolated from a *P. carinii* genomic library and another copy was isolated from a cDNA library (Lugli *et al.,* 1997). The sequences from the two copies were compared and seven introns were identified. The cDNA copy consisted of approximately 790 amino acids and from this seven domains were predicted (Figure 1.2). The first domain is the N-terminal hydrophobic signal sequence, followed by a pro-domain, which contains a potential cleavage site for autoproteolysis. The three essential residues of the active site, aspartic acid, histidine and serine, which were identified from other eukaryotic subtilisin-like serine proteases, are present in the catalytic domain. The P-domain is essential for proteolytic activity while the proline rich region is unique to PRT1 and has not been identified in any other serine proteases. Yet, the neuroendocrine protein, 7B2, has a polyproline helix-like structure and is involved with the interactions of 7B2 with prohormone convertase 2, PC2 (Zhu *et al.,* 1996). There are many potential O-glycosylation sites in the serine-threonine rich region of PRT1 but it has yet to be demonstrated if glycosylation occurs. A GPI-anchor site is present in the C-terminal domain and this has been shown by functional *in vivo* assays to be the means by which the protease is bound to the cell surface (Palmer *et al.,* 2001).

Anti-PRT1 antiserum was used in immunofluorescence studies to demonstrate the specific localisation of PRT1 to the surface of *P. carinii* (Lugli *et al.,* 1999b). PRT1 is present on the cell surface of both the trophozoite and cyst forms of the organism in the rat lung. Also, anti-

Figure 1.2. A schematic diagram of the DNA sequence and proposed protein structure from a *P. carinii PRT1* gene. The nucleotide numbering is based on the genomic *PRT1* copy 3G5, which is part of a cosmid clone. The black arrows, I-VII, indicate the positions of seven introns.



N-terminal hydrophobic signal sequence	Pro-domain Sequence	Catalytic Domain Sequence	Proteolytic Domain Sequence	Proline-rich Region	Ser/Thr rich region	C-terminal Sequence
The sequence is hydrophobic in nature and so is likely to be a signal sequence. Signal sequences are found in all secreted ER and lysosomal proteins as well as many transmembrane proteins.	Possibly autocatalytically cleaved as some copies contain a potent dibasic autocatalytic site, for example, Lys-Arg.	Highly homologous to a subtilisin-like serine protease catalytic domain. Contains the 3 essential active site residues, Asp, His and Ser. **Thought to be the most conserved domain in sequence and size.** Possible that the introns may play a role in maintaining this conservation	Essential for catalytic activity. Shows homology in this region to other subtilisin-like serine proteases. **Appears to be conserved like the catalytic domain**	Present in all *PRT1* genes, but not in other serine proteases. Contains repeated prolines as well as Pro-Glu and Pro-Gln. **Most varied domain in both length and sequence**	Exhibits potential to be modified by O-linked glycosylation	A membrane associated domain, comprised of hydrophobic sequences, containing a GPI anchor site.

PRT1 antiserum was used to show that PRT1 is an important immunodominant protein. Cross reactivity of the anti-PRT1 antiserum was seen with *P. jiroveci.* Part of the catalytic domain from the *P. jiroveci PRT1* gene was cloned (Lugli *et al.,* 1999b).

cDNA and genomic copies of the *PRT1* gene have been characterised (Palmer, 2000a, Lugli, 1999a). The predicted molecular weight for different isoforms of PRT1 ranges from 87 kDa to 102 kDa and a high level of nucleotide heterogeneity has been identified. The proline rich region is the most divergent domain in terms of both length and composition. It has been postulated that PRT1 has a role in the binding of the organism to the epithelium as well as an autocatalytic or possible protein-processing role. Proteases cleave at dibasic residues and there is a dibasic site at the UCS/MSG junction. Some studies have created recombinant PRT1 proteins and attempts were made to cleave MSG with PRT1, but no conclusive result was attained (Rozen and Fuller – personal communication). Karyotypic analysis showed that copies of *PRT1* genes were present on the end of all but one of the 16 *P. carinii* chromosomes. From this data, it can be assumed that there is a minimum of 30 copies of *PRT1* genes per *P. carinii* organism.

Two other groups have characterised copies of the *PRT1* gene. The first group characterised the open reading frame, *ORF-3*, which was identified as part of an *MSG* gene cluster as described in the previous section (Wada *et al.,* 1997). They identified *ORF-3* as a surface subtilisin-like protease, SSP. The second group identified a *P. carinii* multigene family of surface proteases, SPRTs (Russian *et al.,* 1999). The serine endoproteases were localised to the organism surface by immunohistochemical and immunofluorescence studies and to the electron lucent layer of the cyst wall by immunoelectron microscopy. Two cDNA copies of the protease were identified in this study that have a carboxy terminus with a hydrophobic followed by a hydrophilic region, consistent with a transmembrane and intracytoplasmic

region, respectively. This region is thought to be important for localisation of the protease in the endoplasmic reticulum and/or Golgi apparatus, as this is a structural feature found in kexins of several fungi, for example, Kex2 in *Saccharomyces cerevisiae* (Brenner *et al.*, 1994). Most other characterised copies of the *PRT1* gene differ by having a carboxy terminus consistent with a GPI anchor, which is characteristic of most cell surface proteases.

A mouse-derived *Pneumocystis* kexin-like protease, KEX1, has been characterised (Lee *et al.*, 2000). This protein contrasts with PRT1 by being encoded by a single copy gene as determined by Southern blotting. Also, it contains a hydrophobic transmembrane spanning domain and a carboxy-terminal cytoplasmic domain with a conserved tyrosine motif so, KEX1 is unlikely to be expressed at the surface of the organism and as described above is probably bound to the Golgi apparatus. KEX1 may be involved in post-translational processing and maturation of other mouse-derived *Pneumocystis* proteins. If PRT1 has evolved from the kexins it is possible that *P. carinii* also has a functional kexin-like protein, similar to mouse KEX1. Presumably, the protein is sufficiently distinct from PRT1 for the respective gene to have evaded identification.

The *kex1* gene of *P. jiroveci* has also been characterised and was similar to the mouse derived *Pneumocystis kex1* gene (Kutty and Kovacs, 2003); for example, a transmembrane domain was predicted at the carboxy terminus. However, the proline-rich domain present in PRT1 and mouse KEX1 is absent from the *P. jiroveci* Kex1 protein. The deduced amino acid sequence of Kex1 from *P. jiroveci* showed 39% and 37% identity to the two published *PRT1* clones U62910 and U82999 respectively. As previously described, these are the two protease clones characterised as having a transmembrane and intracytoplasmic region (Russian *et al.*, 1999). It also showed 34% identity with KEX1 from mouse-derived *Pneumocystis* and Kex2 of *Saccharomyces cerevisiae*. Southern blotting demonstrated that the *P. jiroveci kex1* gene

is present as a single copy. It is possible that there is a single copy *kex1* gene in *P. carinii* that duplicated after the divergence of mouse-derived *Pneumocystis* and *P. carinii*, to form the *PRT1* multigene family (Kutty and Kovacs, 2003). This is a very different situation from the *MSG* gene family, since the genomes of all *Pneumocystis* species studied to date contain multiple copies of *MSG* genes. *MSG* gene duplication must have occurred in a very early ancestor of most or all of the *Pneumocystis* species.

As part of the *Pneumocystis* Genome Sequencing Project, the Sanger Centre (http://www.sanger.ac.uk) determined the sequence of gene arrays contained within two rat derived *Pneumocystis* genomic cosmid clones, 3G5 and 1B2 (Keely *et al.*, 2001). Both were cloned gene clusters that contained members all of three gene families, *MSG, MSR* and *PRT1* genes, but the order of the genes in the two clones was different. In addition to full size *MSG* and *MSR* genes, both gene clusters contained what appeared to be fragments of *MSG* or *MSR* genes. The library contained additional cosmids, 11H12 and 12D10, that shared a core region of about 15 kb at the telomere-distal end of the sequence with 1B2, and with each other. The shared region may have been inserted into at least two different locations in the genome by gene conversion events. This region consisted of, from the centromere to the telomere, a 5' end of an *MSG* gene, a *PRT1* gene, an *MSR* gene and part of a *PRT1* gene. A single copy chitin biosynthesis gene was present in the non-telomeric region in both 1B2 and 11H12, while 12D10 contained a single copy nuclear protein in the non-telomeric region. Using probes specific to these single copy genes, 11H12 and 1B2, mapped to chromosome 9 and 12D10, mapped to chromosome 4. These linkage data show that the same gene array can be at the ends of two different chromosomes.

There are 10 complete *PRT1* sequences in the Genbank database. Five are from the rat derived *Pneumocystis* genomic cosmid clones that have been sequenced at the Sanger centre,

(http://www.sanger.ac.uk/Projects/P_carinii/). Both 11A11 (the latest cosmid to be sequenced) and 1B2 have two copies of *PRT1* genes and 3G5 has one. Another, *ORF-3*, was described by the Wada group (Wada and Nakamura, 1994) and the two clones U62910 and U82999 were described by the Kovacs group (Russian *et al.*, 1999). Finally, our group has published 2 *PRT1* sequences, pAGA and 73j. All the 10 sequences are genomic apart from the 3 cDNA clones, U62910, U82999 and 73j.

1.5. Aims of Project

The project focuses on two areas of research. The first will examine the epidemiology of *Pneumocystis* infection while the second part will examine the cell surface protease, PRT1. Much is known about the transmission of *P. jiroveci* yet many aspects are not understood, such as the acquisition of organisms and the possibility of asymptomatic carriage of *P. jiroveci* organisms. Three studies were planned to learn more about these aspects. The studies are dependent on clinical samples, so the investigations were done prospectively. The first study will look at the possible direct transmission of *P. jiroveci* between a mother and her infant, both concurrently suffering from PCP. Secondly, a multi-sample study will look at asymptomatic carriage of *P. jiroveci* among HCW who have occupational contact with patients with PCP and will use ITS genotyping methods to determine whether there is transmission of organisms between patients and HCWs. Thirdly, to validate the assumption that respiratory samples are representative of the distribution of *P. jiroveci* genotypes within a human lung, *P. jiroveci* DNA present in an infected whole lung will be genotyped and the distribution of the genotypes will be determined.

The aim of the second part of the project is to investigate a potential virulence factor, the cell surface expressed protein, PRT1. The protease has a variable sequence and is multicopy so

was initially investigated as a potential genotyping marker for epidemiological studies. However, it was latterly thought to have a potential role in the virulence of rat-derived *Pneumocystis*. The aim of the first protease study is to identify and characterise the original rat-derived *Pneumocystis* single copy *PRT1* gene, which encodes the potential progenitor kexin-like protease. A single copy protease gene has been identified in both mouse-derived *Pneumocystis* and *P. jiroveci*. If the progenitor kexin-like protease does exist, it is likely to be constitutively expressed which might complicate data from other *PRT1* studies. The second aim of the protease part of the project is to investigate the diversity and number of genomic and expressed copies of *PRT1* and to determine whether *PRT1* gene expression is regulated. One assay, based on the assay of Sunkin, (Sunkin and Stringer, 1997) as described in section 1.4.2, will be to take a rat-derived *Pneumocystis* population and amplify the length polymorphisms in the proline rich region of the *PRT1* gene and to characterise them using genescan analysis. The amplification profiles from mRNA and genomic DNA can be compared from the same population to determine if regulation of expression has occurred. Alternatively, the catalytic domain of genomic and cDNA *PRT1* copies will be amplified, sequenced and compared in order to investigate the diversity and number of *PRT1* alleles and to determine if *PRT1* gene expression is regulated. The *PRT1* copies will be amplified from potentially clonal rat-derived *Pneumocystis* populations generated from a low-dose inoculation.

2. Materials and Methods

2.1. Units, Abbreviations and Acronyms

Table 2.1. Units, abbreviations and acronyms

AAP	Abridged Anchor Primer
AIDS	Acquired Immune Deficiency Syndrome
ASP	Allele Specific Primer
AUAP	Abridged Universal Amplification Primer
BAL	Bronchoalveolar Lavage
BSA	Bovine Serum Albumen
bp	Base pair
cDNA	Copy DNA
dATP	Deoxyadenosine 5' triphosphate
dCTP	Deoxycytosine 5' triphosphate
ddNTPs	Dideoxynucleotide triphosphates
dGTP	Deoxyguanine 5' triphosphate
DHFR	Dihydrofolate Reductase
DHPS	Dihydropteroate Synthase
dH_2O	Distilled water
dn	Non-synonymous change
DNA	Deoxyribose Nucleic Acid
dNTPs	Deoxynucleotide triphosphates
ds	Synonymous change
DTT	Dithiothreitol
dTTP	Deoxythymine 5' triphosphate
EDTA	Disodium ethylenediaminetetraacetic acid
EST	Expressed Sequence Tag
GPI	Glycosylphosphatidylinositol
HAART	Highly Active Anti-Retroviral Therapy
HCW	Health Care Workers
HIV	Human Immunodeficiency Virus
hr	Hour

HSP70	Heat Shock Protein 70
Inoc	Inoculum
IPTG	Isopropyl-β-D-thiogalacto-pyranoside
IS	Induced Sputum
ITS	Internally Transcribed Spacer region of the nuclear ribosomal RNA operon
kb	Kilobasepair
kDa	Kilodalton
kPa	Kilopascal
LB	Luria Broth
Mb	Megabasepair
min	Minute
MSG	Major Surface Glycoprotein
MSR	Major Surface Related Glycoprotein
mt LSU rRNA	Mitochondrial Large Subunit ribosomal RNA
mt SSU rRNA	Mitochondrial Small Subunit ribosomal RNA
NJ	Neighbour Joining
OP	Oropharnygeal
PaO_2	Oxygen Pressure
PBS	Phosphate Buffered Saline
PCP	*Pneumocystis* Pneumonia
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gel Electrophoresis
PRR	Proline Rich Region
PRT1	Protease1
RFU	Relative Fluorescence Units
RNA	Ribose Nucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RT	Reverse Transcriptase
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SCID mice	Severe Combined Immunodeficient mice
SDS	Sodium Dodecyl Sulphate
secs	Seconds
SIDS	Sudden Infant Death Syndrome

SOD	Superoxide Dismutase
SSC	Sodium chloride/ sodium citrate
SSCP	Single Strand Conformation Polymorphism
TdT	Terminal Transferase
TMP-SMZ	Trimethoprim-Sulfamethoxazole
TS	Thymidylate Synthase
U	Units
UCS	Upstream Conserved Sequence
UTR	Untranslated Region
vol	Volume
X-gal	5-Bromo-4-chloro-3-indoyl-β-D-galactopyranoside

2.2. Samples and Reagents

Table 2.2. Samples donated by other researchers

Researcher	**Sample Donated**
E.M. Aliouat and E. Dei-Cas	Eight clonal *Pneumocystis* populations using the low-dose methodology.
S. Banerji	Genomic λ DNA library
M.T. Cushion	Two preparations of rat-derived *Pneumocystis* genomic DNA, M54#17 and M59#5
Glaxo	Four preparations of rat-derived *Pneumocystis* genomic DNA, 848.3, 876.4, 910.4, 901.3
G. Smulian	cDNA library used to derive EST data Genomic cosmid library
S. Vargas	*P. jiroveci* infected whole lung tissue

Table 2.3. Work done by other researchers or collaborators

Researcher	Contribution to Project
E.M. Aliouat and E. Dei-Cas	Produced potentially clonal *Pneumocystis* populations using the low-dose methodology.
A. Lindley	Carried out the initial cloning and sequencing of some of the PRR amplification products.
R. Miller	Obtained all clinical samples, unless stated elsewhere.
V. Novelli	Obtained infant respiratory sample
R. Palmer	Sequenced 5'UTR of λ genomic DNA clones
J. Stringer and S. Keely	Generated the UCS/MSG data for the *Pneumocystis* simple populations
A. Tsolaki	Genotyped some of the respiratory samples from the PCP patients
A. Wakefield	Extracted DNA from clinical samples where there was an HIV-risk. Genotyped some of the respiratory samples from the PCP patients Extracted DNA from simple populations. Did the control PCR amplifications to determine if there was latent infection present in the *Pneumocystis* simple populations

2.2.1. Chemicals

All standard chemicals were purchased from Merck/BDH or Sigma unless otherwise stated.

Materials were supplied as follows:

Agarose, general purpose	Bioline
Bacto-yeast extract	Oxoid
Complete™ Protease Inhibitors	Roche
Dexamethasone	Organon
DNA molecular weight standards	GibcoBRL
Film for autoradiography	RX: Fuji, Biomax MR: Kodak
Hybond-N^+ nylon membrane	Amersham

Nutrient Agar	Oxoid
Oligonucleotides	Sigma-Genosys
Radiochemicals, α-^{32}P dATP	AP Biotech
SeaPlaque agarose	Flowgen
Sequagel gel system components	National Diagnostics
TRIzol	GibcoBRL
X-gal, IPTG, dNTPs, ddNTPs	Roche

2.2.2. Enzymes

Roche and New England Biolabs supplied restriction enzymes together with the appropriate buffer solutions. Most other enzymes used for molecular biological procedures were also purchased from Roche except for the following:

DNase I	Ambion
Klenow enzyme	Amersham
SuperScript™ RNase H$^-$ Reverse Transcriptase	GibcoBRL
Taq DNA polymerase	Promega

2.2.3. Standard Solutions

100 x Denhardts

2% Ficoll 400; 2% polyvinylpyrrolidone; 2% BSA

Hybridisation buffer

6 x SSC; 0.01 M EDTA; 5 x Denhardts; 0.5% SDS; 0.1 mg mL^{-1} Salmon Sperm DNA

PBS

137 mM NaCl; 2.7 mM KCl; 4.3 mM $Na_2HPO_4.7H_2O$; 1.4 mM KH_2PO_4

Red Blood Cell Lysis buffer for *Pneumocystis* parasite extraction

9 vol NH_4Cl 0.15 M: 1 vol Tris-HCl pH 7.4 20 mM

20 x SSC	3 M NaCl; 0.3 M Tri sodium citrate
1 x TBE	0.9 M tris-borate; 2 mM EDTA, pH 8.0
TE	10 mM tris base pH 8.0; 1 mM EDTA

2.2.4. Bacterial Media

LB	1% Bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl, NaOH to pH 7.5
LB-agar	As above with 1.5% bacto-agar
LB-amp	LB supplemented with 100 μg mL^{-1} ampicillin

LB-agar/amp/X-Gal/IPTG

LB agar supplemented with: 100 μg mL^{-1} ampicillin, 0.004% X-Gal, 40 μg mL^{-1} IPTG

2.3. Clinical Samples

2.3.1. Ethical Guidelines

For studies from samples on humans informed written consent was obtained from each patient or their parent and the studies were carried out within the guidelines of the local research ethics committee for each institution.

2.3.2. Mother and Infant Samples

The mother, a previously well 25 year old black African, had been resident in the U.K. for four years. Her infant was male who had gone to full-term, born April 1st 2000 and had a birth weight of 3.7 kg (50th percentile). The mother had declined an antenatal HIV test. The birth was problem free and the infant was subsequently breast and bottlefed. Thirty-one days after the birth, the infant presented with a 2 day history of non-productive cough, breathing difficulties and poor feeding. From the physical examination the infant had a respiratory rate of 120 min^{-1} and oxygen saturations (breathing room air) = 87%. The chest radiograph

showed interstitial pneumonitis with widespread ground-glass appearance. Mechanical ventilation was commenced on 5th May 2000 and a BAL sample was taken via the endotracheal tube. The sample contained *P. jiroveci* organisms, identified by methenamine silver staining, cytomegalovirus, *Candida albicans* and *Staphylococcus aureus*. The infant started a regime of parenteral high-dose cotrimoxazole, ganciclovir, liposomal amphotericin and adjuvant glucocorticoid, which lead to a full recovery. HIV type 1 antibodies were found in the infant's blood and HIV-1 RNA was detected, the plasma HIV-1 RNA level was 426,460 copies ml^{-1}. The infant had a $CD4^+$-T-lymphocyte count that was 32% of total with an absolute count of 1997 cells μl^{-1}. The child remained well after 17 months of follow up, the $CD4^+$-T-lymphocyte count was 32% of total with an absolute count of 2,929 cells μl^{-1} and the HIV-1 RNA levels were undetectable. The infant is currently receiving antiretroviral therapy and cotrimoxazole as prophylaxis for PCP.

On the same day that the infant was diagnosed with PCP, the mother presented with a thirty two day history of increased cough, exertional dyspnea with onset three days postpartum. Examination was normal apart from seborrhoeic dermatitis on her face. Investigations revealed radiographically diffuse pneumonia, hypoxemia and an oxygen pressure, PaO_2, level of 6.6 kPa (breathing room air). The BAL sample taken on 16th May 2000, 43 days after the onset of symptoms, revealed *P. jiroveci* organisms identified by methenamine silver staining. HIV antibodies were detected in the blood, the $CD4^+$ -T-lymphocyte count was 10 cells μl^{-1} and the plasma HIV-1 plasma RNA level was 87,600 copies ml^{-1}. Treatment was started with parenteral, high-dose cotrimoxazole and adjuvant methylprednisolone. The patient recovered. Cotrimoxazole was then used as secondary prophylaxis against PCP and antiretroviral therapy was started. The patient remains well 17 months later. The father tested negative for HIV.

Dr R.F. Miller performed a bronchoscopy on the mother while the bronchoscopy for the infant was done at Great Ormond Street Hospital.

2.3.3. Health Care Worker Samples

Induced sputum or nasal rinse samples were obtained from 27 health care workers, HCW, of whom 18 had been employed for over three months in a specialist HIV-AIDS inpatient care facility in London. The other 9 worked exclusively in the general medical and respiratory services that were based at a different site within the hospital and had no occupational contact with the HIV-AIDS inpatient facility. In total, samples were obtained on 36 occasions from the 27 HCW over a 2 year period.

To exclude an effect from underlying immunosuppression, HCW were not included in the study if they suffered from any of the following:

- If they were receiving inhaled, topical, or systemic corticosteroids or other immunosuppressive therapy.
- If they had diabetes mellitus, some chronic pulmonary disease or allergic rhinitis or sinustis.
- If they were pregnant.
- If they knew or suspected that they might have HIV infection
- If they had a current or past history of malignancy.
- If they had experienced an upper or lower respiratory tract infection within the 6 weeks prior to the study.

2.3.4. Patient Samples

Respiratory samples (BAL in 35 patients and induced sputum in 1 patient) were obtained from 36 adults who were HIV type 1 antibody-positive diagnosed with PCP. They were

admitted to the University College of London hospitals during the same time period as the HCW provided samples. All the patients had typical clinical presentations, response to specific anti-*P. jiroveci* therapy and positive results from methenamine silver staining of BAL fluid or induced sputum. Dr R.F. Miller performed the majority of the bronchoscopies. His clinical team who followed the protocol as described did the other bronchoscopies.

2.3.5. Post-mortem whole lung tissue sample

The lung tissue came from a girl, aged 2 years and 4 months. She was diagnosed with medulloblastoma of the brain stem in April 1997. Treatment of vincristine and cisplatinum was started on 11th March 1998. The patient subsequently developed multiple metastases to the spinal cord. On 25th May 1998, she was admitted to hospital with pneumonia and empirically commenced a regime of intravenous ceftriaxone and amikacin, together with high-dose trimethoprim sulfame-thoxazol. Two days later, mechanical ventilation was necessary as dyspnoea and hypoxemia were increasing. A BAL sample was taken, from which *P. jiroveci* organisms were identified by methenamine silver staining. Hydrocortisone was then added to the treatment. Unfortunately, the patient progressively deteriorated and by 1st June had developed total radiological opacity of both lungs. The patient died on 15th June.

An autopsy, limited to the lung was performed by Dr S. Vargas. The complete right lung was removed intact within 4 hrs of death. Each lobe was dissected separately and all necessary precautions were taken to eliminate cross-contamination between samples. Each lobe was divided into pieces (2-3 cm^3) which were stored at -70°C until analysis.

2.4. Techniques to Obtain Clinical Samples

At the time of sampling, an aliquot was saved from all samples. It was sealed and frozen immediately and then stored at -20°C until further analysis.

2.4.1. Bronchoalveolar lavage (BAL) samples

BAL samples were obtained using fibreoptic bronchoscopy, using a technique previously described (Miller *et al.,* 1991).

2.4.2. Induced Sputum (IS) samples

Induced sputum samples were obtained by inhaling 20-30 ml of 2.7% saline (Kendall laboratories) using an ultrasonic nebulizer (Ultraneb 99m, DeVilbliss) as previously described (Miller *et al.,* 1991, Wakefield *et al.,* 1991). To avoid contamination the nebulizer was washed in soapy water and sterilized by immersion in 3% glutaraldehyde between each saline induction and single-use disposable tubing and mouthpieces were used for each procedure. The procedure was supervised by Dr R.F. Miller.

2.4.3. Oropharyngeal (OP) samples

OP samples were obtained by Dr R.F. Miller. In summary, the patients gargled and rinsed their mouths with 10 ml sterile N-saline (Wakefield *et al.,* 1993).

2.4.4. Nasal Rinse Samples

Nasal rinse samples were supervised by Dr R.F. Miller. While the patient was standing erect and breathing through the mouth, they occluded the right nostril. Sterile normal saline (10 ml) was slowly instilled into the left nostril by using a 10 ml sterile single-use syringe for 20-30 secs. Refluxing saline was caught in a sterile universal container, which was held against the upper lip below the nostril.

2.5. Extraction of *Pneumocystis jiroveci* DNA from Human Samples

Prof. A.E. Wakefield extracted the DNA where there was an HIV risk from the respiratory samples. A category three laboratory had to be used for the extraction, which Professor Wakefield was trained to use. Samples from the whole lung tissue were mechanically disrupted using sterile scalpels and digested with 1 mg mL^{-1} Proteinase K, in 10 mM EDTA pH 8.0 and 0.5% SDS at 60°C overnight, after which a further 1 mg mL^{-1} Proteinase K was added and the samples were incubated for a further 24 hr. HCW and patient respiratory samples were treated by the same procedure but the Proteinase K and EDTA solution was added directly to the sample. DNA was purified by phenol-chloroform extraction. An equal amount of phenol and chloroform, which was equivalent to the amount of sample, was added to the sample. The sample was then mixed and spun at 10,000 g for 15 mins. The aqueous layer containing the DNA was extracted and then treated with 1 ml of a DNA binding resin, which is part of the Wizard® DNA Clean-Up system kit (Promega) (Palmer *et al.*, 2000b). The manufacturer's protocol for the Wizard® DNA Clean-Up system kit was then followed.

Alternatively, DNA from the mother and infant samples was extracted using the QIAamp DNA mini kit (Qiagen) following the manufacturer's protocol, with the exception of an extended Proteinase K step of 1 hr and DNA was eluted from the column in 50 µl, rather than 200 µl AE buffer (Qiagen). Precautions were taken throughout these procedures to eliminate the possibility of cross-contamination of samples. All handling of the samples took place in a laminar flow cabinet and negative controls where no tissue was included in the sample were included in the extraction procedure to monitor for contamination.

2.6. The Rat Model of *Pneumocystis* Infection

To obtain complex rat-derived *Pneumocystis* populations, *Pneumocystis* organisms were isolated from the lungs of infected rats. Male Sprague-Dawley rats were immunosuppressed using orally administered corticosteroids (2 mg L^{-1}), and bacterial infections were prevented by addition of amoxicillin (200 mg L^{-1}) or tetracycline (500 mg L^{-1}) for alternate months. Rat-derived *Pneumocystis* organisms are acquired naturally (Frenkel *et al.*, 1966). After 10-20 weeks, when the rats showed weight loss and signs of respiratory distress, they were sacrificed by CO_2 asphyxiation and the lungs excised. Rat lungs were washed in distilled water, dH_20 and either used immediately or stored at -80°C in Hank's buffer (Sigma).

2.6.1. Source of Rat-derived Simple *Pneumocystis* Populations

12 to 17 week old nu/nu Fisher rats (Pasteur Institute, Lille, France), free of *Pneumocystis* infections, were treated with dexamethasone for 5 weeks. Afterwards, they were anaesthetised and intratracheally inoculated with approximately 1 to 10 organisms per rat. The same preparation of diluted organisms was used to inoculate each rat. Nude rats were housed in barrier conditions within isolators. A positive pressure was maintained in the filtration chamber and each cage had a filtered cover to prevent airborne forms of *Pneumocystis* entering the cages. 8 weeks post-inoculation, the animals were sacrificed and organism burdens were enumerated by light microscopy. The average number of organisms per animal was $8 \pm 0.75 \times 10^8$. Sentinel rats, nude rats that were not inoculated, were used to monitor for external infection. In total, five simple populations of rat-derived *Pneumocystis* organisms were produced by this method. This work was kindly done by Eduardo Dei-Cas and El Mouktar Aliouat.

2.6.2. Other Rat-derived *Pneumocystis* Samples

Two preparations of rat-derived *Pneumocystis* genomic DNA, M54#17 and M59#5, were extracted from two infected rat lungs from a colony in the United States. They were kindly donated by M.T. Cushion. A selection of rat-derived *Pneumocystis* genomic DNA preparations, for example 876.4 and 848.3 were from a disbanded Glaxo rat colony. A rat colony was maintained in Oxford over a twelve year period. Rat-derived *Pneumocystis* genomic DNA was extracted from infected rat lungs. Each extracted rat-derived *Pneumocystis* population was numbered, the lower the number the earlier the DNA was extracted, for example, 1099, 1441, 1553 and 1697.

2.6.3. Rat-derived *Pneumocystis* DNA Libraries

The cDNA library used to derive EST data was constructed from RNA isolated from a single rat with a form1 rat-derived *Pneumocystis* infection by Dr. G. Smulian (Smulian *et al.*, 2001b). Dr G Smulian also created the genomic cosmid DNA library (Smulian *et al.*, 2001b). Sheared form1 rat-derived *Pneumocystis* genomic DNA was made blunt-ended and ligated into pWEB cosmid vector. Cosmids that contained *MSG* genes and *PRT1* genes were identified by hybridisation of bacterial colonies with DNA probes. The genomic λ DNA library used is described by Banerji (Banerji *et al.*, 1993). Briefly, purified rat-derived *Pneumocystis* DNA was partially digested with *Mbo*1 and size selected to yield 10-15 kb fragments. These were then cloned into EMBL3 bacteriophage vector cut with *Bam*H1.

2.6.4. Rat-derived *Pneumocystis* Parasite Purification from Rat Lung Tissue

The *Pneumocystis* infected rat lung was washed in sterile PBS and then minced using a sterile scalpel, in a laminar flow cabinet. To grind the lung tissue and dissociate the parasites from the host cells, the minced lung was placed in a beaker with PBS and was rapidly stirred at 4°C for an hour. The extraction took place on ice or at 4°C, and as far as possible, using

sterile conditions. The resulting mixture was filtered through gauze and the filtrate was spun for 15 min at 3,000 g. The supernatant was removed and the pellet was resuspended in 10 ml of lysis buffer with shaking for 10 min on ice. The mixture was spun for 15 min at 3,000 g at 4°C. The pellet was resuspended in 5 ml of PBS containing protease inhibitors. The suspension was filtered twice, once through a 70 μm cell strainer (Falcon) and subsequently through a 12 μm cell strainer (Falcon). The filtrate was placed into a VectaSpin™ 10 μm centrifuge filter (Whatman) and spun briefly at 1,000 g for 30 secs. The resulting eluate was spun at 3,000 g for 15 min at 4°C. The supernatant was removed. Aliquots were made from the pellet and were either used directly or were frozen at -80°C.

2.6.5. DNA Extraction from Rat-derived *Pneumocystis* Organisms

Rat-derived *Pneumocystis* DNA was prepared by resuspending approximately 50 μl of parasites in 100 μl of 1.2% molten SeaPlaque agarose in 125 mM EDTA, to prevent subsequent mechanical disruption of the DNA. The agarose blocks were suspended in 0.5 M EDTA, containing 1 mg mL^{-1} Proteinase K and 1% sarcosine and were then incubated at 37°C for 24 hrs. The buffer was then changed and the agarose blocks were left for a further 24 hrs to aid digestion of the cellular material. When digestion was complete the blocks were stored at 4°C in 0.5 M EDTA.

Prior to use, the agarose blocks were repeatedly washed in TE buffer to reduce the EDTA concentration. The agarose was melted at 70°C before 2 vol of water were added. The DNA solution was then stored at -20°C.

2.7. Amplification of *Pneumocystis* DNA Sequences

All Polymerase Chain Reaction (PCR) amplifications were performed using reagents at the following concentrations: 50 mM KCl, 10 mM TRIS (pH8.0), 0.1% Triton X-100, 3 mM $MgCl_2$, 0.4 mM of each of the four dNTPs, 1 μM oligonucleotide primers and 0.025 units mL^{-1} *Taq* polymerase. Template concentration varied depending upon the source of DNA. For second round PCR amplification, the product from the first round was used as a template. 30 μl of mineral oil was added on top of the reaction mix to prevent evaporation. Template DNA was added under the mineral oil using filtered pipette tips.

Great care was taken at all times to monitor for and prevent contamination. Negative controls where dH_2O was used in place of a template were included once every few samples, and all reagents were handled in a laminar flow cabinet, using sterile tubes, sterile pipette tips and aliquoted reagents. A Perkin Elmer Cetus PCR machine was used at all times, with all cycling conditions as follows, unless otherwise stated:

Denaturation	94°C	1.5 minutes	}
Primer Annealing	T_m of primer	1.5 minutes	} 40 cycles
Primer Extension	72°C	2 minutes	}

For the second round of the ITS nested PCR amplification the following reaction cycle was used; 10 cycles at 94°C for 1.5 min, 56°C for 1.5 min and 72°C for 2 min, and 30 cycles at 94°C for 1.5 min, 64°C for 1.5 min and 72°C for 2 minutes.

The approximate melting temperature, T_m of oligonucleotides was calculated from the sequence using the equation:

Tm = Σ ((No. of G bases + No of C bases) x 4) + ((No. of A bases + No. of T bases) x 2))

Table 2.4 shows the sequence, calculated melting temperatures and other information on the primers used.

All PCR amplification products were electrophoretically separated using a BioRad horizontal electrophoresis system, on 0.7-2.0% agarose gels depending on the size of the expected PCR amplification product, 1x TBE buffer and 0.1 μg mL^{-1} ethidium bromide were included in the gels. The DNA was visualised with UV light and the gel images were obtained using the Amersham Pharmacia Biotech Imagemaster VDS system.

2.7.1. PCR Amplification of the *Pneumocystis jiroveci* Genotyping Loci

The primers, ITSF3 and ITS2R3, were used for a single round ITS PCR to amplify the ITS regions, using the conditions described in section 2.7. Two rounds of amplification (nested PCR) were performed on samples when there was no or very little amplification product observed on an ethidium bromide stained agarose gel, after a single round of PCR amplification. In a nested ITS PCR, the first round amplification was performed with primer pair N18SF and N26SRX and the primer pair ITSF3 and ITS2R3 was used in the second round amplification.

The primer pair, pAZ102-H and pAZ102-E was used in a single round mt LSU rRNA PCR amplification of the mt LSU rRNA gene. In a nested mt LSU rRNA PCR amplification, the first round primers pAZ102-H and pAZ102-E were used, followed by the second round primers pAZ102-X and pAZ102-Y or pAZ102-X and pAZ102-Z, depending on which polymorphisms were being investigated.

Table 2.4. Oligonucleotide Primers Used

PRIMER NAME	LOCUS	TARGET DNA[1]	SEQUENCE (5' –3')	LENGTH (bp)	Tm (°C)	REFERENCE
pAZ102-H	mt LSU rRNA	*Pneumocystis*	GTGTACGTTGCAAAGTACTC	20	58	(Wakefield et al, 1990a)
pAZ102-E	mt LSU rRNA	*Pneumocystis*	GATGGCTGTTTCCAAGCCCA	20	62	"
pAZ102-X	mt LSU rRNA	*Pneumocystis*	GTGAAATACAAATCGGACTAGG	22	56	(Wakefield, 1996)
pAZ102-Y	mt LSU rRNA	*Pneumocystis*	TCACTTAATATTAATTGGGGAGC	23	56	"
pAZ102-Z	mt LSU rRNA	*Pneumocystis*	AAGCCCACTTCTTRACTGTC	20	56	"
DHPS3	DHPS	*Pneumocystis*	GCGCCTACACATATTATGGCCATTTTAAATC	31	55	(Beard et al, 2000)
DHPS4	DHPS	*P. jiroveci*	GGAACTTTCAACTTGGCAACCAC	23	55	(Beard et al, 2000)
N18SF	Nuc ITS rRNA	*P. jiroveci*	GGTCTTCGGACTGGCAGC	18	58	(Tsolaki et al, 1999)
N26SRX	Nuc ITS rRNA	*P. jiroveci*	TTACTAAGGGAATCCTTGTTA	21	58	"
ITSF3	Nuc ITS rRNA	*P. jiroveci*	CTGCGGAAGGATCATTAGAAA	21	56	"
ITS2R3	Nuc ITS rRNA	*P. jiroveci*	GATTTGAGATTAAAATTCTTG	21	56	"
HA1/ geno probe	*PRT1*	*P. carinii*	GGAATTTAACTCAAGATATTTT	22	57	new for study
HA2/ cDNA probe	*PRT1*	*P. carinii*	AAATGTAGTTATGCCAATGAT	21	57	"
HA3/ MID probe	*PRT1*	*P. carinii*	GAAATAAGATTGAAAACATTTTT	23	57	"

PRIMER NAME	LOCUS	TARGET DNA[1]	SEQUENCE (5'-3')	LENGTH (bp)	Tm (°C)	REFERENCE
HA4	*PRT1*	*P. carinii*	GATTAGTAAAAACAAGAATAATC	23	57	"
HA5	*PRT1*	*P. carinii*	CTAAAAAGTTAAGAAATTCTAA	22	57	"
PcProt1	*PRT1*	*P. carinii*	TTATTCTTGTAGCTGGGGAC	20	56	(Lugli et al. 1997)
PcProt3	*PRT1*	*P. carinii*	CACGCCATGTAAGATTAGGA	20	56	"
PcProt19	*PRT1*	*P. carinii*	GGTAAYTGGACTATCGATGT	20	56	"
PcProt31	*PRT1*	*P. carinii*	GAAGATGTTGATATTGAGGAG	29	58	lab primer
PcProt32	*PRT1*	*P. carinii*	ATCGTCTCTTATCGCACCC	27	58	"
PcProt48	*PRT1*	*P. carinii*	CCGGAACAAAACCAACATC	20	58	"
PcProt57	*PRT1*	*P. carinii*	CCGTTTTCTTAAAAATCCTATA	22	56	"
PcProt57-1	*PRT1*	*P. carinii*	TCCGTTTTCTTAAAAATCCTATA	23	57	new for study
PcProt58	*PRT1*	*P. carinii*	TTACTCTAACTAAATAGATCCAG	23	60	lab primer
PcProt62	*PRT1*	*P. carinii*	CTAACCTTCAGCCAGATTCAA	21	60	new for study
PcProt63	*PRT1*	*P. carinii*	CAGATCCAGATGCTACACTT	20	59	"
PcProt64	*PRT1*	*P. carinii*	CAGAACCAACTACTACGCTT	20	56	"
PcProt65	*PRT1*	*P. carinii*	TTCTATCTAGATATCCAACTAC	22	48	"
PcProt66	*PRT1*	*P. carinii*	AATCAGAAGGCGTACCGCC	19	66	"

PRIMER NAME	LOCUS	TARGET DNA[1]	SEQUENCE (5'-3')	LENGTH (bp)	Tm (°C)	REFERENCE
PcProt74	*PRT1*	*P. carinii*	CTAAATCAGAAGGCGTACC	19	56	"
PcProt83	*PRT1*	*P. carinii*	TTTACTCGCTTGACC	15	52	"
PcProt84	*PRT1*	*P. carinii*	AAAGTTTGTCACTATTAAACCC	22	56	"
PcProt85	*PRT1*	*P. carinii*	TTTCTTCCTTTTTCTAAACC	20	56	"
PcTel2	Subtelomeres	*P. carinii*	AAGTCACGTGCTCTCTTGGTCA	22	56	(Underwood et al 1996)
T3 Promoter primer	Universal		ATTAACCCTCACTAAAGGGA	20	56	(Sambrook et al,1989)
TS1	TS	*P. carinii*	CATTTATGGGTTTCAATGGAG	21	60	(Mazars et al, 1995)
TS2	TS	*P. carinii*	CGATGCAATATTAAAGGGAAC	21	60	"
18SRNA3	18S rRNA	*P. carinii*	CGCGGATCCAAAGATTAAGCCATGCATGTC	30	52	lab primer
18SrRNA4	18S rRNA	*P. carinii*	CGCGGATCCCTTGTTACGACTTTTACTTCC	30	52	"

[1] Source of DNA that the primers were designed to amplify.

For the DHPS locus only a single round PCR was required, due to the density of *P. jiroveci* organisms in the samples used. The primer pair, DHPS3 and DHPS4, was used in the single round DHPS PCR.

2.8. Cloning of Amplified DNA into Plasmid Vectors, PCR Screen and Plasmid Purification

PCR products were either cloned into a plasmid vector without purification or after they were purified using the SpinPrep™ PCR Clean-up Kit (Novagen) or were extracted from DNA fragments excised from agarose gels using the Wizard® PCR Preps DNA Purification System (Promega).

PCR products were ligated into the plasmid vector pGEM T-Easy (Promega) according to the manufacturer's protocol. The resulting plasmids were then transformed into competent *Escherichia coli* DH5α cells (Novagen) and plated onto LB/amp/X-Gal/IPTG plates and left to grow for 12-16 hrs at 37°C. White colonies not expressing the lacZ gene were picked and streaked onto fresh agar, then were screened by PCR amplification. For a screen by PCR amplification, an additional initial cycle of 94°C for 5 mins was included to break open the bacteria and release the plasmid DNA. For positive clones either the amplified DNA was sequenced directly or plasmid DNA was extracted from the respective colony and sequenced. PCR products were purified using the SpinPrep™ PCR Clean-up Kit (Novagen). For plasmid DNA, colonies were cultured in 10 ml of LB-amp at 37°C for 12-16 hrs. DNA was extracted from the cells using the Wizard® Miniprep kit (Promega) under conditions described by the supplier.

2.9. DNA Sequencing protocols

2.9.1. Genescanning

Genescanning sizes and quantifies DNA fragments to a resolution of 1 bp. PCR products are created using an oligonucleotide fluorescently labelled with 6-FAM. For each DNA sample, 0.5 µl of appropriately diluted PCR product are mixed with 0.25 µl of loading buffer (Applied Biosystems), 0.25 µl of 4 nM GeneScan-2500 TAMRA size standards (Applied Biosystems) and 1.5 µl of deionised formamide. One sample per lane is run on an ABI Prism™ automated sequencer and is subsequently analysed by GeneScan™ Analysis software. The overall results are seen as a gel image while each individual result can be viewed in a tabular format or as an individual electropherogram. The fluorescence is presented as a function of either time or size. Genescanning uses an internal size standard to perform precise size calling. A TAMRA fluorescent dye is used to label the internal size standards. The standards should negate band-shift artefacts and run-to-run variation.

2.9.2. Sequencing by the Dideoxy Chain Termination Method

Plasmid DNA and purified PCR products were sequenced using the BigDye Terminator Cycle sequencing kit (Applied Biosystems). For each sequencing reaction the following cycling conditions were used in a Biometra UnoII Thermoblock PCR machine with heated lid (no oil used):

Denaturation	96°C	30 seconds	}	
Primer Annealing	50°C	15 seconds	}	25 cycles
Primer Extension	60°C	4 minutes	}	

The sequencing reactions were ethanol precipitated according to the kit protocol with the aid of Pellet Paint NF Co-Precipitant (Novagen), and were then run on an ABI Prism 377 automated sequencing machine.

2.9.3. Sequence Analysis

Sequence analyses were performed using the ABI Prism™ 377 DNA sequencer running the data collection Software Version 2.1 (Applied Biosystems, Warrington, UK). Sequence data analysis was performed using Chromas 1.44 software (C.McCarthy, Griffith University, Australia). Sequence alignments were performed using the University of Wisconsin GCG sequence analysis Version 10.1, (Genetics Computer Group, Wisconsin, USA) and Pregap4 and Gap4 (Staden Package).

Phylogenetic analysis was performed using ClustalW running within the GCG package. ClustalW uses the Neighbour Joining (NJ) method of phylogenetic analysis. This produces a tree that has the smallest possible value for the sum of its branches, implying the shortest evolutionary path to explain the differences between the sequences (Saitou *et al.*, 1987). NJ trees can be bootstrapped to test the significance of the relationship between sequences. The same number of residues used for constructing the NJ tree, are randomly sampled and a new tree is produced from the resampled data. Any interior branch of the bootstrap that gives the same partition of sequences as the NJ tree is given a score of 1. Any branch that is different is scored 0. These bootstrap trials are repeated a thousand times. Results are presented as a number out of a thousand of the comparisons that scored 1.

TreeView was used to analyse the phylogenetic trees (Page, 1996). Divergence analyses were conducted using MEGA version 2.1 (Kumar *et al.*, 2001). Each sequence within a nucleotide alignment was compared with all the other sequences within the alignment and

the number of nucleotide differences between each sequence was calculated. Then the average nucleotide difference value was calculated and expressed as a percentage. All other analysis programs, unless stated, are part of the GCG Winconsin package and their use is explained in the text.

2.9.4. Synonymous versus Non-synonymous Nucleotide Changes

A nucleotide change that does not result in an amino acid change is a synonymous change while a nucleotide change that does result in an amino acid change is a non-synonymous change. These changes can be calculated within a population to produce a ratio of synonymous, (ds) and non-synonymous, (dn) changes i.e. ds/dn. These ds/dn ratios are a way of analysing the amount and type of selection a coding region is under. This analysis can be done using SNAP, a synonymous non-synonymous analysis programme (Nei *et al.*, 1986). If the value is greater than 1 the protein is under selective constraint. Synonymous changes are greater than non-synonymous, so the nucleotide sequence is changing more than the amino acid sequence. This implies that the function and structure of the protein cannot vary greatly. If ds/dn equals 1, the protein is evolving without constraint. The structure and function of the protein can tolerate changes both at the amino acid and nucleotide level or little or no change is occurring at all. Finally, if ds/dn is less than 1, positive selection is promoting amino acid replacement within the protein.

2.10. BAL-31 Nuclease digests of rat-derived *Pneumocystis* Genomic DNA

BAL-31 exonuclease degrades both 3' and 5' termini of duplex DNA without generating internal scissions. A digest was performed after melting 100 mg of agarose block containing rat-derived *Pneumocystis* genomic DNA by heating at 65°C for 10 min. Then 200 µl of BAL-31 buffer preheated to 65°C was added. The DNA solution was left to equilibriate at

45°C for 10 min. One 75 µl aliquot of the solution was pipetted into a preheated 0.5 ml tube while 15 units of BAL-31 was added to the remainder of the solution, which was then also aliquoted into 3x 75 µl lots. All four of the aliquots were incubated at 30°C. To monitor the progress of the digestion, the three aliquots containing the BAL-31 were removed at 5, 10, 15 and 20 min, respectively, while the aliquot with no BAL-31 was removed after 60 min. Each reaction was quenched by adding 7.5 µl 250 mM EGTA and was terminated by heating at 75°C for 10 min. The agarose was solidified in block moulds then placed into 400 µl *Pst*1 digestion buffer and left overnight at 4°C. The *Pst*1 buffer was changed three times, ensuring that the agarose blocks were fully equilibriated with buffer. The blocks were melted at 65°C for 10 min then left to cool to 37°C for 5 min. 500 units of *Pst*1 were added to each sample which were then incubated overnight. To confirm that BAL-31 exonuclease had degraded both the 3' and 5' termini of the duplex genomic DNA chromosome ends and that the *Pst*1 digest had gone to completion, the DNA was run on a 0.7% agarose gel and transferred onto a nylon membrane by Southern Blotting and then hybridised as described below.

2.11. Southern Blotting and Membrane Hybridisation

2.11.1. Transfer of DNA and Binding of Probes

Rat-derived *Pneumocystis* genomic DNA embedded in 1.2% SeaPlaque agarose (low-melt) in 125 mM EDTA was equilibriated with the appropriate restriction enzyme digestion buffer and digested overnight with the appropriate restriction enzyme at a concentration of one enzyme activity unit per µl of digest volume. The same amount of enzyme was added the following morning and the digest left for at least a further four hours. The agarose containing the appropriate restriction enzyme digested rat-derived *Pneumocystis* genomic DNA was melted, then run on 0.8% agarose gels and transferred to Hybond-N^+ nylon membrane by

capillary action using the alkali transfer method (Sambrook *et al.,* 1989). The resulting membrane was UV irradiated for 3 minutes to fix the DNA. Probes for hybridisation were either oligonucleotides or purified PCR products. Oligonucleotides were labelled with α-^{32}P dATP using the terminal transferase enzyme under conditions described by the supplier. A Sephadex®G-25 NAP™-5 column was employed to remove any excess radioactive nucleotide from the labelled probe. Purified PCR products were labelled with α-^{32}P dATP using the Klenow enzyme or the megaprime kit (Amersham Pharmacia), under conditions described by the supplier. A Sephadex®G-50 NICK™ column was employed to remove any excess radioactive nucleotide from the labelled probe.

For prehybridisation, the membrane was strategically placed with gauze in a hybridisation vessel, containing approximately 30 ml of hybridisation buffer. Temperatures for prehybridisation, hybridisation and washing were generally 65°C for double stranded probes and 45°C for oligonucleotides. The membrane was incubated in a Hybond oven for at least 4 hours, after which the relevant probe was added and the membrane was hybridised overnight. The membrane was washed at the same temperature as the hybridisation with first 6 x SSC and 0.2% SDS followed by 2 x SSC and 0.2% SDS. Each wash step lasted 15 minutes. Any variation on the washing steps is stated in the text. The membrane was then placed in a cassette with an intensifying screen, wrapped in Saranwrap and exposed to X-ray film. The cassette was placed in a freezer at -80°C for the required amount of time until the film was subsequently developed.

2.11.2. Stripping of Southern Blots

To remove unwanted probe so that a blot could be reused, the membrane was placed into 0.1% SDS at 95°C, and left to cool to room temperature for at least twenty mins. A film image of the membrane was created to check that the stripping had been effective.

2.11.3. Dot Blots

Dot blots were used to verify the efficacy of a probe. This was done using a BioRad BIO-DOT® apparatus, which works as a vacuum manifold. A Hybond-N^+ nylon membrane was cut to the appropriate size and submerged in dH_2O for 10 mins. A piece of Whatman 3 MM filter paper was cut to the same size and was also wetted in dH_2O. The piece of Whatman 3 MM filter paper was placed onto the apparatus and the membrane was placed on top. 500 µl of dH_2O was pipetted into each well of the apparatus and sucked through to wash the membrane. 1 M NaOH and 200 mM EDTA, pH 8.2, were added to each DNA sample to give a final concentration of 0.4 M NaOH/ 10 mM EDTA in a volume of 200 µl. The samples were heated for 10 mins at 94°C. The alkali heat treatment denatures the DNA and promotes binding of the DNA to the membrane. The DNA samples were pipetted into the wells and the DNA transferred onto the membrane under vacuum. Each well was rinsed with 0.4 M NaOH. The apparatus was then dismantled and the membrane was given a final rinse in 2 x SSC and airdried.

2.12. RNA Protocols

For all procedures involving RNA, utmost care was taken throughout to monitor for and prevent contamination. All handling of reagents occurred in a laminar flow cabinet, using sterile tubes, filtered pipette tips and aliquoted reagents.

2.12.1. RNA Extraction from Rat-derived *Pneumocystis* Parasites

1 ml of prewarmed Trizol (GibcoBRL) was added to 50 µl of parasites (frozen or fresh) and was incubated for 5 min at 37°C then placed in a freezer at -80°C overnight. 200 µl of chloroform was added, the mixture was vortexed and then centrifuged at 13,000 g for 30 min at 4°C. The aqueous upper layer containing the RNA was carefully removed and 500 µl isopropanol was added. This was vortexed and then incubated at 4°C overnight. It was spun at 13,000 g for 30 min at 4°C and the supernatant was removed. The resulting RNA pellet was airdried in the laminar flow cabinet then dissolved in 30 µl of RNase free water. The RNA was aliquoted into 3x 10 µl lots and was either used immediately or stored in a freezer at -80°C.

2.12.2. DNase Treatment of RNA

Using the DNA-free™ (Ambion) DNase, 1 µl of 10x DNase1 Buffer and 2 units of DNase1 enzyme were added to 8 µl of the RNA preparation. The reaction was mixed gently and then incubated at 37°C for 20-30 mins. To terminate the reaction, 1 µl of DNase inactivation reagent was added and incubated for 2 mins at room temperature with periodical mixing of the contents of the tube. The mixture was then centrifuged for 1 min at 13,000 g to pellet the inactivation reagent. The supernatant containing the RNA was used for further experimentation. Conditions for an alternative DNase, DNase1 (Ambion), were, 1 µl of 10x DNase1 Buffer and 2 units of DNase1 enzyme were added to 8 µl of the RNA preparation. The reaction was mixed gently and then incubated at 37°C for 20-30 mins. Heat treatment at 75°C for 5 minutes was used to terminate the reaction, with addition of EDTA at a final concentration of 5 mM prior to heating. This step may partially degrade the RNA by chemical scission.

2.12.3. Reverse Transcriptase Polymerase Chain Reaction, RT-PCR

For the first strand synthesis, 1 µl of 10 µM primer, 1 µl RNA and 9 µl of sterile filtered water were incubated at 70°C for 10 mins to denature the RNA. Then 4 µl of 5x reverse transcriptase First-Strand buffer, 1 µl of 10 mM dNTPS and 2 µl of 0.1 M DTT were added and incubated at 37°C for 2 mins. 200 U of reverse transcriptase were then added and the whole reaction was incubated for 1 hr at 37°C. Relevant controls were included to monitor for DNA and RNA contamination.

For the second strand synthesis, 5 µl of the first strand synthesis reaction were added to the corresponding PCR amplification, using reagents as previously described. Genomic DNA was used as a positive control and a negative control where dH_2O was used in place of a template, was prepared to monitor for contamination at this stage of the experiment. The products of RT-PCR were electrophoretically separated on a 1.5% TBE agarose gel.

2.12.4. 5' RACE Amplification of cDNA Ends

5'UTR ends of mRNA were amplified using the 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (GibcoBRL). Briefly, using known *PRT1* gene sequences, a specific primer, PcProt83, was designed to bind approximately 300 bp downstream of the start codon of the *PRT1* gene. This primed first strand cDNA synthesis, using 200 units of SuperScript™II Reverse Transcriptase, 1-5 µg of rat-derived *Pneumocystis* RNA and 100 nM PcProt83 incubated in a total volume of 15.5 µl at 42°C for 40 mins. 10 x PCR buffer, 2.5 mM $MgCl_2$, 400 µM of each dNTP and 10 mM DTT were also included. The reaction was terminated by incubation at 70°C for 15 mins. The RNA template was degraded with 1 µl RNase mix (GibcoBRL) at 37°C for 30 mins. The remaining cDNA was purified with a GlassMAX Spin Cartridge. Using 1 µl terminal transcriptase, TdT (GibcoBRL), a

dCTP tail was added to the cDNA by incubating at 37°C for 10 mins. The reaction was terminated by heat inactivating the TdT for 10 mins at 65°C. The dC-tailed cDNA was PCR amplified using a primer designed to bind to the dC-tail, the Abridged Anchor Primer, AAP, and a nested primer, PcProt84, which bound to the cDNA strand just upstream of PcProt83. The reaction was first heated to 94°C for two mins and then the following cycling conditions were used,

Denaturation	94°C	1 minute	}	
Primer Annealing	55°C	45 seconds	}	35 cycles
Primer Extension	72°C	2 minutes	}	

A final extension cycle of 72°C for 7 mins was added at the end to complete any strand synthesis. The primary PCR product was then reamplified using another set of specific nested primers, the Abridged Universal Amplification Primer, AUAP and PcProt85 with the same cycling conditions as the first amplification. The resulting cDNA product was cloned and sequenced.

Sample aliquots were taken at each stage of the procedure to monitor the efficiency of the experiment. Negative controls were used at the reverse transcriptase stage to monitor for DNA contamination and at the terminal transcriptase stage of the experiment to verify specificity of the PCR with AAP. Reactions using both control DNA and RNA were used to check the PCR reagents and cycling conditions.

3. Epidemiological Studies to Investigate the Transmission of *Pneumocystis jiroveci*

As discussed in chapter one, the acquisition and transmission of *P. jiroveci* infection is not well understood. An effective culture system has yet to be developed for *P. jiroveci* organisms so epidemiological studies are dependent on genotyping methods. The issue has yet to be resolved as to whether or not *Pneumocystis* Pneumonia, PCP, is the result of reactivation of latent infection or results from reinfection of *P. jiroveci* organisms. Asymptomatic carriage of organisms may occur in certain individuals, which could act as an infectious reservoir of organisms. To gain more data with which to clarify these issues three prospective studies were performed. Two of these studies looked at transmission of *P. jiroveci* organisms. The first investigated direct transmission between a mother and her infant and the other study investigated transmission between HCW, and their patients. The third study examined a distribution of *P. jiroveci* organism types within the human lung to assess how representative respiratory samples are of the *P. jiroveci* infection.

3.1.1. Probable Mother-to-Infant Transmission of *Pneumocystis jiroveci* Infection

BAL samples were taken from both the mother and infant who had PCP contemporaneously and *P. jiroveci* DNA was extracted (section 2.3.2). To characterise the infection, genotyping was performed at three independent loci, mt LSU rRNA, DHPS and ITS. For all three loci, there was an abundance of *P. jiroveci* organisms in the samples so a single round of PCR amplification was sufficient to generate enough PCR product for cloning or direct sequencing. Two separate PCR amplifications of the mt LSU rRNA locus were done, using the samples from the mother and infant. A single round of PCR amplification was carried out using the primers pAZ102-X and pAZ102-Y. Both samples were positive for *P. jiroveci* DNA as a product of 266 bp was generated by PCR amplification (Figure 3.1). The

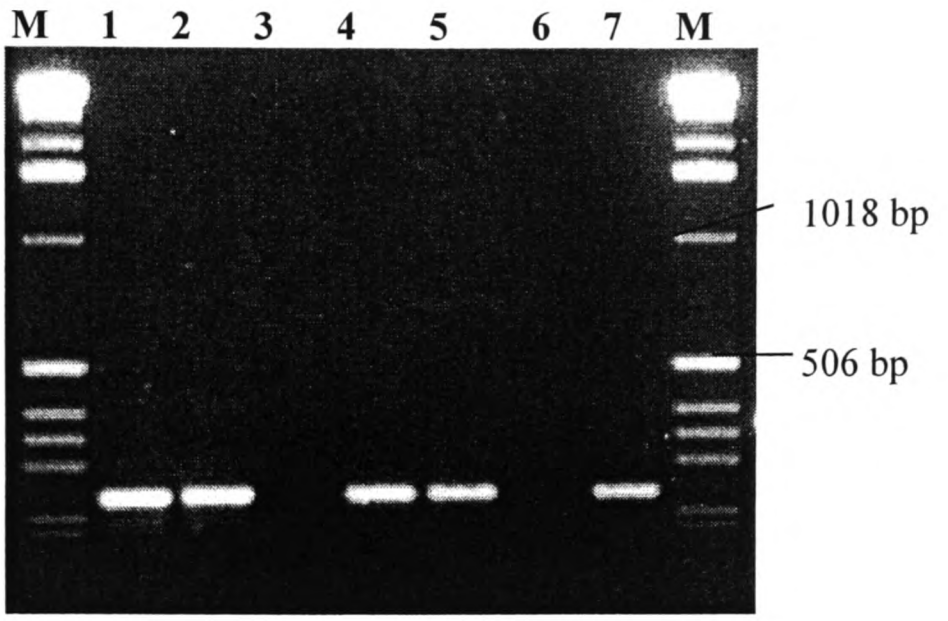


Figure 3.1. Electrophoretic separation of PCR amplification products of the mt LSU rRNA locus using primers pAZ102X and pAZ102Y M, 1kb DNA ladder; lanes 1 and 2: sample C218, (infant) undiluted and diluted x5, respectively; lanes 4 and 5: samples D967, (mother) undiluted and diluted x5, respectively; lanes 3 and 6: negative control using dH_2O in place of a template; lane 7: positive control using a *P. jiroveci* positive sample D129.

resulting PCR products were directly sequenced, again using the primers pAZ102-X and pAZ102-Y. The mt LSU rRNA locus has two scoring positions at nucleotides 85 and 248. This results in four genotypes out of a possible sixteen different genotypes (Beard *et al.*, 2000). Genotyping at the mt LSU rRNA locus, revealed the same type in both the mother and the infant samples. The DNA sequences from two PCR products each from the mother sample and infant sample, all had a C at position 85 and a C at position 248 (Table 3.I). This type is equivalent to genotype 1 (Beard *et al.*, 2000), and is one of the most common types in the United States.

Three separate PCR amplifications of the DHPS locus were carried out on the mother sample while two amplification reactions on the same locus were performed with the infant sample. The primers DHPS3 and DHPS4 were used for both the amplification reactions and for the direct sequencing reactions. Both samples were positive for the *P. jiroveci* DHPS locus as a product of 392 bp was generated (Figure 3.2). The DHPS locus has two scoring positions at nucleotide positions 165 and 171, which correspond to the amino acid residues 55 and 57, when the DHPS gene is translated. Four different genotypes have been defined (Beard *et al.*, 2000). As previously discussed (section 1.3.1), genotype I Thr/Pro, 'wildtype' corresponds to no previous sulphur prophylaxis while genotype 4 Ala/Ser, 'double mutant' corresponds to prior sulphur prophylaxis (Kazanjian *et al.*, 2000). Two 'single mutant' genotypes have also been observed, genotypes 2 and 3, which correspond to Ala/Pro and Thr/Ser respectively. The same DHPS genotype was present in the DNA sequence of three PCR products from the mother sample and in the DNA sequence of two PCR products from the infant sample. Both samples had an A at position 165 which corresponds to threonine at amino acid position 55 and a C at position 171 which corresponds to proline at amino acid position 57 (Table 3.1). These sequences are equivalent to the 'wild-type' DHPS sequence (Lane *et al.*, 1997). This genotype correlates with neither the mother nor the infant

MOTHER			INFANT		
mt LSU rRNA			**mt LSU rRNA**		
PCR Experiment	Nucleotide position/ Identity		PCR Experiment	Nucleotide position/ identity	
1	85/C 248/C		1	85/C 248/C	
2	85/C 248/C		2	85/C 248/C	
DHPS			**DHPS**		
PCR Experiment	Nucleotide position/ identity	Amino acid position/ identity	PCR Experiment	Nucleotide position/ identity	Amino acid position/ identity
1	165/A 171/C	55/Thr 57/Pro	I	165/A 171/C	55/Thr 57/Pro
2	165/A 171/C	55/Thr 57/Pro	2	165/A 171/C	55/Thr 57/Pro
3	165/A 171/C	55/Thr 57/Pro			
ITS			**ITS**		
Clone	ITS genotype		Clone	ITS genotype	
1	B_1a_3		1	B_1a_3	
2	B_1a_3		2	B_1a_3	
3	B_1a_3		3	B_1a_3	
4	B_1a_3		4	B_1a_3	
5	B_1b_1		5	B_1a_3	
6	B_1a_3		6	B_1a_3	
7	B_1a_3		7	B_1a_3	
8	Ca_3				
9	B_1a_3				
10	B_1a_3				

Table 3.1 Combined results of genotyping analysis at three *P. jiroveci* loci performed on BAL samples from a mother and her infant

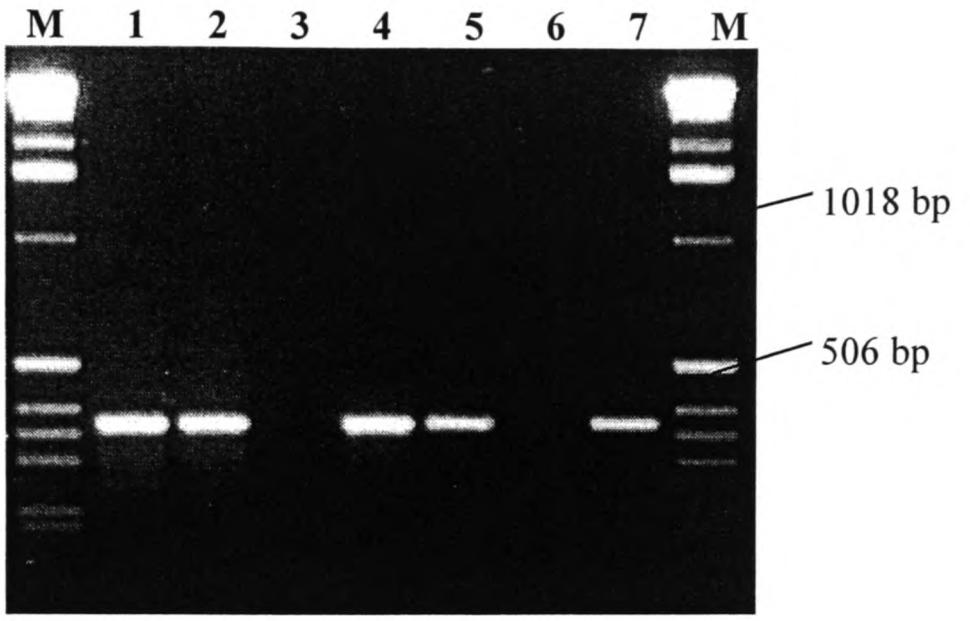


Figure 3.2. Electrophoretic separation of PCR amplification products from the dihydropterate synthase gene, DHPS, using primers, DHPS3 and DHPS4 M, 1kb DNA ladder; lanes 1 and 2: sample C218, (infant) undiluted and diluted x5, respectively; lanes 4 and 5: samples D967, (mother) undiluted and diluted x5, respectively; lanes 3 and 6: negative control using dH_2O in place of a template; lane 7: positive control using a *P. jiroveci* positive sample D129.

having prior sulphur exposure.

The ITS locus was amplified with the primers ITSF3 and ITS2R3. Both samples were positive for *P. jiroveci* DNA as a product of ≈ 500 bp was generated (Figure 3.3). The amplification products were cloned and sequenced as previously described. Ten plasmid clones were sequenced from the mother sample while 7 clones were sequenced from the infant sample. The ITS PCR product divides into two ITS regions, ITS1 and ITS2 with the 5.8S rRNA gene in between. The ITS locus is highly polymorphic with many scoring positions on both ITS1 and ITS2. This results in a number of ITS1 and ITS2 genotypes (Table 3.2 and Table 3.3), which are combined to form an overall ITS type. A total of 42 ITS genotypes have been described by this method of typing (Tsolaki *et al.*, 1996), which is comparable to the numbers described using other ITS typing systems (Nimri *et al.*, 2002, Hosoya *et al.*, 2000, Lee *et al.*, 1998).

The ITS genotype for all seven plasmid clones from the infant sample was B_1a_3 (Table 3.1), which corresponds to type Eg (Lee *et al.*, 1998). This genotype occurs in 16% of United Kingdom samples (Miller and Wakefield, 1999a), and up to 27% of samples from the United States and Denmark (Lee *et al.*, 1998). The majority ITS genotype in the mother samples was also B_1a_3. This was found in 8 out of the 10 plasmid clones. The other two plasmid clones were Ca_3 and B_1b_1, which correspond to Fg and Eb respectively (Lee *et al.*, 1998). This implies that the infection within the mother was not clonal.

The genotyping results suggest that the infant was infected with a single type of *P. jiroveci* while the mother was infected by more than one type of *P. jiroveci*. This single type infecting the infant was the same as the majority type infecting the mother, indicating that

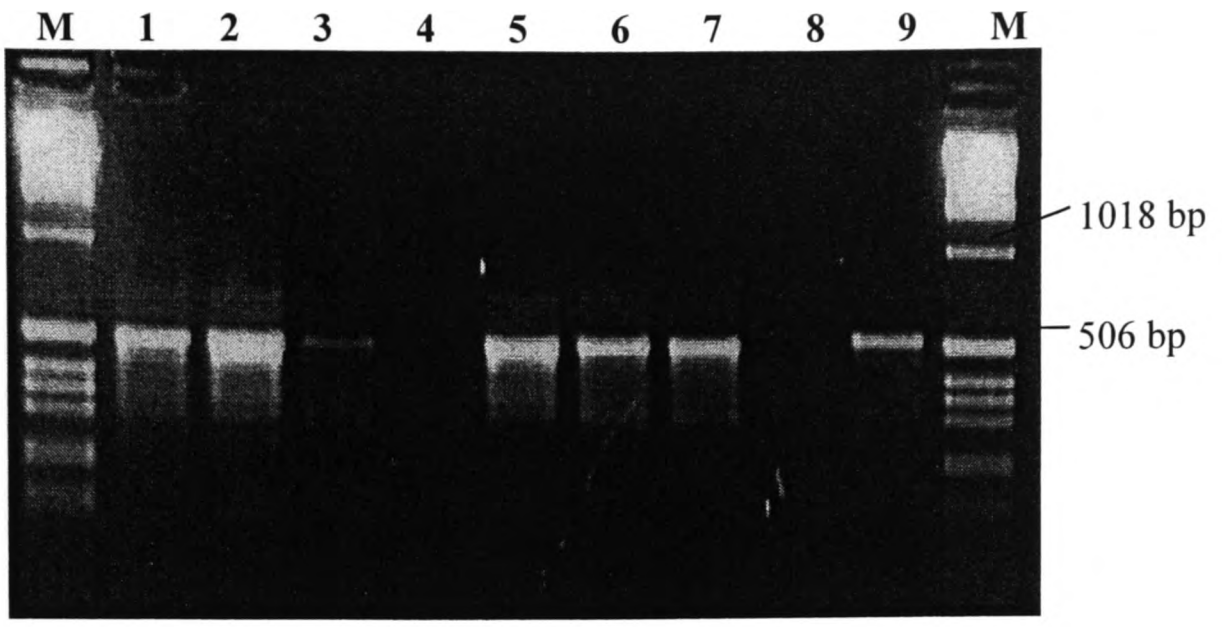


Figure 3.3. Electrophoretic separation of PCR amplification products from the internal transcribed spacer (ITS) regions of the nuclear rRNA, using primers ITSF3 and ITS2R3 M, 1kb DNA ladder; lanes 1, 2 and 3: sample C218, (infant) undiluted, diluted x5 and x25, respectively; lanes 5, 6 and 7: samples D967, (mother) undiluted, diluted x5 and x25, respectively; lanes 4 and 8: negative control using dH_2O in place of a template; lane 9: positive control using a *P. jiroveci* positive sample D129.

Table 3.2. ITS1 sequence types of *P. jiroveci*

	Nucleotide position (bp)						
ITS1 type	**6**	**12**	**21-22 †**	**23**	**28**	**83-84**	**120-122**
A_1	C	T	-	T	T	--	TTA
A_2	C	-	-	T	T	--	TTA
A_3	T	-	-	A	T	--	TTA
A_4	T	-	-	T	T	--	TTA
B_1	T	-	-	T	T	AG	TTA
B_2	T	-	-	A	T	AG	TTA
B_3	C	-	-	T	T	AG	TTA
B_4	T	-	-	G	T	AG	TTA
B_5	T	-	-	T	T	TG	TTA
B_6	T	A	-	T	T	AG	TTA
B_7 *	T	-	TT	T	C	AG	TTA
C	T	-	-	T	T	AG	---

- nucleotide missing at this position; † new scoring position;* new sequence type identified in this study

Table 3.3. ITS2 sequence types of *P. jiroveci*

	Nucleotide position (bp)									
ITS2 type	54-59	63-64 †	68	72-78	91 †	129-130	180-183	184-185†	187	191-197
a_1	TAA---	--	-	AATAT--	T	-T	----	AT	G	-------
a_2	TAA---	--	-	AATAA--	T	-T	----	AT	G	-------
a_3	TAA---	--	-	AATAA--	T	-T	----	AT	A	-------
a_4	TAA---	--	-	AATAT--	T	-T	----	AT	A	-------
a_5	TAA---	--	-	AATATAT	T	-T	----	AT	G	-------
a_6	TAG---	--	-	AATAT--	T	-T	----	AT	A	-------
a_7*	TAA---	--	-	AATAA--	T	C-	----	AT	A	
a_8*	TAA---	--	-	AATAT--	C	C-	----	AT	G	-------
a_9*	TAA---	--	-	AATAA--	T	-T	----	--	A	-------
a_{10}*	TAA---	TA	-	AATAT--	T	-T	----	--	A	-------
b_1	TAA---	--	-	AATAT--	T	-T	--AT	AT	G	-------
b_2	TAA---	--	-	AATAT--	T	CT	--AT	AT	G	-------
b_3	TAA---	--	-	AATGT--	T	CT	--AT	AT	G	-------
c_1	------	--	A	-------	T	-T	----	AT	A	-------
c_2	------	--	A	-------	T	-T	----	AT	G	-------
d_1	TAA---	--	-	AATAT--	T	-T	ATAT	AT	G	-------
d_2	TAA---	--	-	AACAT--	T	-T	ATAT	AT	G	-------
e_1	TAATAA	--	-	AATAA--	T	-T	----	AT	A	CAAAATA
e_2	TAA---	--	-	AATAT--	T	-T	----	AT	A	CAAAATA
e_3	AAATAA	--	-	AATAA--	T	-T	----	AT	A	CAAAATA
f	TAA---	--	-	AATAA--	T	-T	----	AT	A	----ATA

- nucleotide missing at this position; † new scoring position; * new sequence type identified in this study

the infection was transmitted from the mother to the infant. The finding is supported by the data from the 3 genotyping loci. The infant may have been infected by more than one type yet the minority types were undetected.

3.1.2. *Pneumocystis jiroveci* DNA in Immunocompetent Health Care Workers in Contact with Patients with *Pneumocystis* Pneumonia

The possible transmission of *P. jiroveci* from PCP patients to asymptomatic, immunocompetent HCW, with or without occupational exposure to HIV patients with PCP, was investigated. Respiratory samples from HCW (section 2.3.3.), employed in a specialist inpatient HIV-AIDS facility were compared with respiratory samples from a control group of HCW employed in a general medical-respiratory service in the same hospital. Induced sputum and nasal rinse samples were analysed for the presence of *P. jiroveci* DNA, using DNA amplification of the gene encoding mt LSU rRNA. Positive samples were subsequently genotyped at the ITS locus and compared with the ITS genotypes from BAL samples obtained from patients with PCP who were at the hospital at the same time.

At the time of study, patients with suspected or confirmed PCP were cared for on an open ward. Bronchoscopy and sputum induction were carried out in a dedicated room, which shared ventilation with the rest of the ward. Bronchoscopists and staff supervising sputum induction did not wear high efficiency particulate air, HEPA, filter masks.

The mt LSU rRNA locus was analysed to try to detect *P. jiroveci* DNA as previously described (Wakefield *et al.,* 1990). Respiratory samples positive for the presence of *P. jiroveci* DNA were genotyped at the ITS locus. Respiratory samples from 36 adult PCP patients were also genotyped (section 2.3.4.), who were admitted to the hospital during the same time period as the HCW. Prof. A.E. Wakefield and Dr Anthony Tsolaki genotyped 30

of the 36 patient samples at the ITS locus. The remainder were genotyped as part of this project. The ITS PCR products were cloned (section 2.8.) and four recombinant clones were sequenced per positive respiratory sample (three for E101N).

P. jiroveci DNA was detected in 11 HCW samples, 10 were from HCW who had occupational contact with PCP patients and 1 sample was from the control group (Table 3.4). 7 of the 11 samples were from two HCW, both of whom were bronchoscopists (doctor 1 and doctor 10). This compares with the bronchoscopists from the control group, none of who were positive for *P. jiroveci* DNA. Nurse 7 from the contact group was subsequently found to have occult malignancy so was potentially immunocompromised, possibly explaining why the sample was positive for *P. jiroveci* DNA. The reason why one sample from the control group contained *P. jiroveci* DNA was not apparent. Overall, in the contact group, 4 out of 17 HCW (0.24) were positive for *P. jiroveci* DNA. The sample from nurse 7 has not been included. This compares with the control group where 1 out of 9 HCW (0.11) were positive for *P. jiroveci* DNA. In the follow up period (median 98 months, range 15-113 months), none of the HCW developed PCP.

In order to investigate whether direct transmission of *P. jiroveci* was occurring between PCP patients and HCW, ITS genotyping was used to examine the types of *P. jiroveci* detected in contact HCW and infected patients who were hospitalised at the time of sampling (Table 3.5). ITS PCR amplification is not as sensitive as mt LSU rRNA PCR amplification because the mt LSU rRNA locus is multicopy while the ITS locus is postulated to be present in a single copy (Nahimana *et al.,* 2000, Tang *et al.,* 1998, Giuntoli *et al.,* 1994). Therefore, the ITS locus is a more sensitive indicator of the amount of *P. jiroveci* DNA present within the sample, as many *P. jiroveci* organisms need to be present for an ITS locus PCR product to be generated (Tsolaki *et al.,* 1999, Lu *et al.,* 1995). In contrast, only a few *P. jiroveci* organisms

Sample	Date of Sample (mo.yr)	*P.jiroveci* DNA detected	Health Care Worker	Contact with PCP patients
E5	Dec 1991	x	Doctor 4	x
E6	Dec 1991	x	Doctor 5	x
E10	Dec 1991	x	Doctor 6	x
E56	March 1993	x	Nurse 3	x
E67	Apr 1993	x	Technician	x
E68	Apr 1993	x	Pharmacist	x
E73	Apr 1993	x	Nurse 8	x
E74	Apr 1993	✔	Nurse 9	x
E79	Apr 1993	x	Social Worker	x
E1	Dec 1991	x	Physio 1-1	✔
E2	Dec 1991	✔	Doctor 1-1	✔
E3	Dec 1991	x	Doctor 2-1	✔
E4	Dec 1991	x	Doctor 3	✔
E7	Dec 1991	x	Nurse 1	✔
E11	Dec 1991	x	Doctor 7-1	✔
E32	March 1992	x	Physio 1-2	✔
E33	March 1992	✔	Doctor 1-2	✔
E43	March 1993	✔	Doctor 1-3	✔
E49	March 1993	x	Nurse 2	✔
E57	March 1993	x	Doctor 8	✔
E58	March 1993	x	Nurse 4	✔
E60	March 1993	✔	Secretary	✔
E61	Apr 1993	✔	Nurse 5	✔
E62	Apr 1993	x	Nurse 6	✔
E63	Apr 1993	✔	Nurse 7	✔
E66	Apr 1993	x	Physio 2	✔
E75	Apr 1993	x	Doctor 7-2	✔
E76	Apr 1993	x	Doctor 9	✔
E78	Apr 1993	x	Doctor 2-2	✔
E81	Sept 1993	x	Nurse 10	✔
E86	Sept 1993	x	Nurse 11	✔
E88	Sept 1993	✔	Doctor 1-4	✔
E94	Nov 1993	✔	Doctor 10-1	✔
E97	Jan 1994	✔	Doctor 10-2	✔
E101	March 1994	✔	Doctor 1-5	✔
E116	July 1994	x	Doctor 1-6	✔

Table 3.4. Presence of *P. jiroveci* DNA in respiratory samples from Health Care Workers x – no *P. jiroveci* DNA detected, ✔ – *P. jiroveci* DNA detected

Date of Sample (day.mo.yr)	Patients with PCP		HCW	
	Sample Code	ITS genotype	Sample	ITS Genotype
24.5.91	D74	A_2c_1		
26.9.91	D122	Ca_3		
16.10.91	D129	B_1a_4 B_1e_1		
22.11.91	D138	B_1a_3 B_1c_1 B_1b_1		
17.12.91			Doctor 1-1	B_1a_3 A_4a_3
6.1.92	D148	B_1a_3 Ca_3		B_1a_{10} A_4a_9
29.5.92	D204	B_2a_1 B_2a_2		
22.7.92	D222	B_1d_1 B_2a_1		
29.7.92	D224	B_2a_1		
2.9.92	D229	B_1a_3		
4.11.92	D253	B_1a_3 B_1c_1		
27.11.92	D262	B_2a_1		
5.2.93	D283	B_2a_1		
12.2.93	D288	B_1b_1 B_2a_1		
19.2.93	D289	B_1b_2		
19.2.93	D290	A_3a_1 B_1a_1 B_2a_1 B_2a_4		
24.2.93	D295	B_2a_1		
24.2.93	D296	B_2a_8 B_2a_1		
19.3.93	D298	B_1d_1 B_2a_1		
22.3.93			Doctor 1-3	A_4a_3 B_7a_3
19.4.93	D303	B_2a_5		
5.5.93	D310	B_1a_1 B_1a_3 B_1e_1 B_1e_2 B_1e_3 B_2a_1 B_2a_2 B_2a_3 B_5a_3		
9.6.93	D321	B_1b_2 B_2a_1		
15.6.93	D328	B_2a_1		
15.6.93	D329	B_2a_1		
15.7.93	H27	B_1b_1		
6.8.93	D342	B_2a_1		
8.9.93	D347	B_2a_1		
23.9.93			Doctor 1-4	B_7a_{10}
29.9.93	D349	B_1a_1 B_1a_2 B_1a_3 B_2a_1 B_2a_3 B_4a_3		
4.10.93	D351	B_1b_1		
6.10.93	D352	A_4a_3		
18.11.93			Doctor 10-1	A_1c_1
5.1.94	D386	A_2b_2 A_2c_1 B_1b_2 B_3b_2		
16.3.94	D410	B_1a_3		
23.3.94			Doctor 1-5	B_1a_7 B_1a_{10}
25.3.94	D413	B_1a_3		B_7a_{10}
3.6.94	D438	B_2a_1		
6.7.94	D445	B_2a_1		
5.10.94	D474	B_1b_2		
26.10.94	D482	B_2a_1		

Table 3.5. ITS genotypes of *P. jiroveci* from asymptomatic HCW and patients with PCP

need to be present for an mt LSU rRNA PCR product to be generated. Often this difference in PCR amplification sensitivity corresponds with the clinical distinction between carriage and infection. Only 5 of the 11 positive HCW samples contained enough *P. jiroveci* DNA to be genotyped at the ITS locus. All 5 samples were from the two bronchoscopists, doctor 1 and doctor 10.

From all the respiratory samples, 31 ITS genotypes were detected, 5 of which were first identified in this study. A new 'B' genotype was designated to ITS1 sequence types and 4 new 'a' genotypes were designated to ITS2 sequence types (Table 3.2 and Table 3.3, new genotypes in blue). 26 ITS genotypes were identified from the patient samples. Mixed infection occurred in 16 out of 36 patients (44%). B_2a_1 was the most frequent ITS genotype detected in 20 out of 36 patients (55%). It was present as a single infection in 11 of the 20 samples (55%). Conversely, B_2a_1 was not identified in the HCW samples. Eight different ITS genotypes were detected in the HCW samples. Mixed infections were present in 3 out of 5 samples (60%). Three genotypes identified were present in the patient samples, A_1c_1, B_1a_3 and A_4a_3 while five genotypes identified were not present in the patient samples, A_4a_9, B_1a_7, B_1a_{10}, B_7a_3 and B_7a_{10}. There was a predominance of ITS1 'A' genotypes in the HCW samples.

Within the time frame of this study, the ITS genotyping data presents a complicated and dynamic picture of *P. jiroveci* transmission. The genotype B_1a_3 occurs in the patient group and the HCW group. It is present in a patient in Nov 1991 and is then identified in a HCW in Dec 1991 and finally is detected in patient samples in Jan 1992, Sept 1992 and Nov 1992. Yet, many other ITS genotypes were detected in this time period and B_1a_3 is a common genotype, occurring in 20% of samples (Lee *et al.*, 1998). The overall data set is complex and gives no clear indication of direct transmission of *P. jiroveci* organisms.

Two of the six positive contact HCW were bronchoscopists. Doctor 1 maintained asymptomatic *P. jiroveci* DNA carriage for 27 months, from which 7 ITS genotypes were identified (Table 3.6). During that time the HCW had no intercurrent illnesses and did not take immunosuppressive drugs or any antimicrobials. ITS genotyping was not possible with two of the samples that were taken from these HCW. This is a reflection of the low levels of *P. jiroveci* DNA present within these samples, and strongly supports the concept of low level carriage of *P. jiroveci* organisms.

P. jiroveci DNA was detected in 11 HCW samples and an ITS genotype was characterised from five of those samples. There was no direct correlation between the ITS genotypes of the HCW samples and the ITS genotypes of the patient samples, implying that transmission and acquisition of *P. jiroveci* organisms is not simple. Seven of the samples containing *P. jiroveci* DNA were from two bronchoscopists, suggesting that HCW in close occupational contact with PCP patients may become colonised with *P. jiroveci.* The results do not exclude the possibility of other transmission routes, such as transmission from asymptomatic carriers of *P. jiroveci* to immunocompromised patients, or the possibility of exposure to exogenous environmental reservoirs of infectious organisms.

3.1.3. Distribution of *Pneumocystis jiroveci* Types in the Lung of a Child Dying of *Pneumocystis* Pneumonia

A whole lung tissue was used to look at the relative distribution of *P. jiroveci* organisms. The lung tissue was from a Chilean child, immunosuppressed by cytotoxic drugs rather than HIV infection (section 2.3.5.). ITS genotyping was carried out on ten different samples from the whole lung to investigate the pattern of *Pneumocystis* organisms in the lung, to identify whether or not there was more than one genotype present and to examine the distribution of the different genotypes in the lung.

HCW	Sample	Date	mt LSU rRNA	ITS	ITS types
Doctor1	E2N	17.12.91	✔	✔	B_1a_{10}, B_1a_3, A_4a_3, A_4a_9
	E33	6.3.92	✔	x	
	E43N	22.3.93	✔	✔	A_4a_3, B_7a_3
	E88	23.9.93	✔	✔	B_7a_{10}
	E101N	23.3.94	✔	✔	B_1a_7, B_1a_{10}, B_7a_{10}
	E116N	15.7.94	x	x	
Doctor 10	E94	18.11.93	✔	✔	A_1c_1
	E97	24.1.94	✔	x	

Table 3.6. Detection of *P. jiroveci* DNA and subsequent results from ITS genotyping in two asymptomatic HCW who performed bronchoscopies on patients with PCP. ✔ - *P. jiroveci* DNA was amplified at this locus, x. – *P. jiroveci* DNA was not amplified at this locus.

The diagnosis of PCP was made by methenamine silver staining of the BAL sample. The diagnosis was confirmed by PCR at the mt LSU rRNA locus, (Wakefield *et al.,* 1990). Samples were analysed from the right lung, four from the upper lobe, three from the middle lobe and three from the lower lobe. *P. jiroveci* DNA was present in all ten samples. This reflects the severity of the disease.

Four ITS recombinant clones were sequenced per sample (Table 3.7.) the genotype Ca_3 was present in three of the samples; B_1b_2 in another three and four samples contained a mixture of the two types. Ca_3 was mainly detected in the lower lobe, B_1b_2 was mainly detected in the middle lobe and the upper lobe was a mixture of the two types.

3.2. Discussion

Three separate studies have been completed. In the first study, the genotyping results strongly suggest that the infant was infected with a single type of *P. jiroveci,* which was the same as the majority type infecting the mother. The data implies that the infection was transmitted from the mother to the child. In the second, *P. jiroveci* organisms were detected in HCW. The carriage is asymptomatic and it is possible but not proven that the organisms may have come from patients who were in close contact with the HCW. The whole lung study characterised a representative distribution of *P. jiroveci* infection and hints at the interaction of two different genotypes distributed within the human lung.

3.2.1. Probable Mother-to-Infant Transmission of *Pneumocystis jiroveci* Infection

It is a unique clinical observation that the mother and infant had PCP contemporaneously. Yet, it is a common clinical phenomenon for the mother to be unaware of her own and her infant's HIV status as PCP is a frequent AIDS indicator disease. Among infants born in the

Site in Lung	Presence of *P. jirovecii*	*P. jiroveci* ITS genotype
Upper Lobe		
Apical segment	✔	Ca_3, B_1b_2, B_1b_2, B_1b_2
Anterior segment, upper	✔	Ca_3, Ca_3, B_1b_2, B_1b_2
Anterior segment, lower	✔	Ca_3, Ca_3, Ca_3, Ca_3
Posterior segment	✔	B_1b_2, B_1b_2, B_1b_2, B_1b_2
Middle Lobe		
Medial segment, upper	✔	B_1b_2, B_1b_2, B_1b_2, B_1b_2
Medial segment, lower	✔	Ca_3, B_1b_2, B_1b_2, B_1b_2
Lateral segment	✔	B_1b_2, B_1b_2, B_1b_2, B_1b_2
Lower Lobe		
Anterior basal segment	✔	Ca_3, Ca_3, Ca_3, Ca_3
Medial basal segment	✔	Ca_3, Ca_3, Ca_3, Ca_3
Posterior basal segment	✔	Ca_3, Ca_3, Ca_3, B_1b_2

Table 3.7. Distribution of *P. jiroveci* genotypes in the whole lung tissue ✔ - *P. jiroveci* DNA was detected.

UK and Ireland with PCP, 79% of the mothers were unaware of their HIV status (Williams *et al.,* 2001). A study from the United States showed that 44% of infants with PCP had unknown HIV serostatus at diagnosis (Simonds *et al.,* 1993), while a similar figure of 40% was cited in a study from the British Isles (Gibb *et al.,* 1994).

The early presentation of PCP in the infant and the infant's survival despite severe illness is unusual. The US study showed that over 50% of PCP cases occur between the ages of 3–6 months (Simonds *et al.,* 1993). Three other studies calculated the median age of onset to be 4.1 months (Gibb *et al.,* 1994), 5.0 months (Scott *et al.,* 1989), and 7.7 months (Connor *et al.,* 1991). There is approximately 7 – 20% risk of contracting PCP in the first year of life for infants with perinatally acquired HIV (Simonds *et al.,* 1993). There is a high mortality rate for infants with PCP. The British Isles study quotes a figure of 82% (Gibb *et al.,* 1994) and another United States study has a comparable figure of 89% (Connor *et al.,* 1991).

From the time course of the clinical symptoms, it is highly probable that the mother had *P. jiroveci* infection before the time of delivery. By analogy with animal models (Walzer, 1984, Hughes *et al.,* 1983) the airborne transmission of *P. jiroveci* is the most likely explanation of the infection passing from the mother to the infant. Also, the mother was breast-feeding so she and the infant would have been in close proximity to each other. The mother had a mixed infection identified at the ITS locus while the infection in the infant was clonal. It is possible that the infant acquired a subset of *P. jiroveci* types from the mother. There are alternative explanations for the disease transmission between the mother and infant. There could have been a common exogenous source, or transmission may have occurred via the placenta, blood during delivery, or via the breast milk.

In animal models, *Pneumocystis* is acquired at various ages. In the rat model, rat-derived *Pneumocystis* organisms are acquired after birth, presumably via airborne transmission from the mother (Hughes, 1982). Rat-derived *Pneumocystis* DNA is detected in 78% of pups, 1-2 hours after birth and is present in 100% of pups, 24 hours after birth (Icenhour *et al.*, 2002). However, none of the 48 rat foetuses, placentas or deciduae investigated was positive for *Pneumocystis* DNA. Rabbit-derived *Pneumocystis* is transmitted *in utero* (Cere *et al.*, 1997) as detected by PCR amplification, histology and immunofluorescence and can be found as early as the 10^{th} day of pregnancy. Yet, rabbits are unusual in that they acquire *Pneumocystis* without being immunosuppressed (Soulez *et al.*, 1988). With *P. jiroveci*, it is not known when the organisms can first be acquired.

The possibility of transplacental acquisition of *P. jiroveci* infection in infants is not ruled out. An HIV positive infant was born with meconium aspiration (Beach *et al.*, 1991) and PCP was diagnosed 19 days later. An HIV positive mother died of PCP when 27 weeks pregnant (Mortier *et al.*, 1995). The foetus was born dead and *P. jiroveci* organisms were identified in the lungs but without associated pneumonitis, and *P. jiroveci* organisms were also detected in the placenta. This raises the intriguing possibility that the infant inhaled the organisms from the amniotic fluid. Yet, a critique of the study (Hughes, 1995a) noted that the colour photomicrograph used to demonstrate the presence of *Pneumocystis* cysts does not show organisms recognisable as *Pneumocystis* and suggests the organisms present may be nonbudding yeast that can resemble *Pneumocystis* cysts. In another family, three siblings each developed PCP, one within three days of birth (Bazaz *et al.*, 1970). There is a possible genetic immunological incompetence in the family, or possible *in utero* transmission of *P. jiroveci*. However, none of the placentas were examined histologically and the mother did not have PCP.

3.2.2. *Pneumocystis jiroveci* DNA in Immunocompetent Health Care Workers in Contact with Patients with *Pneumocystis* Pneumonia.

P. jiroveci DNA was identified in respiratory samples taken from immunocompetent HCW, the majority of whom had occupational contact with PCP patients. *P. jiroveci* DNA is not usually found in non-invasive respiratory tract samples from immunocompetent individuals without PCP (Oz *et al.*, 2000). Serum titres of *P. jiroveci* antibodies from immunocompetent HCW caring for patients with AIDS were raised when compared with the serum titres of *P. jiroveci* antibodies from HCW exclusively treating elderly patients (Leigh *et al.*, 1993). The serum titre data contrasts with data from another study that demonstrated that *P. jiroveci* DNA was not detected in oropharyngeal (OP) samples from immunocompetent HCW treating patients with PCP (Lundgren *et al.*, 1997). However, a potentially less sensitive single rather than nested PCR amplification of the mt LSU rRNA locus was used.

The data presented supports the hypothesis that *P. jiroveci* organisms can be transmitted from a PCP patient to a HCW. Yet, the samples used in this study are unique. They were collected in the early 1990s when prevention of *P. jiroveci* transmission was limited. Since then, concerns about the nosocomial transmission of tuberculosis have resulted in changes in the management of HIV adults at risk of developing PCP. An HIV adult with any respiratory symptoms who is admitted to hospital is cared for in a side room with negative pressure ventilation. Bronchoscopists now wear HEPA filter masks and the procedure is carried out in a room with negative pressure ventilation. These changes would limit the transmission of airborne organisms.

A total of 31 ITS genotypes were detected in this study; 26 were identified in the patient samples and eight were identified in the HCW samples. Three ITS genotypes were identified in both sample groups and five new ITS genotypes were characterised. Few corresponding

ITS genotypes were detected in both the patient and HCW samples, implying that direct transmission between the two groups may not be solely responsible for the presence of *P. jiroveci* organisms in the HCW samples. Five ITS genotypes were detected only in the HCW samples, implying that different ITS genotypes may have adapted to exist in the two different host environments. One set of ITS genotypes may represent *P. jiroveci* organisms that have adapted to survive in PCP patients while another set of ITS genotypes may represent *P. jiroveci* organisms that have adapted to survive only in immunocompetent carriers. It is also possible that different *P. jiroveci* types have adapted to exist in different microniches of the human lung as depicted by different types of respiratory samples. The respiratory samples from the HCW were a mixture of induced sputum samples and nasal rinse samples while all the respiratory samples from the patients were BAL samples. This might explain the lack of correlation between the HCW samples and the patient samples. The hypothesis that ITS genotypes represent different types of *P. jiroveci,* is supported by the suggestion that some ITS genotypes are associated with disease severity, for example, B_2a_1 is often found in mild disease (Miller and Wakefield, 1999a).

For the majority of the ITS genotyping, four recombinant clones per sample were sequenced. This is considered a reasonable number to identify the main genotypes present (Tsolaki *et al.,* 1996). In the mother and infant study, ten ITS recombinant clones were sequenced originating from the sample of the mother, B_1a_3 was identified in eight out of the ten clones while two minority types B_1b_1 and Ca_3 were also detected. It is possible that if four recombinant clones had been sequenced in the mother and infant study that the minority types would not have been detected. In this study, it is possible that if more recombinant clones had been sequenced per sample that they may have been more correlation between the ITS genotypes from the patient and HCW samples.

One HCW maintained carriage of *P. jiroveci* organisms for a period of 27 months. No *P. jiroveci* organisms were detected in the last respiratory sample taken, suggesting that the HCW managed to clear the organisms (Table 3.6). It is possible that the *P. jiroveci* organisms were present before the start of the study and the respiratory samples taken as part of this study are just a representation of continuous acquisition and clearage of *P. jiroveci* organisms within the lungs of this HCW, especially as the ITS genotypes change throughout the 27 months. Reinfection could either be from a PCP patient, an HIV patient colonised with *P. jiroveci* or another HCW with asymptomatic carriage of *P. jiroveci.*

In another study, *P. jiroveci* ITS genotypes were determined in 17 patients in the same hospital, 11 were PCP patients and 6 were HIV-negative patients colonised with *P. jiroveci* organisms (Nevez *et al.,* 2003). Similar to the HCW study, out of a total of 14 ITS genotypes only 3 ITS genotypes were common to both groups indicating that different ITS genotypes are associated with PCP and pulmonary colonisation. Another study looked at asymptomatic carriage in multiple BAL fluid samples from 16 HIV infected patients, collected between 1990 and 2001 (Wakefield *et al.,* 2003). Genotyping was done at the mt LSU rRNA, mt SSU rRNA, DHPS and SOD loci. Seven patients had asymptomatic carriage of *P. jiroveci* over periods of nine months or more, following an episode of PCP. During carriage, in all seven patients, a change in genotype from that in the original episode of PCP was observed. Two patients had asymptomatic carriage for up to nine months or more prior to an episode of PCP with no change in genotype and five patients were asymptomatic carriers for up to two years without developing PCP. This implies that *P. jiroveci* organisms can lie dormant within the lungs for extended periods of time. Asymptomatic carriage in HIV infected patients and colonisation of human populations by *P. jiroveci* may play a role in transmission of *P. jiroveci* and whatever the clinical manifestation, might act as a reservoir for the fungus.

Using the animal model, Dumoulin *et al.,* showed that asymptomatic carriage of *Pneumocystis* occurred in mice (Dumoulin *et al.,* 2000). Immunocompetent BALB/c mice, transiently parasitized by *Pneumocystis* organisms after close contact with *Pneumocystis* infected mice, were able to transmit the infection to *Pneumocystis*-free mice with severe combined immunodeficiency. The minimum amount of contact required for the BALB/c mice to transmit *Pneumocystis* organisms to the SCID mice was only one day. This data shows that there is circulation of infectious *Pneumocystis* organisms between infected, carrier and susceptible mice but the question remains of whether the same is true for *P. jiroveci* and humans.

Studies looking at DHPS mutations within *P. jiroveci* infection have shown a correlation not only with prior sulphur prophylaxis and treatment but also with geographical location (Huang *et al.,* 2000). The DHPS mutations were associated with place of diagnosis rather than place of birth, suggesting reinfection rather than latency of *P. jiroveci* organisms. Also, there was a high proportion of mutant genotypes among persons never prescribed prophylaxis, including those newly diagnosed with HIV infection, implying the mutations originated in another PCP patient. This provides indirect evidence that organisms with these mutations are transmitted from person to person either directly or through a common exogenous source (Beard *et al.,* 2000). Futhermore, geographic clustering of PCP cases has been documented among HIV-infected individuals with the analysis based on zip codes, (Dohn *et al.,* 2000, Morris *et al.,* 2000).

Vargas *et al.,* detected *P. jiroveci* DNA in nasal samples from a child with PCP, his mother and two contact health care workers but not in respiratory samples from 30 hospital staff members who did not enter the room of the patient (controls), (Vargas *et al.,* 2000). The results document that person-to-person transmission of *P. jiroveci* is possible. Like the

present study, it also raises the question as to whether healthy contacts can transmit the disease to immunocompromised patients at risk.

3.2.3. Distribution of *Pneumocystis jiroveci* Types in the Lung of a Child Dying of *Pneumocystis* Pneumonia

A representative distribution of ITS genotypes in the human lung was characterised and comparable to ITS genotypes usually detected in respiratory samples, two previously described ITS genotypes were identified. The ITS locus was used because it is the most informative genotyping locus. It contains more polymorphisms in comparison to other *P. jiroveci* genotyping loci (Tsolaki *et al.,* 1996) and it is present as a single copy gene (Nahimana *et al.,* 2000). The lung contained a mixed infection of two different ITS genotypes, Ca_3 and B_1b_2, corresponding to Fg and Ec, respectively (Lee *et al.,* 1998). This conclusively shows, by using human lung tissue, that infection with *P. jiroveci* is not always clonal and infection with more than one type of *P. jiroveci* is possible. The lower lobe consisted mainly of type Ca_3 while the middle lobe consisted mainly of type B_1b_2. The upper lobe was a mixture of the two types.

The data suggest that the infection was initiated by two separate foci of organisms; type Ca_3 in the lower lobe and type B_1b_2 in the middle lobe. The two sites of infection spread and met in the upper lobe where they proliferated. This idea is consistent with the upper lobe having a heavier burden of infection (Grebski *et al.,* 1994, Baughman *et al.,* 1993). Competition between two different types of *P. jiroveci* organisms may have lead to different *P. jiroveci* organisms adapting to survive in different environmental microniches within the lung.

A similar study was carried out using respiratory samples to investigate the distribution of *P. jiroveci* organisms within the lung (Keely *et al.,* 1996). From each of the ten patients, two

BAL samples were taken during one episode of PCP. One was taken from the right upper lobe and the other from the right middle lobe and both were genotyped at the mt LSU rRNA locus. In every case, the genotype was the same regardless of the site of origin. However, the mt LSU rRNA locus is not as informative as the ITS locus as only four different polymorphic phenotypes have been identified (Beard *et al.,* 2000) as opposed to 21 ITS1 genotypes and 29 ITS2 genotypes (Nimri *et al.,* 2002).

A Danish study looked at the distribution of *P. jiroveci* organisms in three different whole *P. jiroveci* infected lung tissues (Helweg-Larsen *et al.,* 2001a). Three loci were used for genotyping, ITS, mt LSU rRNA and DHPS. Two of the three lungs contained a mixed infection. Not all the genotypes identified in the lung tissue were detected in the relevant BAL respiratory sample. This conclusion has implications for reinfection studies on recurrent PCP, which rely on respiratory samples such as induced sputum or BAL, because the genotypes detected in the respiratory samples may not represent all the genotypes present in the lung. The Danish study also recorded different genotypes being detected in different areas of the lung.

Animal studies have limited use for examining the distribution of *Pneumocystis* types because there is a dearth of appropriate markers in non-human-derived *Pneumocystis*. In the rat, there is no variation within the rat-derived *Pneumocystis* ITS locus (Palmer *et al.,* 2000b). By contrast, up to five different special forms of rat-derived *Pneumocystis* have been identified by karyotype analysis (Palmer *et al.,* 2000b) and co-infection with more than one type is common. The potential dynamic interaction of two special forms of rat-derived *Pneumocystis* might be equivalent to the co-infection with two different types of *P. jiroveci* identified at the ITS locus. However, the human-derived and rat-derived *Pneumocystis*

organisms are very different, not only in genetic composition but also in their population structure and natural history of infection (Wakefield, 1998).

In humans, different ITS genotypes of *P. jiroveci* have been found to reflect different levels of disease severity (Miller and Wakefield, 1999a). For example, B_2a_1 is associated with mild disease while B_1a_3 was found only in first and not in subsequent episodes of PCP. It is intriguing that one of the ITS genotypes identified in this study, Ca_3 is uncommon in *P. jiroveci* isolates from the UK (4%) and from samples from Denmark and the United States (1%). Furthermore, it is of note that the human lung tissue came from a fatal PCP case. Unlike the majority of PCP samples, the lung tissue was from a Chilean child, immunosuppressed by cytotoxic drugs rather than by HIV infection, yet, *P. jiroveci* ITS genotypes equivalent to those from Northern Hemisphere HIV adults were identified.

3.2.4. Overall Discussion

To summarise, a mother and her 4½-week-old infant had PCP contemporaneously. Genotyping of *P. jiroveci* DNA at three independent loci detected the same genotype in the sample from the infant as the majority genotype present in the sample from the mother. These data are consistent with direct airborne transmission of *P. jiroveci* organisms from the mother to the infant but other possible sources cannot be ignored.

The second study detected *P. jiroveci* DNA in respiratory samples from immunocompetent HCW. *P. jiroveci* DNA carriage is asymptomatic and tends to occur in HCW with the closest, most direct contact with PCP patients; in this study, it was the bronchoscopists. The genotyping results show little evidence of direct transmission between HCW and the corresponding patient group, reflecting the complexity of the transmission of *P. jiroveci*. There are many likely routes of transmission of *P. jiroveci*; from an immunocompromised

PCP patient to either a susceptible immunocompromised host or to an immunocompetent close contact. Alternatively, *P. jiroveci* could be transmitted from asymptomatic carriers to immunocompromised patients or an exogenous source of *P. jiroveci* might exist.

The third study characterised a representative distribution of *P. jiroveci* within a whole human lung. The number of genotypes detected was comparable to the number of genotypes detected in respiratory samples. Two different ITS genotypes were identified which may represent two different forms of the organism that could exist in different environments in the lung. The data also directly demonstrates that infection of the human lung is not necessarily clonal and the same ITS genotypes are present in young Chilean immunosuppressed child as in adult HIV sufferers from the Western World.

The three genotyping loci have different levels of utility. The mt LSU rRNA locus is multicopy so is sensitive to PCR amplification, however, it is not very polymorphic as only four different genotypes have been defined. In contrast, the ITS locus is single copy so is not as sensitive as the mt LSU rRNA locus to PCR amplification, however, it is highly polymorphic so is very discriminatory as many different genotypes have been defined. There is evidence though that the polymorphisms may be unstable over time (Nimri *et al.,* 2002). The DHPS locus is both single copy and only has two polymorphic nucleotides so is neither discriminatory nor sensitive to PCR amplification.

The *MSG* gene has been proposed as a possible epidemiological marker. Two studies have developed assays using PCR amplification of *MSG* genes to aid diagnosis of *P. jiroveci* (Huang *et al.,* 1999a, Mei *et al.,* 1998), while another study has analysed the variation of tandem repeats in the intron of the UCS locus of *P. jiroveci* (Ma *et al.,* 2002). Quantifying the number of repeats is a simple and rapid-typing method. It is possible that *PRT1* genes

could also be used as an epidemiological marker. All the loci described can be used alone or in combination with other methods to study the epidemiology of *P. jiroveci.*

Multilocus genotyping is the use of combinations of the different genotyping loci to characterise an infection. When describing differences between isolates of an organism, it is a more powerful tool than genotyping at a single locus (Feavers *et al.,* 1999, Maiden *et al.,* 1998). The chief advantage of multilocus typing is that there is a greater sensitivity for detecting differences when several loci are examined concurrently rather than when a single locus alone is examined. Additionally, by looking at polymorphisms at multiple gene loci, more information is available for strain characterisation so that strain typing has been performed rather than simply gene typing.

All molecular typing methods are limited by the fact that it is only possible to show which type of DNA is present; there is no information on the viability and infectious nature of the organisms. Several studies have attempted to develop a viability assay. In the rat model, microscopy experiments have demonstrated that rats with *Pneumocystis* infection release specific life cycle forms of *Pneumocystis* organisms that can remain dormant at room temperature for at least five months outside the mammalian host (Chin *et al.,* 1999). One sample was used to initiate a *Pneumocystis* culture, biological proof that the organisms remain viable.

In the rat model, researchers have tried to design a molecular viability assay (Maher *et al.,* 2000). The method is based upon the detection of rat-derived *Pneumocystis* mRNA by an RT-PCR assay that employs specific primers to amplify a member of the heat shock protein 70, HSP70, family. The RT-PCR assay was capable of detecting as few as 10^2 viable trophozoites as determined by ethidium bromide staining. In contrast, no HSP70 mRNA was

detected in the RT-PCR assay from 10^6 nonviable trophozoites killed by high heat, desiccation or UV exposure. However, the published data using this method was flawed by lacking the correct controls. The same assay has been successfully used for examining *P. jiroveci* organisms (Latouche *et al.*, 2001) but it has yet to prove biologically that the presence of HSP70 mRNA in a preparation of rat-derived *Pneumocystis* organisms correlates with the preparation being viable or infectious. To be shown to be viable, a biological endpoint is necessary.

There are many different aspects to the possible life cycle of *P. jiroveci* organisms (Dei-Cas, 2000). It is possible that *P. jiroveci* organisms circulate asymptomatically in human populations, being able to infect hosts with diverse susceptibility levels. Patients with PCP are the only part of the transmission cycle that is presented clinically. Asymptomatic carriage of *P. jiroveci* has now been identified in immunocompetent HCW and susceptible immunocompromised hosts. However, an exogenous source has yet to be characterised but *Pneumocystis* DNA was detected in the English countryside (Wakefield, 1996). Also, when primary infection occurs in children, it is possible that they can act as a reservoir of *P. jiroveci* organisms (Vargas *et al.*, 2001).

The three studies have added data to understand more about the transmission and acquisition of *P. jiroveci*. Direct transmission has been shown and asymptomatic carriage seems likely. A representative distribution of *P. jiroveci* has been characterised within human lung tissue. Epidemiology studies on *P. jiroveci* will always be limited by the lack of an effective culture system, which could produce large numbers of organisms or allow the derivation of clonally derived stocks. Samples need to be grown in a controlled environment where they can be studied and analysed further and be demonstrated to be viable and infectious.

4. Analysis of Telomeric and Non-telomeric Copies of *PRT1*.

PRT1 is a multicopy gene that is predominantly subtelomeric and encodes a surface expressed protease. It is most closely related to the single copy kexin genes (Lugli *et al.*, 1997), which are located in the internal regions of chromosomes. A hypothesis has been proposed that *PRT1* was originally a single copy kexin gene (Kutty and Kovacs, 2003), located in the internal chromosomal regions, that was duplicated and the new copy migrated to the subtelomeric regions. Here, multiple replications occurred and the multicopy gene family arose. The aim of this chapter is to identify the original rat-derived *Pneumocystis* single copy *PRT1* gene that encodes the progenitor kexin-like protease because if the progenitor kexin-like protease did exist and was constitutively expressed it would be likely to complicate data from other protease studies.

4.1.1. Southern Hybridisations to Identify Non-telomeric Copies of *PRT1*

The first step was to identify copies of *PRT1* that are located in the internal regions of chromosomes. Genomic DNA extracted from rat-derived *Pneumocystis* population 876.4 was digested with BAL-31 and aliquots were taken at set times to monitor the progress of the digest (section 2.10). The genomic DNA was digested with the restriction enzyme *Pst*1. *Pst*1 should cut predominantly in the internal regions of the chromosomes because there are fewer restriction sites at the telomeres because of multiple repeat regions (Underwood *et al.*, 1996). Thus, the majority of larger fragments from the digest contain the subtelomeres while most of the smaller fragments contain internal coding sequence (Sunkin and Stringer, 1996). The resulting DNA was electrophoretically separated on an agarose gel (Figure 4.1a). The rat derived *Pneumocystis* genomic DNA was fully digested by *Pst*1 and was highly concentrated, except for the sample where the BAL-31 digest had been left for 15 minutes,

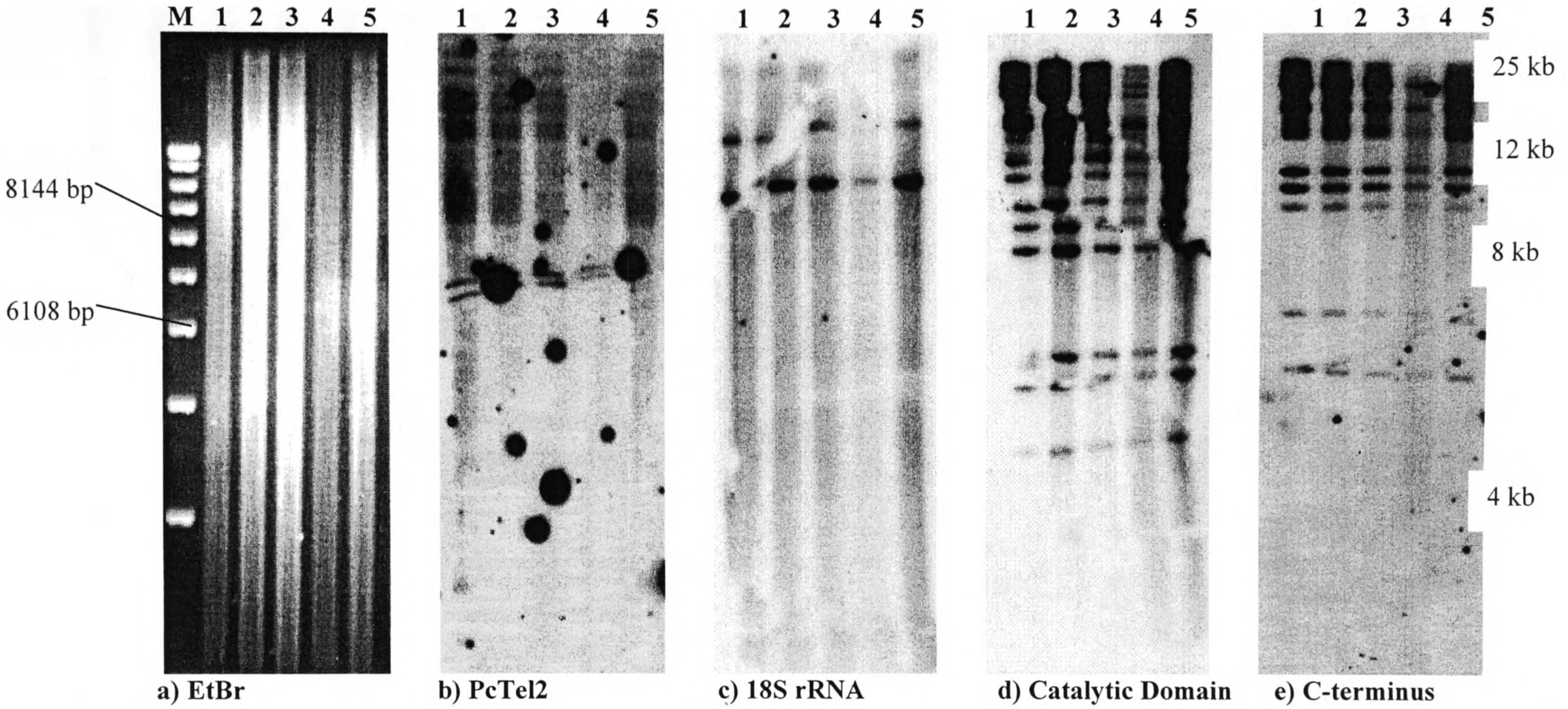


Figure 4.1. Southern hybridisation analysis of the sensitivity of rat-derived *Pneumocystis* genomic DNA to BAL-31 digestion. Rat-derived *Pneumocystis* genomic DNA was extracted from population 876.4, M, 1 kb DNA ladder. lane 1: DNA not treated with BAL-31; lanes 2-5: DNA treated with BAL-31 for 5, 10, 15 and 20 minutes respectively. Panel a: ethidium bromide stained agarose gel; panel b: the DNA has been hybridised with the oligonucleotide PcTel2 which binds to rat-derived *Pneumocystis* subtelomeric repeats; panel c: the DNA was been hybridised to a double stranded probe specific to rat-derived *Pneumocystis* 18S rRNA; panels d and e: the DNA was been hybridised to a double stranded probe specific to the catalytic domain and C-terminus domain of the rat-derived *Pneumocystis* multicopy gene *PRT1*, respectively.

here, the DNA was slightly fainter. The gel was Southern blotted and the resulting membrane was hybridised four times with different probes and was stripped between each hybridisation.

The oligonucleotide PcTel2 binds to subtelomeric repeats and therefore was used to monitor the progress of the BAL-31 digest. Figure 4.1(b) shows the results for this hybridisation. The uppermost bands in the 25 kb region appeared to be degraded as the BAL-31 digest progressed. However, there was no striking diminishment, implying that the digest was not fully successful.

A double stranded DNA probe, specific for the rat-derived *Pneumocystis* 18S rRNA locus was generated by amplifying genomic DNA extracted from a rat-derived *Pneumocystis* population, 901.3, with the primers 18S rRNA3 and 18S rRNA4. This probe was used to monitor the overall efficiency of the DNA transfer and hybridisation, see Figure 4.1(c). The 18S rRNA locus was unaffected by BAL-31 digestion as it is located on an internal region of the chromosomes. Two clear strong bands were obtained for every sample (except the 15 minute aliquot), as there is a *Pst*1 site within the 18S rRNA locus.

Two double stranded DNA probes were made to monitor the position and number of the *PRT1* copies. The first was specific to the catalytic domain of the *PRT1* gene, the catalytic probe, and was made by PCR amplification of the *PRT1* genomic clone pAGA21 with the primers PcProt1 and PcProt3. As seen in Figure 4.1(d), a large number of bands were detected in the upper part of the blot, ranging from 8 kb to 25 kb in size. This region of the blot is assumed to correspond to subtelomeric fragments. Three bands were identified in the region of the blot where the fragments ranged from 4 kb in size to 8 kb. As previously described, these fragments were assumed to be internal regions of the chromosomes because

*Pst*1 will cut predominantly in the internal regions of the chromosomes therefore most of the smaller fragments contain internal coding sequence (Sunkin and Stringer, 1996). Hence, it is likely that not all *PRT1* copies are subtelomeric.

The second double stranded DNA probe used to monitor the position and number of the *PRT1* gene copies was specific to the C-terminus domain of *PRT1*, the C-terminus probe. It was made by PCR amplification of rat-derived *Pneumocystis* genomic DNA extracted from population 876.4 with the primers PcProt48 and PcProt57. A result similar to that obtained with the catalytic domain probe was observed, as shown in Figure 4.1(e) but the band pattern was slightly different to the catalytic probe in the 8 kb to 25 kb region. The C-terminus probe hybridised to two bands in the 4 kb to 8 kb region of the gel, where smaller DNA fragments are located and which may correspond to internal regions of the chromosomes. This blot also indicated that not all copies of *PRT1* genes are subtelomeric.

4.1.2. Further Investigation of Non-telomeric Copies of *PRT1*

The copies of *PRT1* genes from the hybridisation experiments that were indicated to be located on the internal regions of the chromosomes were investigated further. The catalytic domain of PRT1 is the part of the protein where enzymatic activity occurs. If the *PRT1* multigene family has evolved from a single copy *PRT1* gene that encodes a progenitor kexin-like protease, it is the catalytic domain that is most likely to have been conserved because of the constraints on its function. Hence, it could be hypothesised that the catalytic probe will bind to the progenitor single copy *PRT1* gene. Kexins (Redding *et al.*, 1991) and the mouse-derived and human-derived *Pneumocystis* single copy kexin-like protein Kex1 (Kutty and Kovacs, 2003, Lee *et al.*, 2000) bind to the endoplasmic reticulum, or the Golgi apparatus, via a transmembrane domain. It is thought likely that a rat-derived *Pneumocystis* progenitor kexin-like protease would interact in the same way. PRT1 is surface expressed via

a GPI anchor (Palmer and Wakefield, 2001). Therefore, the C-terminus of the progenitor kexin-like protease may have evolved to accommodate the potential change from a transmembrane domain to a C-terminus domain, which contains a GPI anchor binding site. Hence, it could be hypothesised that the C-terminus probe would not hybridise to the progenitor single copy *PRT1* gene. This could explain the discrepancy of the catalytic probe binding to three copies of *PRT1* genes identified in the internal regions of the chromosomes while the C-terminus probe only binds to two copies of *PRT1* genes in this region. The *PRT1* copy to which the catalytic probe hybridises (but not the C-terminus probe) could potentially encode the progenitor kexin-like protease.

Two samples of *Pst*1 digested rat-derived *Pneumocystis* genomic DNA extracted from population 876.4 were run on an agarose gel. For both samples, an area of the gel containing DNA of a size, 6-7 kb, was excised. This is the area of the gel that corresponds to the size of the DNA fragment that might contain a potential single copy *PRT1* gene (Figure 4.2). The excised agarose block from the first sample was subdivided into five pieces, 1a-e, while the excised agarose block from the second sample was subdivided into three pieces, 2a-c. To pinpoint the fragment of DNA containing the potential single copy *PRT1* gene all eight pieces of agarose were used as templates in a PCR amplification with the universal primers PcProt1 and PcProt3, which are specific to the catalytic domain (Figure 4.3). For a positive control, a piece of agarose was also excised from the part of the gel containing many subtelomeric copies of *PRT1* genes and for a negative control a piece of agarose was excised from part of the gel where there were no copies of *PRT1* genes. PCR amplification was successful for pieces of agarose, 1c, 1e, 2a and 2c. To verify that the eight pieces of agarose did not contain a C-terminus end encoding a GPI anchor site, another PCR amplification was performed using the primers PcProt48 and PcProt57, which are specific to the C-terminus domain. Only the positive control generated a PCR product.

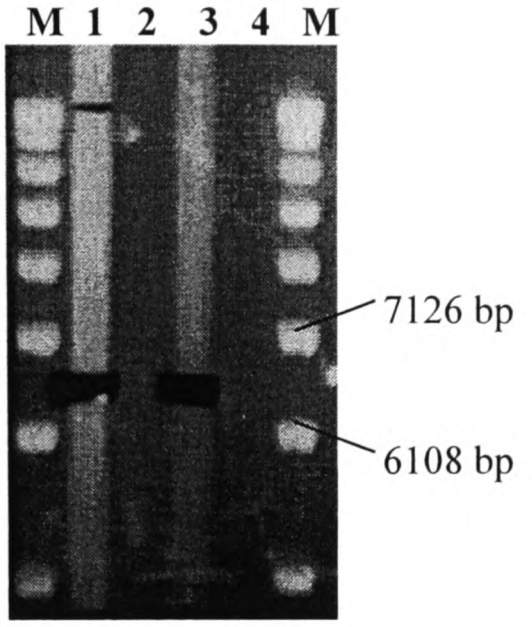


Figure 4.2. Agarose gel of *Pst*1 digested rat-derived *Pneumocystis* genomic DNA M, 1 kb DNA ladder. Lanes 1 and 3: *Pst*1 digested rat-derived *Pneumocystis* genomic DNA extracted from population 876.4; lanes 2 and 4: negative control using dH_2O. Genomic DNA was accurately excised from the 6-7 kb region, lanes 1 and 3.

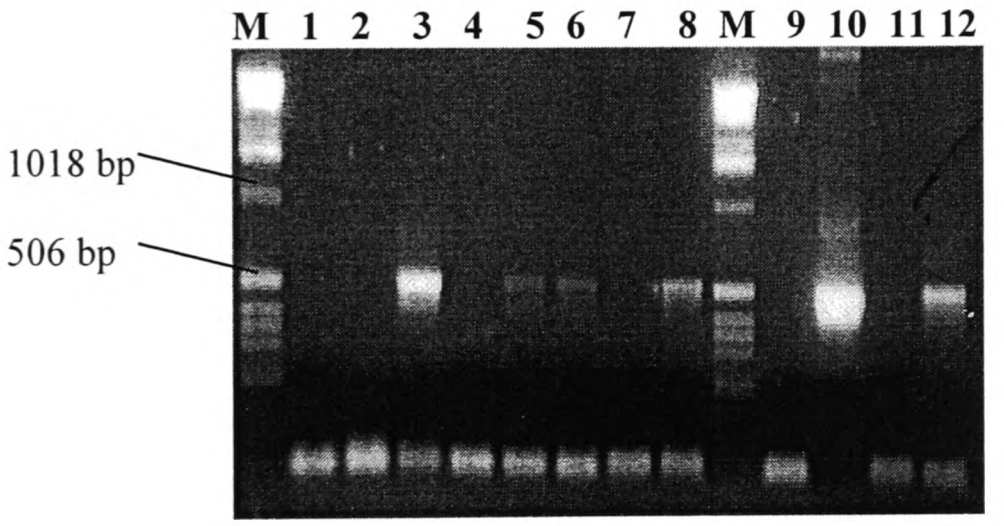


Figure 4.3. Electrophoretic separation of DNA amplification products using the subdivided excised agarose blocks as templates Each of the eight fragments from the two excised blocks was used as a template for the PCR amplification with the primers PcProt1 and PcProt3. M, 1 kb DNA ladder. Lanes 1-5: 5 fragments from block 1, 1a-1e; lanes 6-8: 3 fragments from block 2, 2a-2c; lane 9: a piece of agarose that should not contain a protease gene used as a negative control; lane 10: a piece of agarose that should contain a protease gene used as a positive control; lane 11: negative control, using dH_2O in place of a template; lane 12: positive control using pAGA21 as a template

Two of the four catalytic domain PCR products were analysed further. The two PCR products obtained from the agarose 1e and 2a were cloned without prior purification and for each PCR product, the inserts from three plasmids were sequenced. One representative catalytic domain sequence from each agarose piece with a potential single copy *PRT1* protease was compared with the ten published catalytic domain sequences. Figure 4.4 shows the ClustalW alignment of the translated sequences. All of the amino acid residues that are conserved throughout the kexin family were present in the two potential single copy *PRT1* cat domain sequences, 1c71 and 2a71, including the asparagine at residue 45 which forms part of the oxyanion hole and serine at residue 116 which is part of the catalytic triad. There is one exception in sequence 2a71, which has an alanine rather than a glycine at residue 124. This difference is unlikely to have huge structural implications.

Two motifs have been highlighted in Figure 4.4. The first is the KRGAY or GENGS motif. This motif is striking because there are two very different alternative sequences. The first is the basic KRGAY sequence which contains a potential autocatalytic site while the second is the more acidic option GENGS which contains a negatively charged residue, glutamate. Both 1c71 and 2a71 follow the more acidic trend with the sequences GENGS and GESGR respectively. The negatively charged glutamate would affect the structure of the catalytic domain. However, this is a common feature in the *PRT1* gene family, so is not something unique indicating a potential single copy *PRT1* gene.

The second motif, PLAAG is less diverse. The 1c71 sequence, AITAG, is unique by having a threonine residue in the middle. Threonine contains a hydroxyl group as opposed to alanine, which contains a methyl group. Therefore, it has the potential to form a hydrogen bond in the catalytic domain, which could alter the structure. The 2a71 sequence, PIASA, is

```
          1                                                                  ✸                                  70
   orf3   SCSWGPADTG NLTQDIFYTT YSAIIKGINQ GRNGLGSIYV FGSGNGGYFD NCNYDGYANS PYTITIAAID
  paga3   SCSWGPADTG NLTQDIFYTT YSAIIKGINQ GRNGLGSIYV FGSGNGGYFD NCNYDGYANS PYTITIAAID
   1b22   SCSWGPADTG NLTQDIFYTT YSAIIKGINQ GRNGLGSIYV FGSGNGGYFD NCNYDGYANS PYTITIAAID
 11a112   SCSWGPPEYG SLAMPMMPTT YSAIIKGVKE GRNGLGSIYV FGTGNGGLLD DCNYDGYANS PYTITIGAIN
   1b21   SCSWGPPDYG YVAMPMMPTT YSAIIKGIKE GRNGLGSIYV FGSGNGGLLD DCNYDGYANS PYTVTIAAID
    3g5   SCSWGPADTG NLIQDISYTS YSAIIKGINE GRNGLGSIYV FGSGNGGYFD DCNYDGYANS PYTITIAAID
 u82999   SCSWGPADDG NSIEGISYST YSSIIRGINE GRNGLGSIYV FGSGNGGYYD DCNYDGYANS PYTITIAAID
    73j   SCSWGPPGDG YAAIPMYPTT YSAIIKGIKE GRNGLGSIYV FGTGNGGSLD GCNYDGYANS PYTITIAAID
 u62910   SCSWGPADTG NLIQDISYTS YSAIIKGINE GKNGLGSIYV FGSGNGGYFD DCNYDGYANS PYTITIAAID
   1c71   SCSWGPPDYG YVAMPMMPTT YSAIIKGIKE GRNGLGSIYV FGSGNGGLLD DCNYDGYANS PYTVTIAAID
   2a71   SCSWGPSDLG NKIENIFYST YTAIIDGINK GRKGLGSIYV FGSGNGGDLD HCNYDGYANS PYTITIAAIN
 11a111   .......... .........T YSAVIKGIKE GRNGLGSIYV FGSGNGGLLD DCNYDGYANS PYTVTIGAID
            β4 | Hypervariable  |<── αE ──>|   |<── β5 ──>|                    |<── β6

                                                              ✸                                  140
   orf3   AEEKRFIFSE PCPCILASTY SGKRGAYIYT TDVGTTECSI RHTGSSASTP LAAGVIALLL SACPNLTWRX
  paga3   AEEKRFIFSE PCPCILASTY SGKRGAYIYT TDVGTTECSI RHTGSSASTP LAAGVIALLL SACPNLTWRX
   1b22   AEEKRFIFSE PCPCILASTY SGKRGAYIYT TDVGTTECSI RHTGSSASTP LAAGVIALLL SACPNLTWRX
 11a112   SEDERPYFSE SCPCILASTY SGGGNGSIYT TDIGKTNCTT QHSGSSASTA IAAGIIALVL SANPNLTWRX
   1b21   AEDKKFDTSE SCPCILASTY SGGENGSIYT TDLGKTNCTT QHSGTSASTA IAAGIIALVL SVNPNLTWRX
    3g5   AEEKRFRFSE PCPCILASTY SGKSGAYIYT TDVGTTECTF LHTGTSASTP LAAGVIALLL SACPNLTWRX
 u82999   AEEKRFIFSE PCPCVLASTY SGGSNGRIYT TDVGTTECTS RHTGTSASTP IAAGIVALVL SVRPDLTWRX
    73j   SEDKNFYFSE SCPCILASTY SGGENGSIYT TDLGKEGCTT EHTGASASTP LAAGIIALVL SANPNLTWHX
 u62910   AKEERFMFSE PCPCILASTY SGRRGAYIYT TDVGTTECTF LHTGTSASTP LAAGVIALLL SACPNLTWRX
   1c71   AEDKKFDTSE SCPCILASTY SGGENGSIYT TDLGKTNCTT RHSGTSASTA ITAGIIALVL SVNPNLTWRX
   2a71   ARERKFRFSE PCPCILASTY SGGESGRIYT TDVGSTACTS IHTGISAAIP IASAIIALVL SVNPNLTWRX
 11a111   SEDKNFYFSE SCPCILASTY SGGENESIYT TDIGKTNCTT EHSGTSASTA IAAGIIALVL SVNPNLTWHX
      <──>|          |  β7  |                                  |<── αE ──>|
```

Figure 4.4 Amino acid alignment of the catalytic domain of different copies of published *PRT1* sequences compared with the catalytic domains of two representative potential progenitor single copy *PRT1* gene sequences. The ClustalW alignment consists of the catalytic domain of the 10 published *PRT1* sequences, and the two catalytic domain sequences of the two potential single copy *PRT1* gene sequences, 1c71 and 2a71. The secondary structure is marked in blue at the bottom; β - B-sheet and α - α-helix, (Lugli *et al.*, 1997). The residues marked in purple are conserved residues. N45 forms the oxyanion hole and S116 is part of the catalytic triad, both are marked ✸. Two conserved motifs are indicated in blue and magenta.

unique by having an alanine residue rather than a glycine residue at position five. As discussed previously this does not have huge structural implications.

There is little convincing sequence data that either of the two sequences correspond to the progenitor single copy *PRT1* gene. A phylogenetic tree was constructed via the neighbour joining method for all twelve sequences (data not shown) and demonstrated that 1c71 is most closely related to 1b21, as both sequences are part of a clade at the bottom of the tree. There is little to suggest that 1c71 is a unique sequence. 2c71, however, is apart from the main structure of the tree and is not closely related to any sequence. This implies that it may be unique, supporting the argument that it could be the progenitor single copy *PRT1* gene.

To further investigate the progenitor single copy *PRT1* gene, more PCR amplifications were performed on the excised pieces of agarose, which could potentially contain the relevant DNA fragments containing the progenitor single copy *PRT1* gene. PCR amplification using the primers specific to the 5' domain of the *PRT1* genes, PcProt31 and PcProt32 was successful. Again agarose fragments 1c and 2a generated PCR products, which were cloned without prior purification, and four inserts were successfully sequenced from six resulting plasmids.

All four inserts produced the same sequence, so a representative sequence, h22, from the cloning of the 5' domain of the potential single copy *PRT1* gene was compared with the 5' domain sequences of the nine published *PRT1* sequences (the sequence for U82999 did not extend as far as the 5' domain). Figure 4.5 shows the ClustalW alignment for the translated sequences. The whole sequence is very conserved. The sequence, h22, has two unique residues, an arginine at position 16 and a glutamate at position 80. The other residues at position 16 are either histidines or tyrosines. All three amino acids, arginine, histidine and

```
          1                                                    40
11a112    DFSRALGFKH YMKLEYLDNQ HIFSIEKSVL EDEIKEKIEN
11a111    DFSRALGFKH YMKLEYLDNQ HIFSIEKSVL EDEIKEKIEN
u62910    DFSRALGFKH YMKLEYLDNQ HIFSIEKSVL EDEIKEKIEN
   3g5    DFSRALGFKY HEKLEYLDNQ HIFSIEKGVL EDEIKEKIEN
  orf3    DFSRALGFKY HEKLEYLDNQ HIFSIEKGVL EDEIKEKIEN
  1b22    EFSRAVGLKY HMKVEYLDNQ HIFFIEKGVL EDEIKEKIEN
 paga3    EFSRAVGLKY HMKVEYLDNQ HIFFIEKGVL EDEIKEKIEN
  1b21    EFSRAIGFKY HMKVDHLDNH HIFFIEKGVL EDEIKEKIEN
   h22    EFSRAIGFKY HMKVDRLDNH HIFFIEKGVL EDEIKEKIEN
   73j    EFSRAVGFKY HMKVDHLDNH HIFFIEKGVL EDEIKEKIEN


                                                        80
11a112    YFGLEKGRNA IDGFNSDKLF HYEKQKLVKR EKRNVIKDDI
11a111    YFGLEKGRNA IDGFNSDKLF HYEKQKLVKR EKRNVIKDDI
u62910    YFGLEKGRNA IDGFNSDKLF YYEKQKLVKR ENRGAIRDDI
   3g5    YFGLEKGRNA IDGFNSDKLF YYEKQKLVKR ENRGAIRDDI
  orf3    YFGLEKGRNA INGFNSDKLF YYEKQKLVKR ENRGVIRDDI
  1b22    YFGLEKGRNA IDGFNSDKLF YYEKQKLVKR VNRGVIRDDI
 paga3    YFGLEKGRNA IDGFNSDKLF YYEKQKLVKR VNRGVIRDDI
  1b21    YFGLEKGRNA INGFNSDKLF YYEKQKLVKQ ENRGAIRDDI
   h22    YFGLEKGRNA INGFNSDKLF YYEKQKLVKP ENRGAIRDDE
   73j    YFSLEKGRNA IDGFNSDKLF YYEKQKLVKP VNRGAIRDDI
```

Figure 4.5. Amino acid alignment of the 5' domain of different copies of published *PRT1* sequences compared with the 5' domain of a representative potential progenitor single copy *PRT1* gene sequence. The ClustalW alignment consists of the 5' domain sequences of 9 published *PRT1* sequences and the one representative sequence of the 5' domain of a potential single copy *PRT1* gene sequence, h22. Two residues unique to the h22 allele are coloured red and the potential autocatalytic site is highlighted in magenta.

tyrosine have strikingly different structures. So h22 would have a slightly different overall structure as a result of having arginine at position 16.

There is a potential autocatalytic site at positions 79 and 80. The majority of sequences are KR, lysine arginine while h22 and 73j are KP, lysine proline. It is somewhat unusual that the representative sequence, h22 follows the pattern of one of the cDNA clones, implying it too may be expressed. Also, KP does not consist of two basic residues unlike KR so may not be autocatalytic, implying that the 5' leader sequence of the protein is not cleaved.

A phylogenetic tree was constructed via the neighbour-joining method for all eleven sequences (data not shown). Like the catalytic domain sequence 1c71, h22 is most related to 1b21. 1b21 is from a genomic subtelomeric cosmid clone, negating any ideas it might be the progenitor single copy *PRT1* gene. Being in a clade at the bottom of the tree, h22 is also related to other copies of *PRT1*.

4.2. Discussion

In summary, copies of *PRT1,* located on internal regions of the chromosomes, were identified by Southern hybridisation. One hybridisation band, specific to the catalytic probe and not the C-terminus probe, was investigated further and sequence data was generated. One catalytic domain sequence 2a71 was phylogenetically unique but the sequence data results did not identify conclusively as to whether this was the progenitor single copy *PRT1* gene.

A slight reduction in the intensity of the upper bands was seen in the hybridisation with PcTel2. To fully digest the telomeric regions of the rat-derived *Pneumocystis* genomic DNA

the digestion with BAL-31 could have been run for a longer period of time. Also, the optimal digestion time for BAL-31 could have been characterised further.

The catalytic and 5' domain sequence data do not necessarily originate from the same *PRT1* copy because there may be more than one copy of a *PRT1* gene within the piece of agarose. Two different *PRT1* copies would also explain the presence of the two different catalytic domain sequences that were characterised. From the sequence data (Figure 4.4 and Figure 4.5) the 5' domain appeared to be more conserved than the catalytic domain. It was presumed that the catalytic domain was the most crucial to the function of the protein; therefore, it was thought to be the most conserved domain. This may not be true. The 5' domain might be involved in transporting PRT1 to the surface of the organism and could be under more functional constraint. Alternatively, the catalytic domain might be more exposed on the surface of the organism so needs to vary more to evade the immune system of the host. This is discussed in more detail in chapter 6.

Sequence data was generated that may have been from the progenitor single copy *PRT1* gene. Unique residues were identified from candidate sequences and 2c71 stood alone on a phylogenetic tree. However, there was little conclusive evidence that the progenitor single copy *PRT1* gene was identified. More *PRT1* sequence data needs to be generated to try to detect a domain or motif, which has not been previously characterised, that may encode part of a protein with a function not associated with a surface expressed protease. Probes could be designed from this sequence data to investigate the genomic context of this *PRT1* copy, for example, is it single copy and located internally on the chromosomes. Alternatively, the internal protease may be described when the *Pneumocystis* genome sequencing project is completed, http://gene.genetics.uga.ehu.

Two *PRT1* gene copies have been identified that do not possess a GPI anchor, instead they have a carboxy terminus with a hydrophobic region followed by a hydrophilic region, consistent with a transmembrane and intracytoplasmic region, respectively (Russian *et al.*, 1999). This type of domain is thought to be important for localisation of proteins in the endoplasmic reticulum or the Golgi apparatus, as this is a structural feature found in the kexins of several fungi, for example kex2 in *Saccaromyces cerevisiae* (Brenner *et al.*, 1994) and in the mammalian furin PC7 (Nakayama, 1997). Therefore, the genomic context of these sequences needs to be further characterised, to investigate if these *PRT1* clones could be the progenitor single copy *PRT1* gene.

In mouse-derived *Pneumocystis,* a single copy kexin-like protease, KEX1, has been identified (Lee *et al.*, 2000). This also contains a hydrophobic transmembrane spanning domain and a carboxy-terminal cytoplasmic domain with a conserved tyrosine motif thought to be important for localisation of the protease in the endoplasmic reticulum and/ or Golgi apparatus. This protein was identified by expression screening with two specific monoclonal antibodies, mAbs, of a cDNA library generated from a mouse lung infected with *Pneumocystis*. The mAbs were initially selected based on their ability to recognise antigens other than mouse derived *Pneumocystis* MSG, and to cross-react with *Pneumocystis* from different hosts, including ferrets, humans, rats and Rhesus macaques. Such mAbs that are cross-reactive with *Pneumocystis* derived from different hosts are rarely isolated (Gigliotti *et al.*, 1998). The rat-derived *Pneumocystis* antigen, to which these mAbs cross-hybridise to might be the progenitor kexin-like protease, assuming the rat progenitor kexin-like protease is analogous to the mouse KEX1. However, probes specific to the catalytic domain, the P-domain and the C-terminus domain of the gene encoding the mouse KEX1 protein were hybridised against a Southern Blot of rat-derived *Pneumocystis* chromosomes, which had been separated by PFGE. The probe hybridised to many different chromosomes, indicating

that the gene encoding the mouse KEX1 has many analogous genes in the rat-derived *Pneumocystis* genome. Notably, a region on *P. carinii PRT1* homologous to the mouse *KEX1* C-terminus probe has not yet been reported.

A similar single copy *kex1* gene has also been identified in *P. jiroveci* (Kutty and Kovacs, 2003). The complete genomic sequence was obtained by using primers based on the previously reported *P. jiroveci PRT1* gene fragment (Lugli *et al.*, 1999b). *P. jiroveci* genomic DNA was extracted from a PCP infected human lung and a combination of techniques including a two-step PCR with magnetic beads, inverse PCR and genomic library screening were used to identify the *P. jiroveci kex1* sequence. These techniques could be applied to the fragments of DNA obtained in this study to try to obtain their full length gene sequence.

To conclude copies of *PRT1* genes have been identified on the internal regions of chromosomes, demonstrating that not all copies of *PRT1* genes are subtelomeric. This supports the idea that they may be a progenitor single copy *PRT1* gene. Further hybridisation and sequence data has been generated to try to characterise the progenitor single copy *PRT1* gene however, not enough data has been obtained to conclude that the non-telomeric *PRT1* copy was the progenitor kexin-like protease.

5. Characterisation of the Complex Pattern and the Possible Regulation of *PRT1* Gene Expression within the Context of Different Populations of Rat-derived *Pneumocystis*

The aim of this part of the study was to investigate the diversity and number of genomic and expressed copies of *PRT1* and to determine whether *PRT1* gene expression is regulated. An assay was devised which compared cDNA and genomic *PRT1* copies in different populations of rat-derived *Pneumocystis* by analysing the length polymorphisms within the proline rich region (PRR), of *PRT1* genes. The lengths of the PRR sequences divide into two groups, 500-600 bp and 800-1000 bp. The PRR assay compares approximately 30 *PRT1* loci per *Pneumocystis* organism. If the PRR is amplified by PCR, using rat-derived *Pneumocystis* genomic DNA, then this should result in a profile of PCR products comprised of a range PRRs, each with different lengths. A unique profile of cDNA and genomic PRRs should be amplifiable from each rat-derived *Pneumocystis* organism or population, which can then be directly compared.

5.1.1. Visualisation of PRR Amplification Products

An initial attempt to visualise the products of the PRR amplification was made by running the products on a 5% denaturing acrylamide gel, then staining with ethidium bromide, similar to the method of Sunkin and colleagues (Sunkin and Stringer, 1997). This gel proved to be difficult to manipulate and handle (data not shown) so a method using genescanning was used as an alternative.

For genescanning, PCR products were created using an oligonucleotide fluorescently labelled with the dye 6-FAM (6-carboxyfluorescein). One sample per lane was run on an ABI Prism™ automated sequencer machine and was subsequently analysed by GeneScan™ Analysis software (section 2.9.1). The size of each PCR product could be resolved to 1 bp using internal size standards.

To set up the genescan procedure, the primers PcProt19 and PcProt57 (fluorescently labelled) were used to amplify three DNA templates; rat-derived *Pneumocystis* genomic DNA from population 876.4, the genomic λ6.16 clone and the genomic cosmid clone 3G5 (Figure 5.1). The rat-derived *Pneumocystis* genomic DNA produced a profile of products, ranging from 300 bp to 600 bp while the PCR products from the two clones corresponded to the relevant size of their PRR. The PCR products were run on a genescan gel at twofold serial dilutions from undiluted to sixty four fold dilution. The genescanning software characterised five different PCR products amplified from the rat-derived *Pneumocystis* genomic DNA. Only one product was detected from each of the two clones and corresponded to the product seen on the agarose gel. Each PCR reaction had a different optimal running concentration on the genescan gel.

5.1.2. Template Choice for the PRR Assay Control Experiments.

Eight different templates were used to determine the relative importance of the different parameters for the PRR assay (Table 5.1.a). The 8 templates encompass a range of PRRs of different sizes and complexity. There are 3 different preparations of rat-derived *Pneumocystis* genomic DNA. There are genomic and cDNA clones containing both short and long PRRs. There is also an equimolar mix of a short PRR clone and a long PRR clone, which can be used to verify that there is no preferential amplification in a PCR of a short PRR over a long PRR. This can then be extrapolated to the genomic DNA from rat-derived

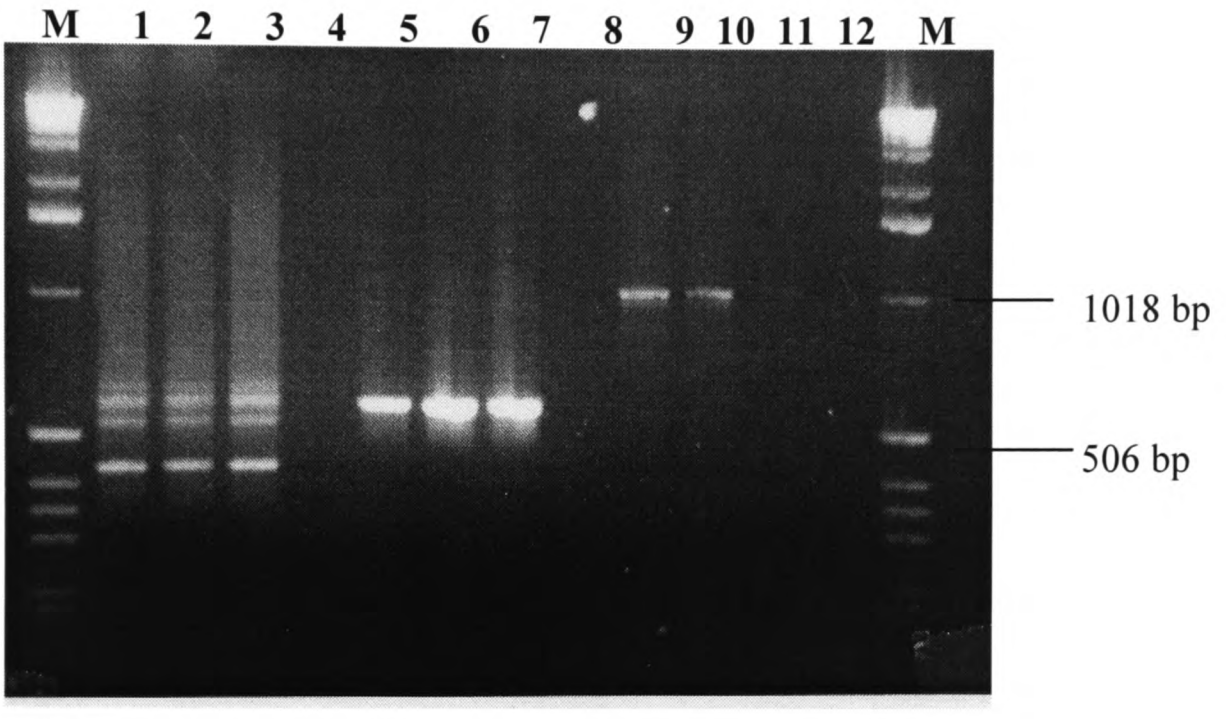


Figure 5.1. Electrophoretic separation of products from a PCR amplification using the primers PcProt19 and PcProt57. Lanes 1-3: rat-derived *Pneumocystis* genomic DNA 876.4, undiluted, x5 dilution and x25 dilution respectively; lanes 5-7: genomic λ6.16 clone, undiluted, x5 dilution and x25 dilution respectively; lanes 9-11: genomic cosmid clone 3G5, undiluted, x5 dilution and x25 dilution respectively; lanes 4, 8 and 12: negative control using dH_2O in place of a template. M, 1 kb DNA marker. The rat-derived *Pneumocystis* genomic DNA produced a profile of bands while the clones of DNA generated PRR products of the predicted size.

Template	Type of Template	Anticipated Result of PcProt19 and PcProt57 PCR
876.4	Complex genomic DNA	Profile of products
1553A	Complex genomic DNA	Profile of products
848.3	Complex genomic DNA	Profile of products
pAGA21	Genomic clone	≈ 800 bp
1B2	Genomic cosmid clone	2x ≈ 800 bp
λ6.16	Lambda genomic clone	≈ 500 bp
Mix of 1B2 and λ6.16	As above	2x ≈ 800 bp, 1x ≈ 500 bp
73j	cDNA clone	≈ 500 bp

Table 5.1(a) Description of the eight templates for the PRR PCR.

	***PRT1* Clones**						
Primer	1B21[2] (L)	1B22[2] (L)	pAGA (L)	*ORF-3*[1]	3G5 (L)	U62910[1]	U82999[1]
PcProt19		✔	✔	✔	✔	✔	
PcProt62			✔	✔			
PcProt63		✔	✔	✔			
PcProt64	✔				✔	✔	✔
PcProt65		✔	✔	✔		✔	
PcProt66		✔	✔	✔	✔	✔	
PcProt74					✔	✔	

Table 5.1(b) Specificity of primers to different *PRT1* clones. ✔ - the primer is specific to that clone, L – the clone has a long PRR, [1] – the length of the PRR has yet to be determined, [2] – this clone was not included in the original ClustalW sequence alignment when the primers were designed.

Pneumocystis populations where there should be an equal probability of each PRR being amplified by the PCR.

Figure 5.2(a) shows an example of a PCR amplification using the primers PcProt19 and PcProt57 to amplify the templates in table 5.1(a). A profile of products resulted from the amplification of rat-derived *Pneumocystis* genomic DNA. Ideally, this profile represents each PRR allele from each *Pneumocystis* genome within the population. It is possible that a single band from the profile on an agarose gel relates to several PRR products that are close in size. Differently sized PCR products of the PRRs should be resolved when the sample is run on a genescan gel. Analysis of amplification products of UCS/MSG junctions, show that one DNA band on an agarose gel can be equivalent to many different sequences (Keely *et al.*, 2003). So, a PCR product once it is resolved to 1 bp may represent different PRR alleles that are the same size but consist of different sequences. Ultimately, cloning and sequencing of a mixed band will determine how many PRR copies of different sizes and sequences are present. Some products that may not be visible on an agarose gel might be detected via the genescan method. Also, DNA samples have different migration patterns in agarose and acrylamide gels so the two different types of gel cannot be directly compared.

5.1.3. Primer Design for the PRR Assay

The primers PcProt19 and PcProt57 seem to preferentially amplify the shorter PRRs, for example the profile from population 876.4 in Figure 5.1. This phenomenon was investigated further using the genomic cosmid clone 3G5, which has a long PRR and the λ6.16 clone, which has a short PRR. Equimolar preparations of each clone were mixed in the following ratios of 3G5 to λ6.16; 8:1, 4:1, 2:1 and 1:1, and were amplified with PcProt19 and PcProt57. λ6.16 was always preferentially amplified, indicating that PcProt19 and PcProt57 preferentially amplify shorter PRRs (data not shown).

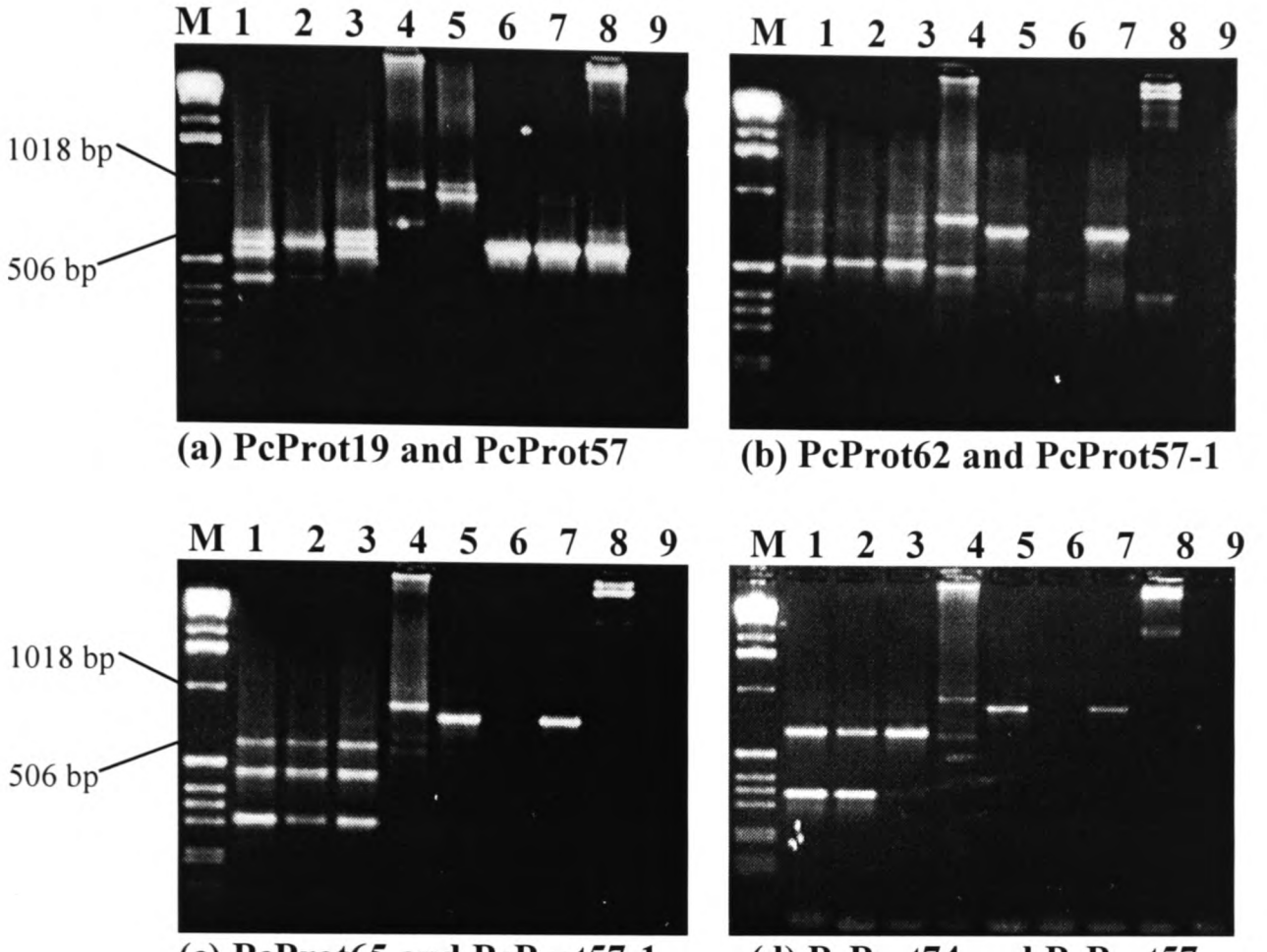


Figure 5.2. Electrophoretic separation of products from the PCR amplifications using different *PRT1* templates with different combinations of proline rich region specific primers. The primer pairs are indicated below each agarose gel. M, 1 kb DNA ladder. The templates for agarose gels a to d are; lanes 1-3: rat-derived *Pneumocystis* genomic DNA from populations 876.4, 1553A and 848.3 respectively; lane 4: genomic *PRT1* clone pAGA21; lane 5: genomic cosmid clone 1B2; lane 6: genomic *PRT1* clone λ6.16; lane 7: an equimolar mix of clones 1B2 and λ6.16; lane 8: cDNA *PRT1* clone 73j; lane 9: negative control using dH_20 in place of a template. The templates for agarose gels e and f are; lane 10: rat-derived *Pneumocystis* DNA 876.4 x10 dilution; lane 11: rat-derived *Pneumocystis* DNA 901.3; lane 12: genomic *PRT1* clone λ6.16; lane 13: genomic cosmid clone 1B2; lane 14: genomic cosmid clone 3G5; lane 15: cDNA *PRT1* clone 73j; lane 16: genomic *PRT1* clone pAGA21; lane 17: an equimolar mix of clones 3G5 and λ6.16; lane 18: negative control using dH_20 in place of a template. Different products were generated with each set of primers.

Five primers were designed to replace PcProt19, (Figure 5.3). Each primer binds to a different subset of sequences, (Table 5.1(b)). When the primers were designed, the sequences of the cosmid clones 1B2 and 11A11 had yet to be published. The sequences of these clones have subsequently been released and have been included in this analysis. None of the primers were specific for either of the two *PRT1* copies on cosmid clone 11A11, or cDNA clone 73j. Primers PcProt65 and PcProt66 were designed to bind to the maximum number of PRRs while PcProt62, PcProt63 and PcProt64 were designed to bind to subsets of PRR, preferably long PRRs.

Adding an extra thymine to the 5' end optimised the annealing temperature, T_m, of PcProt57. The new primer PcProt57-1 has the same T_m as the new PRR primers. PcProt66 has a strong secondary structure because of two runs of cytosine and guanine bases. An oligonucleotide, PcProt74, was designed that would bind 3 bp upstream of PcProt66, negating the secondary structure. PcProt74 has a T_m compatible with PcProt57. To determine primer specificity, the PRR specific primers were used to amplify by PCR the templates described in section 5.1.2 (Figure 5.2). PcProt63 with PcProt57-1 and PcProt64 with PcProt57-1 seemed to preferentially amplify the smaller PRRs, within the complex genomic DNA. Therefore, these primers were discarded. The remaining three primers, PcProt62, PcProt65 and PcProt74 when used in combination with either PcProt57 or PcProt57-1 seem to preferentially amplify the larger PRRs. Therefore, based on the *PRT1* sequence data available, the amplification products generated using these four primer sets should include the majority of the PRR alleles present in rat-derived *Pneumocystis* genomic DNA

5.1.4. Adapting the PRR assay to Characterise cDNA copies of the PRR

An RNA extraction method using TRIzol (GibcoBRL) was developed based on the work of Kyes et al (Kyes *et al.*, 2000). Briefly, the *Pneumocystis* parasites were broken down by

```
                    PcProt19                                                                2250
PC11A11PRT-1A  TTGTAGGTAA CTGGACTATT GATGTTGAAG ATAAAAATAT GTGGAATGAA CAAAGTTCTA
  PC1B2PRT-2   TTGTAGGTAA TTGGACTATC GATGTTGAAG ATAAAAAGGA TGAGAATCTA GATGGTGGAG
 PAGAGENBANK   TTGTAGGTAA TTGGACTATC GATGTTGAAG ATAAAAAGGA TGAGAATCTA GATGGTGGAG
  ORF3PRTDNA   TTGTAGGTAA TTGGACTATC GATGTTGAAG ATAAAAAGGA TGAGAATCTA GATGGTGGAG
    PC3G5PRT   TTGTAGGTAA TTGGACTATC GATGTTGAAG ATAAAAAGGA TCAGAATCTA GATGGTGAAG
  PC1B2PRT-1   TTGTAGGTAC CTGGACTATC GATGTTACAG ATAAAAATAA CTGGAATGAA CAAGGTTATA
      U62910   TTGTAGGTAA TTGGACTATC GATGTTGAAG ATAAAAAGGA ACCGAACCTA GATGGTGAAG
         73J   TTGTAGGTAA CTGGACTATC GATGTTGAAG ATGAAAAGGT TTCGAATCTA GATGGTGAAA
 PC11A11PRT-2  TTGTAGGTTA CTGGACTATC GATGTTAAAG ATAGAAATAG TTGGGATGAA GAAAGTCATA
      U82999   TTGTAGGTAA CTGGACTATC GATGTTGAAG ATAAAAAGGA TCAGAATCTA GATGGTGAAG

                                                          PcProt66 and PcProt74 2310
PC11A11PRT-1A  TTTTTAGTTG GCAACTTCAT TTTTTCGGAG AGTCTATTAA TTCAAGTAAA GCACAACCTC
  PC1B2PRT-2   TTTTTGATTG GCAACTTCAT TTTTTCGGGG AGTCTTGTGA ATCA...GAA GGCGTACCGC
 PAGAGENBANK   TTTTTGATTG GCAACTTCAT TTTTTCGGGG AGTCTTGTGA ATCA...GAA GGCGTACCGC
  ORF3PRTDNA   TTTTTGATTG GCAACTTCAT TTTTTCGGGG AGTCTTGTGA ATCA...GAA GGCGTACCGC
    PC3G5PRT   TTTTTTATTG GCAACTTCAT TTTTTCGGGG AGTCTTCTAA ATCA...GAA GGCGTACCGC
  PC1B2PRT-1   TTTTTAATTG GCAACTTCAT TTTTTCGGAG AGTCTATTAA TTCA...AGT AAAGCACTAC
      U62910   TTTTTGATTG GCAACTTCAT TTTTTCGGGG AGTCTTCTAA ATCA...GAA GGCGTACCGC
         73J   TTTTTGATTG GCAACTTCAT TTTTTCGGAG AGTCTATTGA TTCAAGTAAA GCAGAACTTC
 PC11A11PRT-2  TTTTTCTTTG GCAACTTCAT TTTTTCGGAG AGTCTATTAA TTCAAGTAAA CCACAACCTC
      U82999   TTTTTGATTG GCAACTTCAT TTTTTTGGAG AGTCTTTTGA TTCAAATAAA ...GTACCAG

                    PcProt65                                  2360 PcProt63
PC11A11PRT-1A  TCCATATCC TTTTGCGCAT AAACAACCAA CTACTGCGCC TCCACCAGAT CCAGGTGCTA
  PC1B2PRT-2   TCCTTCATA TCCTTTTCTA TCTAGATATC ...CAACTAC TACGCCTCCA CCAGATCCAG
 PAGAGENBANK   TCCTTCATA TCCTTTTCTA TCTAGATATC ...CAACTAC TACGCCTCCA CCAGATCCAG
  ORF3PRTDNA   TCCTTCATA TCCTTTTCTA TCTAGATATC ...CAACTAC TACGCCTCCA CCAGATCCAG
    PC3G5PRT   TGCTTCATA TCCTTTTCTA TCTATATATC ...CAACTAC TATGCCTCCG CCAGAACCAA
  PC1B2PRT-1   TCCTCCATA TCCTTTTGTG CATAAACAAC ...CAACTAC TATGCCTCCG CCAGAACCAA
      U62910   TGCTTCATA TCCTTTTCTA TCTAGATATC ...CAACTAC TACGCCTCCG CCAGAACCAA
         73J   TCCTCCATA TCCTTTTAAG CCTCAACCTC CTTCAAAACC TGCGCCTCCA TCAAAACCAG
 PC11A11PRT-2  TCCTCAATA TCCTTTTAAG CATATCCTTC ...CAACTAC TGCGCCTCCG CCAGATCCAA
      U82999   TCTTTCATT TCCTTTTATA TCTAGATATC ...CAACTAC TACGCCTCCG CCAGAACCAA

                and PcProt64                  PcProt62                          2430
PC11A11PRT-1A  CGCCTCCACC AGATCTAGAT GCTAACCCTC AACCTCAACC AGATCCTGGG TCTCCGCCTT
  PC1B2PRT-2   ATGCTACACC TTCTCCAGAT CTGGATGCTA .......... .......... ..........
 PAGAGENBANK   ATGCTACACC TTCTCCAGAT CTGGATGCTA .......... .......... .......ACC
  ORF3PRTDNA   ATGCTACACC TTCTCCAGAT CTGGATGCTA .......... .......... .......ACC
    PC3G5PRT   CTACTACGCT TCCACCAACT CCAAGTTCTA CATCTTCATC AAG...ACCA ...CCACCAC
  PC1B2PRT-1   CTACTACGCT TCCATCAGAT CTAGATGCTA ACCCTCAACC TCA...ACCA GATCCTGGGT
      U62910   CTACTACGCT TCCACCAACT CCAAGTTCTA CACCTTCATC AAG...ACCA CCACCGCCGC
         73J   ATCCTAACCC TCCATCAGAT CCTAGCTCTC AGCAAGATTC AGA...TACT TCGCTTTCAT
 PC11A11PRT-2  CTACTGCGCC TCCGCCAAAT CCAATTACTA CGCCTCCGCC T...TCATCA GATCCTGGGT
      U82999   CTACTACGCT TCCACCAACT CCAAGTTCTA CATCTTCATC AAGACCACCA CCACCTGCAC

                                  2450
PC11A11PRT-1A   CATCAGATCC TGAGTCTCCG
  PC1B2PRT-2    .......... ..........
 PAGAGENBANK    TTCAGCCAGA TTCAAATGCT
  ORF3PRTDNA    TTCAGCCAGA TTCAAATGCT
    PC3G5PRT    CTGCACCTCA ACCACAACCG
  PC1B2PRT-1    CTCCGCCTTC ATCAGATCCT
      U62910    CACCACCACC TGCACCTGCA
         73J    CAACTCCAAC TTCTACATCT
 PC11A11PRT-2   CTCAACCTCC ATCAAAACCT
      U82999    CTCAACCACA ACCACAACCA
```

Figure 5.3. ClustalW alignment of the beginning of the PRR of published *PRT1* sequences with the new PRR primers marked. The sequence is numbered from the beginning of the U62910 sequence. The oligonucleotide sequences are highlighted in pink except PcProt74 and PcProt64, which are in red. The names of the oligonucleotides are marked in bold above the relevant sequence and they are all in the sense direction, as indicated by the arrows. Residues that differ from the oligonucleotides are marked in blue

TRIzol, which contains guanidine isothiocyanate. This is acidic so preferentially solubilises RNA rather than DNA. The RNA was further purified by extraction with chloroform and isopropanol (section 2.12.1.).

The thymidylate synthase gene was used to develop an RT-PCR technique. This single copy housekeeping gene encodes a constitutively expressed protein (Edman *et al.,* 1989). It contains two introns so the cDNA and DNA RT-PCR products can be distinguished. Two primers, TS1 and TS2 (Table 2.4) were designed to amplify across these two introns. A control without enzyme was included in the first strand synthesis to monitor for contaminating DNA within the RNA preparation. Negative controls, using dH_2O in place of a template, were included at both stages of synthesis to monitor for contaminating RNA and DNA in the first and second strand synthesis stages respectively. An example of an RT-PCR amplification using the TS primer pair is shown in Figure 5.4(a). The technique was extended to the PRR primers. Figure 5.4(b) shows an example of an RT-PCR amplification using the PRR primers.

Profiles of PCR amplification products with length polymorphisms were being investigated so it was not possible to differentiate between DNA-derived and cDNA products especially as an intron has not been identified in the PRR. For this reason DNase treatment of the RNA is a crucial step. Two different DNases were considered, DNase I (Ambion) and DNA-free™ (Ambion) (section 2.12.2). Aliquots of the same RNA preparation from rat-derived *Pneumocystis* population 1966 were used to test each treatment. The DNase-treated RNA was used in an RT-PCR with the primers TS1 and TS2, (Figure 5.4 (a)). The size difference easily distinguished the cDNA and DNA products. The RNA preparation had very little contaminating DNA, as there is no product in any of the controls without enzyme. Possible presplicing cDNA products are faintly visible above the main cDNA transcript. There is very

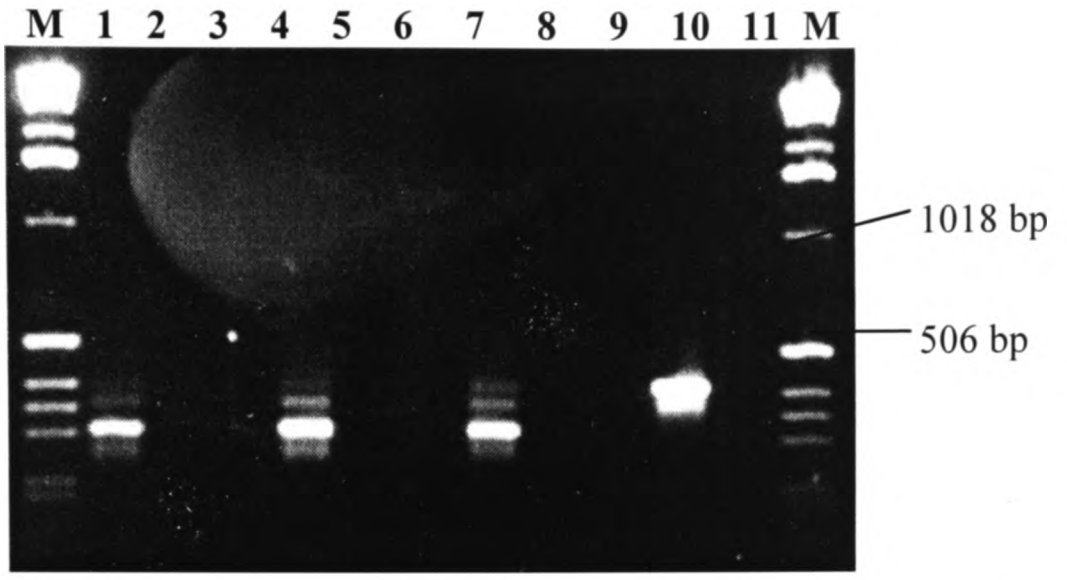


Figure 5.4(a) Electrophoretic separation of RT-PCR products, using RNA treated with different DNases. TS1 and TS2 primers, which amplify the single copy housekeeping gene, thymidylate synthetase, were used. M, 1 kb DNA ladder. Templates for cDNA synthesis were all extracted from rat-derived *Pneumocystis* population, 1966, lanes 1 and 2: RNA treated with DNase I; lanes 4 and 5: RNA treated with DNase-free™ and lanes 7 and 8 untreated RNA. Lanes 3, 6 and 9: negative control using dH_2O in place of a template. Lanes 1, 3, 4, 6, 7 and 9: plus reverse transcriptase; lanes 2, 5 and 8 minus reverse transcriptase. The templates for DNA-derived synthesis, lane 10: genomic DNA extracted from rat-derived *Pneumocystis* population 1966; lane 11 negative control using dH_2O in place of a template.

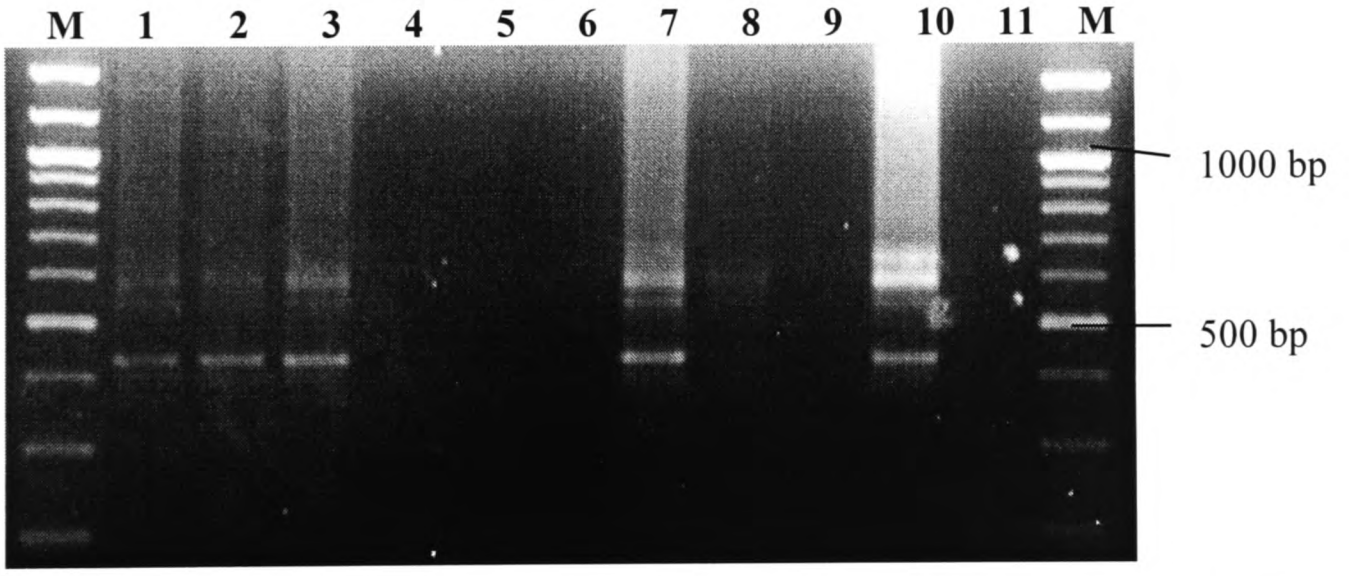


Figure 5.4(b). Electrophoretic separation of RT-PCR products, using RNA at different dilutions. M, 100 bp DNA ladder. Templates for cDNA synthesis, lanes 1-5: DNase treated rat-derived *Pneumocystis* RNA, extracted from rat-derived *Pneumocystis* population 1984, at undiluted, x5, x25, x125 and undiluted, respectively; lanes 6 and 9: dH_20 in place of a negative control; lanes 7 and 8: untreated, undiluted rat-derived *Pneumocystis* RNA, extracted from rat-derived *Pneumocystis* population 1984. Lanes 1-9 all reactions contained reverse transcriptase except lanes 5 and 8. PcProt19 was the first strand synthesis primer. Templates for DNA-derived synthesis, lane 10: rat-derived *Pneumocystis* genomic DNA undiluted, extracted from rat-derived *Pneumocystis* population 1984; lane 11: negative control using dH_2O in place of a template. The primers PcProt19 and PcProt57 were used.

little difference between the untreated RNA and the RNA after each of the two treatments. The product after DNA-free™ treatment appears to be more intense. Also, there is less risk of RNA degradation in the DNA-free™ enzyme inactivation step compared to the heat inactivation of the DNase I so the DNA-free™ treatment was chosen for future work.

To investigate the effect of RNA concentration on the efficiency of RT-PCR, RNA extracted from rat-derived *Pneumocystis* population 1984 was serially diluted and amplified with the primers PcProt19 and PcProt57, (Figure 5.4 (b)). The products are faint but the same cDNA profile can be seen for undiluted, x5 and x25 diluted RNA. There are no products for the x125 dilution. Therefore, RNA concentration produces an all or nothing effect; it is either fully amplified or not at all, under the conditions used. It can be assumed that there is no diminishment in the range of cDNA products amplified as the concentration of the RNA is reduced. In conclusion, it is possible to reproducibly extract RNA from *Pneumocystis* parasites and to amplify the PRR with specific primers, resulting in cDNA profiles.

5.1.5. Genescan and Proline Rich Region PCR Amplification Reproducibility.

Two different parameters influencing the reproducibility of amplification of PRR were investigated (data not shown). The first was repeat PCR amplifications run on the same genescan gel and the second was PCR amplification products run on two separate genescan gels. The set of templates described in section 5.1.2 minus the pAGA21 clone were amplified with the primer pairs PcProt74 with PcProt57 and PcProt65 with PcProt57-1. The templates generated identical products upon repeat PCR amplification. All the PCR products were run on a genescan gel at x 8 dilution. The genescan analysis demonstrated that the two PCR amplifications with primers PcProt65 and PcProt57-1 generated the same products, with only a small discrepancy of a couple of base pairs between products. The same was true for the PCR amplification with the primers PcProt74 and PcProt57, except rat-derived

Pneumocystis DNA from population 848.3 that had two PCR products in the first amplification and only one in the second.

The products from the two PCR amplifications with primers PcProt65 with PcProt57-1 and PcProt74 with PcProt57 were run on two separate genescan gels. As with the PCR reproducibility experiments, there were discrepancies between products of up to 6 bps. The products from the amplification with PcProt74 and PcProt57 generated similar results on both genescan gels. The same is true of the amplification with PcProt65 with PcProt57-1; however, 73j produced many peaks on the first genescan gel. These products were not visible on the agarose gel so presumably it was an artefact of that particular genescan gel.

5.1.6. Comparison of PRR DNA Profiles from Rat-Derived *Pneumocystis* Genomic DNA, Extracted from Populations from the Same Rat Colony

Rat-derived *Pneumocystis* genomic DNA was extracted from eight different rat-derived *Pneumocystis* populations, which were sacrificed within the same three month period from the rat colony in Oxford. PCR amplifications were planned, using four primer sets, to amplify the eight rat-derived *Pneumocystis* genomic DNA preparations. These PCR amplifications were repeated generating a total of eight different experiments. This data are summarised in Table 5.2.

	PcProt19 and PcProt57	PcProt65 and PcProt57-1	PcProt62 and PcProt57-1	PcProt74 and PcProt57
First Amplification	**1**	**2**	**3**	**4**
Second Amplification	**5**	**6**	**7**	**8**

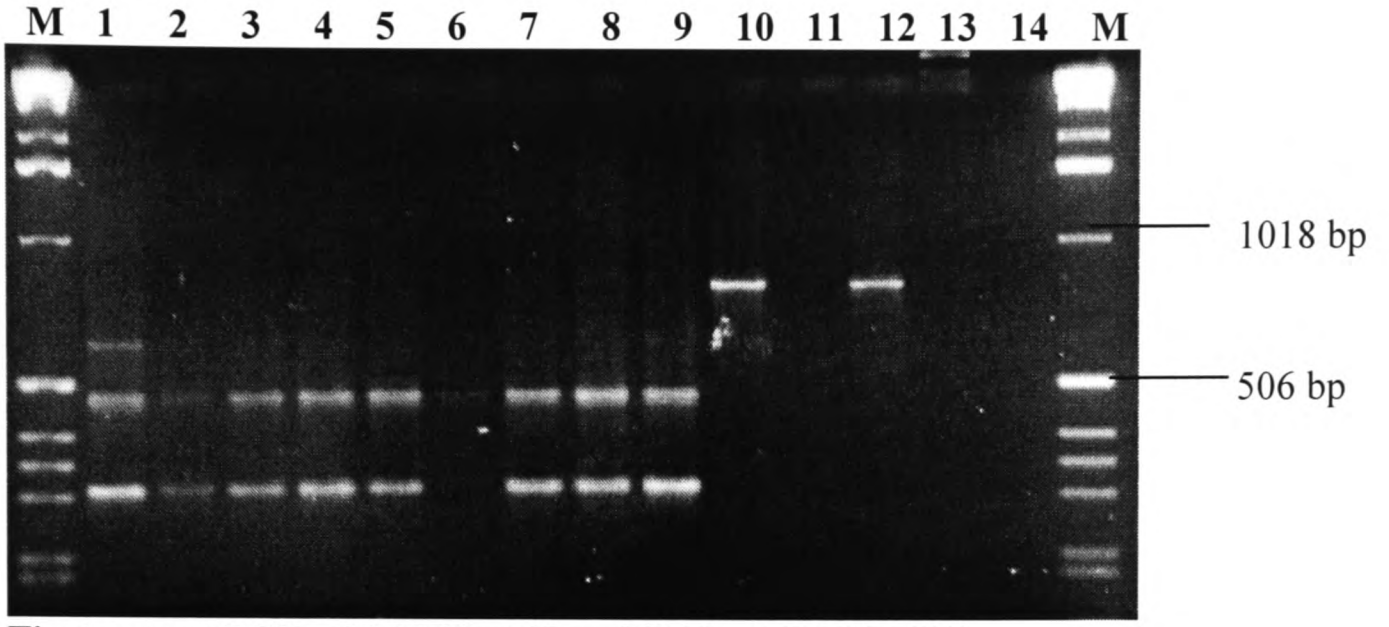
Table 5.2. Numbering of the eight different PCR amplifications with the four primer sets.

PCR amplification produced a series of PRR profiles for each rat-derived *Pneumocystis* genomic preparation. As an example, PCRs 2 and 6 will be described in detail. Figure 5.5. shows the results from the two repeat PCR amplifications. The genomic DNA from rat derived *Pneumocystis* population 1975 was of poor quality in both PCR amplifications, so no PRR profile had been generated for this population. The other profiles are reproducible apart from preparation 1984, which generated two bands on the first attempt and three on the second. The genescan results are shown in Table 5.3. Peak intensity was included, measured in relative fluorescence units, RFU, reflecting the amount of DNA in the sample.

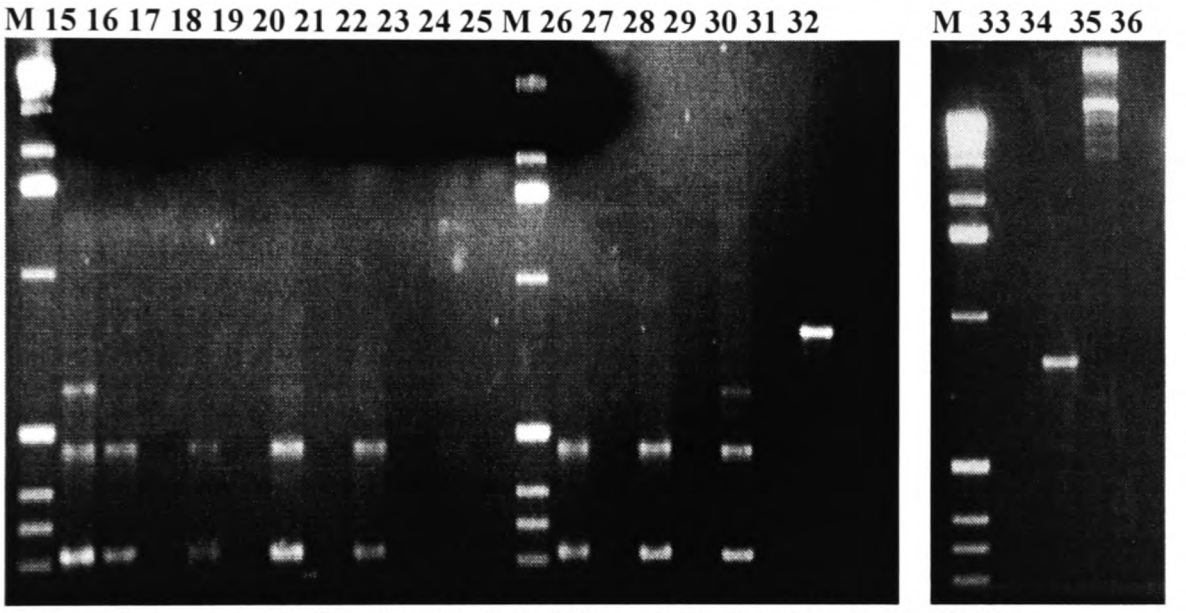
There were three main products in all eight rat-derived *Pneumocystis* genomic DNA preparations, except 1975, at 311 bp, 469 bp and 484 bp. There were many other minor peaks ranging in size from 115 bp to 841 bp and in fluorescence from <100 RFU to >1000 RFU. None of these were visible on the agarose gels, making their significance uncertain. Populations 876.4, 1968, 1977, 1978 and 1984 had a further PRR allele amplification product at ≈ 619 bp, This peak was visible on the agarose gel for populations 876.4 and 1984. There is very little variation of the PRR profiles between the populations.

Control DNA samples of known sizes were also included on the genescan gel. These were from the previous amplifications with PcProt65 and PcProt57-1. The same sized peaks were produced for population 876.4 apart from a couple of base pair discrepancies and minor peaks. A band of 830 bp was detected with all three 1B2 products but again some minor peaks were apparent. No significant products are detected with the λ6.16 samples and the MIX samples reflected the 1B2 results. Finally, there were no products with the 73j samples or the negative controls as expected.

Therefore, the genescan analysis generated accurate sizes for the amplification products and



First repeat PCR with PcProt65 and PcProt57-1 (PCR2)



Second repeat PCR with PcProt65 and PcProt57-1 (PCR6)

Figure 5.5. Electrophoretic separation of products from two PCR amplifications using the primers PcProt65 and PcProt57-1, (PCRs 2 and 6). M, 1 kb DNA ladder. A series of rat-derived *Pneumocystis* genomic DNAs extracted from different rat-derived *Pneumocystis* populations were amplified; lanes 2 and 16: population 1966; lanes 3 and 18: population 1967; lanes 4 and 20: population 1968; lanes 5 and 22: population 1970; lanes 6 and 24: population 1975; lanes 7 and 26: population 1977; lanes 8 and 28: population 1978; lanes 9 and 30: population 1984. All templates were diluted 1 in 5 apart from population 1975, which is undiluted. Lanes 1 and 15: control rat-derived *Pneumocystis* genomic DNA, 876.4. Clones of *PRT1* were amplified, lanes 10 and 32: genomic cosmid clone 1B2; lanes 11 and 33: genomic clone λ6.16; lanes 12 and 34: equimolar mix of 1B2 and λ6.16; lanes 13 and 35: cDNA clone 73j. Lanes 14, 17, 19, 21, 23, 25, 27, 29, 31 and 36: negative controls using dH_2O in place of a template.

PCR no	Template	Comments	Peak1	Peak2	Peak3	Peak4	Peak5	Peak6	Peak7	Peak8	Peak9
2	876.4			297		311	316	419		469	
6				298	305	311		419		469	
Control			296		305	311		419		469	
2	1966	Faint PCR			305	311			463	469	
6						311	316	419		469	
2	1967		115			311	316	420		469	
6					305	311			463	469	
2	1968				305	311		420	463	469	472
6						311	316	419		469	472
2	1970			298		311		420		469	472
6					305	311		420		469	
2	1975	Very Bad PCR	115		305	311			463	469	
6						311	316				
2	1977					311	316	419		469	471
6					305	311			463	469	472
2	1978			297	305	311		419	463	469	472
6				298		311	316	419		469	471
2	1984				305	311		419		469	471
6					306	311				469	471
2	1B2				305				463		
6							316				
Control					305						
2	6.16										
6											
Control			192								
2	MIX										
6											
Control											
2	73j		-								
6			-								
Control			-								
2	neg		-								
6			-								

Table 5.3. Genescan results for the two PCR amplifications using the primers PcProt65 and PcProt57-1. PCR2, first repeat, PCR6, second repeat, control, products from previous reactions whose sizes are already known. A series of rat-derived *Pneumocystis* genomic DNAs from different rat-derived *Pneumocystis* populations was used, 1966, 1967, 1968, 1970, 1975, 1977, 1978 and 1984. 876.4 was a control preparation of genomic rat-derived *Pneumocystis* DNA. DNA clones of *PRT1* were amplified, as controls of known sequence; genomic cosmid clone 1B2, genomic clone λ6.16, an equimolar mix (MIX) of 1B2 and λ6.16 and cDNA clone 73j. Neg, negative control using dH_20 in place of a template. Peaks are the sizes of each PCR product, measured in base pairs. The intensity of the peak is indicated by the colour the peak size is written in. The intensity of the fluorescent signal measured in relative fluorescent units (RFU) and has been graded into three

PCR no	Template	Peak10	Peak11	Peak12	Peak13	Peak14	Peak15	Peak16	Peak17
2 6 Control	876.4	478	484 484 484			619 619 622			
2 6	1966	478	484 484						
2 6	1967	478	484 484						
2 6	1968	478	484 484	489		619 619			
2 6	1970	478	484 484						
2 6	1975	478	484						
2 6	1977	478	484 484			619 618			
2 6	1978	478	484 484			619 618			841
2 6	1984	479	484 484			618 619	822		837
2 6 Control	1B2	478			607 607		817	830 831 831	
2 6 Control	6.16						822		835
2 6 Control	MIX			564	607		817	830 830 831	
2 6 Control	73j								
2 6	neg								

Table 5.3. cont. groups, <100, highlighted in green, >100 to <1000, highlighted in orange and >1000 highlighted in red.

gave an indication as to the amount of DNA present. However, the presence of minor peaks questions the validity of the analysis. Often there was an inconsistency of results between the two attempts at the PCR and it was hard to classify the validity of minor genescan peaks. Also, there was little difference seen between the eight populations to be able to distinguish them by their *PRT1* alleles.

To overcome the problem of a lack of variation between populations, a more diverse range of rat-derived *Pneumocystis* populations was analysed. Two rat-derived *Pneumocystis* genomic DNA samples, M54#17 and M59#5 were extracted approximately five years ago from a U.S.A. rat colony (kindly donated by M.T. Cushion). Four rat-derived *Pneumocystis* genomic DNA samples, 1099, 1441, 1553, 1697 were extracted from the Oxford rat colony over a six year period by other researchers and another population 1964 was also included. The last rat-derived *Pneumocystis* genomic DNA sample, 848.3, was extracted many years ago from a Glaxo rat colony. As previously, eight PCR amplifications were done and the results from the primers PcProt65 and PcProt57-1 are shown as an example (Figure 5.6). From the agarose gel, the main products at 311 bp and 469/ 484 bp were visible, however, here the 619 bp product was generated from every population and in samples 1553 and 1697, there is a possible further three minor products larger than 650 bp. However, there were no striking differences between the populations. The same was true for the other primer pairs. These PCR products were not subject to genescan analysis.

5.1.7. Comparison of PRR Profiles Amplified from cDNA and DNA Extracted from the Same Rat-Derived *Pneumocystis* Population

RNA was extracted from the previously described eight rat-derived *Pneumocystis* populations, section 5.1.6. The quality of the RNA was verified by RT-PCR using the TS primers and four preparations, 1966, 1967, 1968 and 1984, generated cDNA of a good

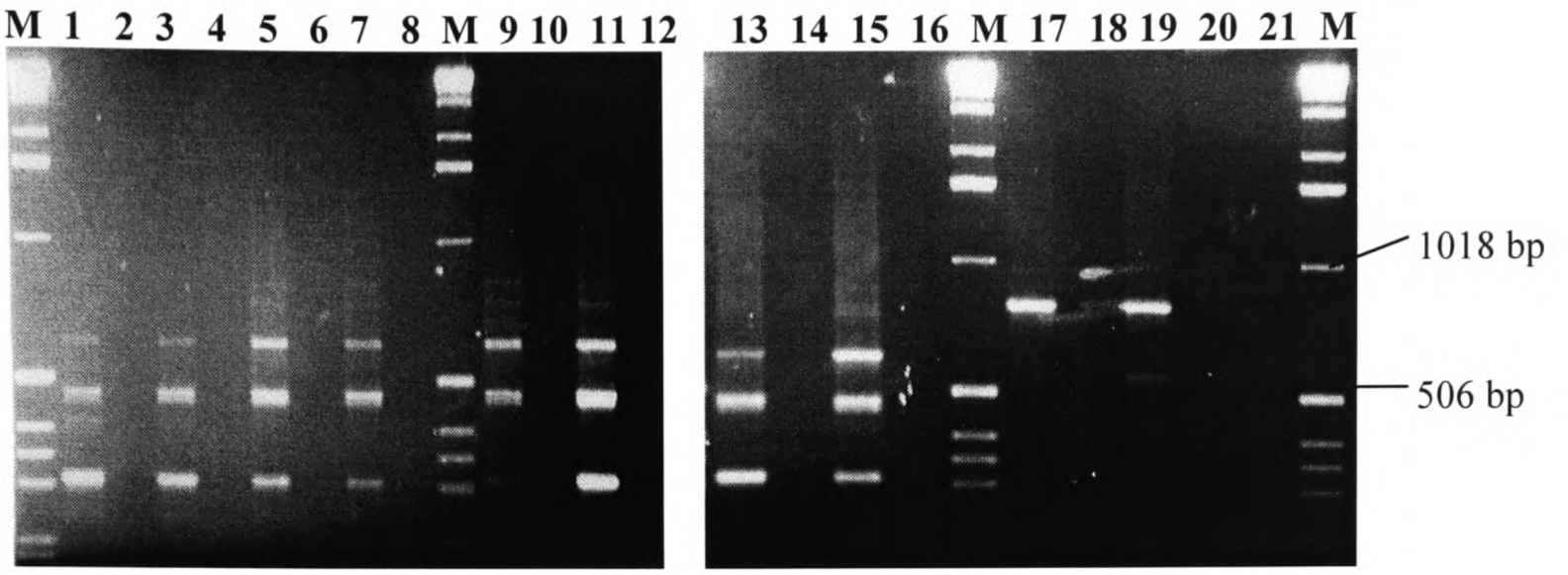


Figure 5.6. Electrophoretic separation of products from the PCR amplification, using different rat-derived *Pneumocystis* genomic DNA preparations and the primers PcProt65 and PcProt57-1. M, 1 kb DNA ladder. Lane 1 and 3: M54#17 and M59#5, respectively, both rat-derived *Pneumocystis* genomic DNA preparations from a rat colony in the U.S.A. from several years ago; lanes 5, 7, 9, 11 and 13: rat-derived *Pneumocystis* genomic DNA preparations 1099, 1441, 1553, 1697 and 1964 respectively, all from the rat colony in Oxford, covering a six year period. Lane 15: rat-derived *Pneumocystis* genomic DNA 848.3 from a rat colony developed over ten years ago at Glaxo; different *PRT1* clones, lane 17: genomic cosmid clone 1B2; lane 18: genomic clone λ6.16; lane 19: an equimolar mix of 1B2 and λ6.16; lane 20: cDNA clone 73j; lanes 2, 4, 6, 8, 10, 12, 14, 16 and 21: negative controls using dH_20 in place of a template.

quality. Two of the four primer sets were not optimal for RT-PCR as they failed to generate reproducible cDNA profiles. So, the two primer pairs PcProt19 with PcProt57 and PcProt74 with PcProt57 were used. The RT-PCR amplifications are summarised in Table 5.4.

Primer Pair	**RNA from Rat-Derived *Pneumocystis* Population**			
	1966	1967	1968	1984
PcProt19 and PcProt57	1	2	3	4
PcProt74 and PcProt57	5	6	7	8

Table 5.4. Numbering of the eight different RT-PCR amplifications with two primer pairs and four RNA preparations from rat-derived *Pneumocystis* populations

The results for RT-PCR amplifications 4 and 8, RNA from population 1984 amplified with the primers PcProt19 with PcProt57 and PcProt74 with PcProt57 respectively, are shown in Figure 5.7. The DNase treatment was effective because in both RT-PCR amplifications, there is no contaminating DNA as shown by the control without enzyme for the treated RNA compared with the untreated RNA control without enzyme. There is little difference between the cDNA and DNA PRR profiles with the primers PcProt19 and PcProt57 while there is a distinct difference for primers PcProt74 and PcProt57. The same was true for the other three RNA preparations, implying that with PcProt19 and PcProt57, there is similarity between the expressed and genomic copies of PRR across the populations but with PcProt74 and PcProt57, there is similarity between the populations yet, there is a difference between expressed and genomic copies of PRR. PcProt19 specifically binds to the majority of PRRs while PcProt74 specifically binds to a subset of PRRs as identified within the limits of the known *PRT1* sequence data (Table 5.1(b)) Therefore, PcProt74 amplifies less PRR copies, so any difference between expressed and genomic PRR copies may be easier to detect. These results demonstrate the complexity of the PRR assay, which is aiming to compare the differences between a multicopy gene family, within a non-clonal rat-derived *Pneumocystis*

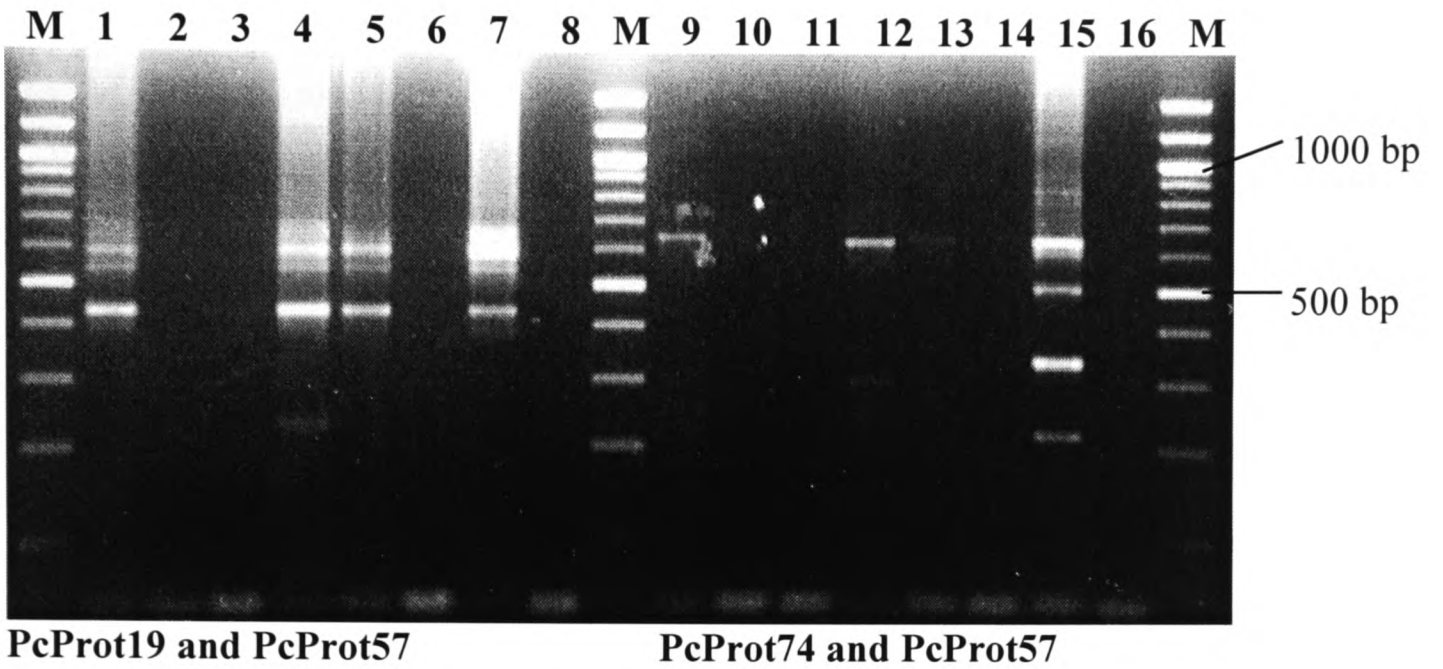


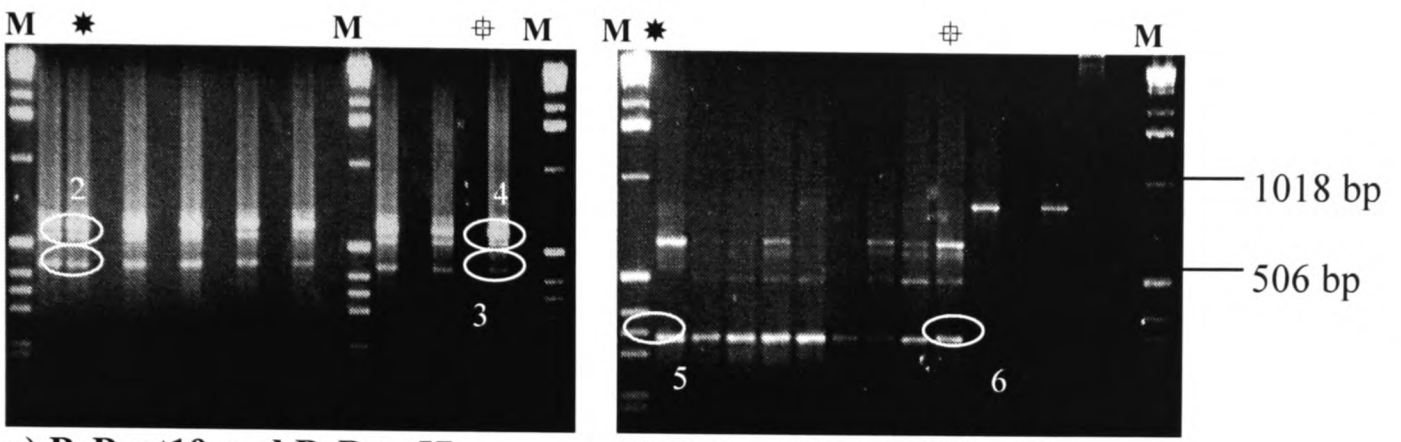
Figure 5.7. Electrophoretic separation of RT-PCR products M, 100 bp DNA ladder. Templates for cDNA synthesis, lanes 1, 2, 9 and 10,: DNase treated rat-derived *Pneumocystis* RNA, extracted from rat-derived *Pneumocystis* population 1984; lanes 4, 5, 12 and 13: untreated rat-derived *Pneumocystis* RNA, extracted from rat-derived *Pneumocystis* population 1984; lanes 3, 6, 11 and 14: negative control using dH_2O in place of a template; Lanes 1-6 and 9-14 contained reverse transcriptase except lanes 2, 5, 10 and 13. PcProt19 was used as the first strand synthesis primer for lanes 1-6 and PcProt74 was used as the first strand synthesis primer for lanes 9-14. Templates for DNA-derived synthesis, lanes 7 and 15: rat-derived *Pneumocystis* genomic DNA from population 1984; lanes 8 and 16: negative control using dH_2O in place of a template. The primers for lanes 1-8 were PcProt19 and PcProt57 while the primers for lanes 9-16 were PcProt74 and PcProt57. cDNA and DNA PRR profiles were generated with both sets of primers.

population. This apparent difference between a cDNA and DNA PRR profile from an RT-PCR amplification may possibly result from regulated *PRT1* gene expression, whereby only a subset of *PRT1* genes are expressed. None of these RT-PCR products were analysed by the genescan method.

5.1.8. Cloning and Sequencing of PRR Amplification Products

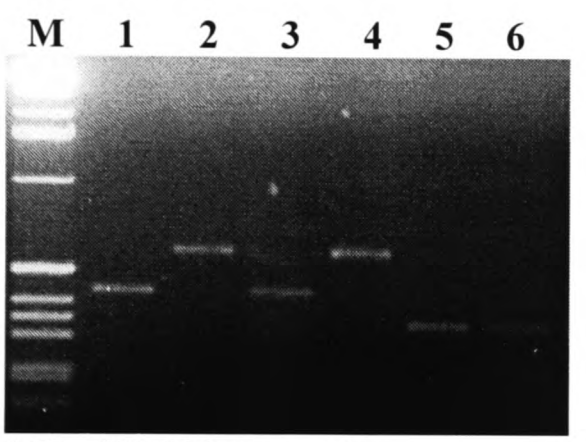
As discussed in section 5.1.2, many of the bands from the PRR profiles are very intense when visualised on an agarose gel. It is possible that they consist of more than one PRR allele. For this reason attempts were made to characterise these bands. From the RT-PCR amplification, which used the primers PcProt74 and PcProt57 to amplify the RNA and genomic DNA extracted from rat-derived *Pneumocystis* population 1984, the cDNA and genomic DNA-derived products were cloned without prior purification. Ten plasmids were extracted and six were sequenced; however, the sequence data obtained was not recognisable as rat-derived *Pneumocystis* DNA but from a BLAST search algorithm (Altschul *et al.*, 1997), had most homology with pyruvate dehydrogenase genes. This is probably due to an error or artefact in the cloning or sequencing steps.

A more precise approach was to excise specific PRR products from an agarose gel, then to purify and clone the DNA fragments (Figure 5.8). Six different bands were extracted from a number of profiles derived by PCR from templates of rat-derived *Pneumocystis* DNA extracted from populations 1966 and 1984. The first four were from two PRR profiles resulting from PCR amplifications using the primers PcProt19 and PcProt57. From both profiles, the lower band of ≈ 440 bp was excised and the middle band of ≈ 600 bp from the triplet of bands above the 500 bp marker. Two further bands were excised from the two PCR amplifications using the primers PcProt74 and PcProt57. The lower band of 320 bp was excised from both profiles. All six DNA bands were purified and cloned into the pGEM



a) PcProt19 and PcProt57

b) PcProt74 and PcProt57



c) Purified Fragments

Figure 5.8. Extraction of PRR products from gels after PCR amplification. The agarose gels a and b show the products from a PCR amplification using the primers as labelled under each gel. The lanes of interest are, M, 1kb DNA ladder. The template in lanes marked with ✸ is rat-derived *Pneumocystis* genomic DNA extracted from population 1966, and the template in the lanes marked with ⌖ is rat-derived *Pneumocystis* genomic DNA extracted from population 1984. The other lanes contain either products from PCR amplifications using genomic DNA from other rat-derived *Pneumocystis* populations or are negative controls using dH_2O in place of a template. The gel was run further and six products were excised as marked on the gels and then purified. Agarose gel c shows the six purified fragments. These were then cloned and sequenced.

vector. In total, fourteen plasmids were extracted and sequenced. At least two clones per fragment were sequenced.

Figure 5.9 shows the sequence alignment of the ≈ 440 bp products from the PCR amplification using the primers PcProt19 and PcProt57. Seven plasmids were successfully sequenced, four from population 1966 and three from population 1984. Both primers are evident at the ends of the sequence. There is very little sequence variation, no more than 10 variant nucleotides between all sequences. It could be assumed that this band represents only one PRR product or a selection of very invariant PRR alleles.

The larger fragment of ≈ 600 bp gave a different result; there is far more sequence and length variation. Figure 5.10 shows the sequence alignment for the larger PRR product from the PCR amplification using the primers PcProt19 and PcProt57. Two plasmids from each fragment were sequenced. Due to the differing length of the fragments and the nucleotide variation, the sequences were difficult to align. Each sequence is unique. Therefore, this band does consist of more than one sequence. It is possible that the sequence and length variation arises from sampling DNA from bands above and below the ≈ 600 bp band on the gel or that the variation is the result of artefacts induced during the preparation of the sequencing reactions by the repetitive nature of the proline rich region. More plasmids need to be sequenced to identify the extent of this variation. Like the previous attempt at sequencing the PRR products from amplifications using the primers PcProt74 and PcProt57, rat-derived *Pneumocystis* DNA sequence was not identified from the cloning of the final two fragments. Three plasmids were successfully sequenced but the same non-*Pneumocystis* sequence as before was generated.

```
            1                                                    50
plasmid3d   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG
plasmid3b   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG
plasmid3a   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG
plasmid1a   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG
plasmid1c   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG
plasmid1b   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG
plasmid1d   GGTAACTGGA CTATCGATGT TGAAGATGAA AAAGATCCGA ATCTAGATGG

            51                                                  100
plasmid3d   TGAAGTTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT TCTGATTCAG
plasmid3b   TGAAGNTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT TCTGATTCAG
plasmid3a   TGAAGTTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT TCTGATTCAG
plasmid1a   TGAAGTTTTT GATTGGCGAC TTCATTCTTT CGGAGAGTCT TCTGATTCAG
plasmid1c   TGAAGTTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT TCTGATTCAG
plasmid1b   TGAAGTTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT TCTGATTCAG
plasmid1d   TGAAGTTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT TCTGATTCAG

            101                                                 150
plasmid3d   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT
plasmid3b   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT
plasmid3a   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT
plasmid1a   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT
plasmid1c   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT
plasmid1b   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT
plasmid1d   AAAACGTACC ACCTCCTTCA TATCCTTTTC TATCTAGATA TCCAACTACT

            151                                                 200
plasmid3d   ACTCCTCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC
plasmid3b   ACTCCTCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC
plasmid3a   ACTCCTCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC
plasmid1a   ACTCCTCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC
plasmid1c   ACTCCTCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC
plasmid1b   ACTCCCCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC
plasmid1d   ACTCCTCCTC CAGATCCAGA TGCTACTCTT CCATCGACTC CAAGTTCTAC

            201                                                 250
plasmid3d   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA
plasmid3b   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA
plasmid3a   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA
plasmid1a   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA
plasmid1c   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA
plasmid1b   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA
plasmid1d   ATCTTCATCA AAAATAACTT CAAGCAAAAC TAAAACATCA ACCACTCGAA

            251                                                 300
plasmid3d   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
plasmid3b   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
plasmid3a   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
plasmid1a   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
plasmid1c   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
plasmid1b   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
plasmid1d   AAGCTTCATC TACTACAAAG ACCTCTGCAC GGCCATCTCC TACTAAGGGC
```

Figure 5.9 ClustalW sequence alignment of the ≈ 440 bp product from the PCR amplification of rat-derived *Pneumocystis* genomic DNA with PcProt19 and PcProt57. Plasmids 1a-d and plasmids 3a, b and d, from the reactions with rat-derived *Pneumocystis* genomic DNA from the populations, 1966 and 1984 respectively. PcProt19 is highlighted in red and PcProt57 is highlighted in green. Any nucleotide variation is highlighted in blue.

```
            301                                                       350
plasmid3d   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA
plasmid3b   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA
plasmid3a   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA
plasmid1a   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA
plasmid1c   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA
plasmid1b   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA
plasmid1d   ACTTCTACTG GATCAGGTGC TTCTCATCTT CCATTCTTCG AAAAAAGGCA

            351                                                       400
plasmid3d   TTTGTTACTT CAAATGATAT TATTCTTATT CTTTTTCTTA TTTTTGGGTT
plasmid3b   TTTGTTACTT CAAATGATAT TATTCTTATT CTTTTTCTTA TTTTTGGGTT
plasmid3a   TTTGTTACTT CAAATGATAT TATTGTTATT CTTTTTCTTA TTTTTGGGTT
plasmid1a   TTTGTTACTT CAANNGATAT TATTCTTATT CTTTTTCTTA TTTCTGGGTT
plasmid1c   TTTGTTACTT CAAATGATAT TATTCTTATT CTTTTTCTTA TTTTTGGGTT
plasmid1b   TTTGTTACTT CAAATGATAT TATTCTTATT CT~~~~~~~~ ~~~~~~~~~~
plasmid1d   TTTGTTACTT CAAA~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

            401                                                 436
plasmid3d   ACTCTTTTTA ACCTATAGGA TTTTTAAGAA AACGG
plasmid3b   ACTCTTTTTA ACCTATAGGA TTTTTAAGAA AACGG
plasmid3a   ACTCTTTTTA ACCTATAGGA TTTTTAAGAA AACGG
plasmid1a   ACTCTTTTNA AC~~~~~~~~ ~~~~~~~~~~ ~~~~~
plasmid1c   ACTCTTTTTA ACCTATAGGA TTTTTAAGAA AACGG
plasmid1b   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~
plasmid1d   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~
```

Figure 5.9. cont…

```
            1                                                    50
plasmid4a   GGTAACTGGA CTATCGATGT TGAAGATAAA AAGGATCAGA ATCTAGATGG
plasmid2a   GGTAACTGGA CTATCGATGT TGAAGATAAA AAGGATCAGA ATCTATATGG
plasmid2b   GGTAACTGGA CTATCGATGT TGAAGATAAA AAGGATCAGA ATCTAGATGG
plasmid4d   GGTAACTGGA CTATCGATGT TGAAGATGAA AAGGTTTCGA ATCTAGATGG

            51                                                  100
plasmid4a   TGAAGTTTTT GATTGGCAAC TTCATTTTTT TGGAGAGTCT TTTGATTCAA
plasmid2a   TGAAGTTTTT GATTGGCAAC TTCATTTTTT TGGAGAGTCT TTTGATTCAA
plasmid2b   TGAAGTTTTT GATTGGCAAC TTCATTTTTT TGGAGAGTCT TTTGATTCAA
plasmid4d   TGAAATTTTT GATTGGCAAC TTCATTTTTT CGGAGAGTCT ATTGATTCAA

            101                                                 150
plasmid4a   ATAAAGTA.. .CCAGCTCTT TCATTTCCTT TTATATCTAG ATATCCAACT
plasmid2a   ATAAAGTA.. .CCAGATCTT TCATTTCCTT TTATATCTAG ATATCCAACT
plasmid2b   ATAAAGTA.. .CCAGCTCTT TCATTTCCTT TTATATCTAG ATATCCAACT
plasmid4d   GTAAAGCAGA ACTTCATCCT CCATATCCTT TTAAGCCTCA ACCTCCTTCA

            151                                                 200
plasmid4a   A...CTACGC CTCCGCCAGA ACCAACTACT ACGCTTCCAC CAACTCCAAG
plasmid2a   A...CTACGC CTCCGCCAGA ACCAACTACT ACGCTTCCAC CAACTCCAAG
plasmid2b   A...CTACGC CTTCGCCAGA ACCAACTANT ACGCTTTCAC CAACTNCAAG
plasmid4d   AAACCTGCGC CTCCATCAAA ACCAGATCCT AACCCTCCAT CAGATCCTAG

            201                                                 250
plasmid4a   TTCTACATCT TCATCAAGAC CACCACCACC TGCACCTCAA CCACATCCAC
plasmid2a   TTTTACACCT TCATCAAGAC CACCACCACC TGNACCTCAA CCACAACCAC
plasmid2b   TTCTACATTT TTATCAA... .......... .......... ..........
plasmid4d   CTCT.CAGCA AGATTCAGAT .ACT.TCGCT TTCA..TCAA CTCCAACTTC

            251                                                 300
plasmid4a   AACCAGAGCC AGAACCAAGA AACAGAACAA ATACCACC.. ..........
plasmid2a   AACCAGAGCC AGAACCA.GA ACCAGAACAA ATACCACCTN GANCAACANC
plasmid2b   .......... .......... .......... .......... ..........
plasmid4d   TAC...ATCT ....TCA... .TCAAAACTG TCACCACCAC CTACA.....

            301                                                 350
plasmid4a   ..CTTCGACC AACAACA... ......CCTT GACAAAAACC AACATCAATA
plasmid2a   AGCTCTGACC AACATCAGAA CCACAACCGG AACAAAAACC AACATCAATA
plasmid2b   .......... ......A... .......... .ACAAAAACC ANCATCAATA
plasmid4d   ..CCTCAACC AAAGCCAGAA CCACAACCGG AACAGAAACC GACATCAATA

            351                                                 400
plasmid4a   GCTTCATCTA CGACTTCAAC TAATTTAATT TCACCAGCTC CCACATCTTC
plasmid2a   GCTTCATCTA CGACTTCAAC TAATTTAATT TCACCAGCTC CCACATCTTC
plasmid2b   GCTTCATNTN CGANTTCAAC TAATTTAATT TCACCAGCTC CCACAT.TTC
plasmid4d   GCTTCATCTA CAACATCAAC TAATTTAATT CCACCAGCTC CCACATCTTC
```

Figure 5.10. ClustalW sequence alignment of the larger product from the PCR amplification of rat-derived *Pneumocystis* Genomic DNA with PcProt19 and PcProt57. Plasmids 2a-b and plasmids 4a and b, from the reactions with rat-derived *Pneumocystis* genomic DNA from the populations, 1966 and 1984 respectively. PcProt19 is highlighted in red and PcProt57 is highlighted in green. Any nucleotide variation is highlighted in blue

```
           401                                                   450
plasmid4a  ATCAAGCAAA ACTGAAATAT CAACCACTCA AAAAGTTTCA TCTACTACAA
plasmid2a  ATCAAGCAAA ACTGAAATAT CAACCACTCA AAAAGTTTCA TCTACTACAA
plasmid2b  ATCAAGCAAA ACTGAAATAT CAACCACTCA AAAAGTTTCA TCCACTACAA
plasmid4d  ATCAAGCAAA ACTAAAACAT CAACCACTCG AAAAGCTTCA TCTACTACAA

           451                                                   500
plasmid4a  CAGCCTCTAC GCTGCCATCT CCTACTCAGG GCACTTCTAC TGAATCAGGT
plasmid2a  CAGCCTCTAC GCTGCCATCT CCTACTCAGG GCACTTCTAC TGAATCAGGT
plasmid2b  CAGCCTCTAC GCTGCCATCT CCTACTCAGG GCACTTCTAC TGAATCAGGT
plasmid4d  AAACCTCTAC ACGGCCGTCT CCTACTGAGG GTACTTTTAC TGGATCAGGT

           501                                                   550
plasmid4a  ACTTCTC.TC TTTCATTCTT TAAAAAAAAG GAATTTGTTA TTTCGATTGT
plasmid2a  ACTTCTCGTC TTTCATTCTT TAAAAAAAAG GAATTTGTTA TTTCGATTGT
plasmid2b  ACTTCTCGTC TTTCATTCTT TAAAAAAAAG GAATTTGTTA TTTCGATTGT
plasmid4d  GCTTCTCATC TTTCATTCTT CGAAAAAA.G GCATTTGTTA CTTCAAATGA

           551                                                   600
plasmid4a  TATTGTTATT ATTCTT.... ..ATTTTTGG GTTACTCTTT TTA.CTTATA
plasmid2a  TATTGTTATT ATTCTT.... ..ATTTTTGG GTTACTCTTT TTA.CTTATA
plasmid2b  TATTGTTATT ATTCTT.... ..ATTTTTGG GTTACTCTTT TTA.CTTATA
plasmid4d  TATTATTGTT ATTCTTTTTC TTATTTTTGG GTTACTCTTT TTAACCTATA

           601            618
plasmid4a  GGATTTTTAA GAAAACG
plasmid2a  GGATTTTTAA GAAAACG
plasmid2b  GGATTTTTAA GAAAACG
plasmid4d  GGATTTTTAA GAAAACG
```

Figure 5.10 cont.

The phenomenon of PcProt74 and PcProt57 not amplifying rat-derived *Pneumocystis* DNA was investigated further. Three preparations of rat derived *Pneumocystis* genomic DNA extracted from the populations 1966, 1984 and 876.4 were used as templates for the PCR amplification with the primers PcProt74 with PcProt57 and the upper band of ≈ 600 bp was excised, purified and cloned. Six plasmids were extracted and four were successfully sequenced, with at least one successful sequence per fragment. The sequencing alignment is shown in Figure 5.11. All four plasmids contain *Pneumocystis* DNA so in this instance PcProt74 and PcProt57 are specific to and have amplified *Pneumocystis* DNA. Each individual sequence is sufficiently divergent to be considered unique so the band appears to represent a different PRR allele from each population and two different alleles from population 876.4. Again the variation could be due to artefacts induced during the preparation of the sequencing reactions because of the repetitive sequence of the proline rich region. More plasmids need to be sequenced to characterise this variation.

In order to try to explain phenomenon seen with the primers PcProt74 and PcProt57, the three purified ≈ 600 bp PRR products were used as templates for reamplification with these primers. The unexpected result was a series of smaller bands for each template. In order, to verify this result, one of the larger bands from this series of bands was reamplified and again the same effect of a series of smaller bands was generated, indicating some likely artefact during PCR amplification.

5.1.9. Assessment of PRR Assay

At this point the PRR assay was rejected as a suitable assay. The experiments had shown that cDNA and DNA PRR profiles of a high enough quality could be generated with selected primers, however, too many amplification artefacts were generated to permit reliable analysis using the genescan software.

```
          1                                                          50
 1966_2h  CTAAATCAGA AGGCGTACCG CCTGCTTCAT ATCCTTTTCT ATCTAGATAT
876.4_1h  CTAAATCAGA AGGCGTACCG CCTGCTTCAT ATCCTTTTCT ATCTAGATAT
 1984_3a  CTAAATCAGA AGGCGTACCG CCTGCTTCAT ATCCTTTTCT ATCTAGATAT
876.4_1e  CTAAATCAGA AGGCGTACCG CCTGCTTCAT ATCCTTTTCT ATCTAGATAT

          51                                                        100
 1966_2h  CCAACTACTA CGCCTCCGCC AGATCCAGAT GCTACGCCTC CACCAGATCC
876.4_1h  CCAACTATTA CGCCTCCGCC AGATCCAGAT GCTACGCCTC CACCAGATCC
 1984_3a  CCAACTACTA CGCCTCCGCC AGATCCAGAT GCTACGCCTC CACCAGATCC
876.4_1e  CCAACTACTA CGCCTCCGCC AGATCCAGAT GCTACGCCTC CACCAGATCC

          101                                                       150
 1966_2h  GGATGCTAAC CTTCAGCCAG ATTCAAATGC TGACCCTCAA CCTCAACCAG
876.4_1h  GGATGCTAAC CTTCAACCAG ATTCAAATGC TGACCCTCAA CCTCAACCAG
 1984_3a  GGATGCTAAC CTTCAGCCAG ATTCAAATGC TGACCCTCAA CCTCAACCAG
876.4_1e  GGATGCTAAC CTTCAGCCAG ATTCAAATGC CGACCCTCAA CCTCAACCAG

          151                                                       200
 1966_2h  ATGTTAAGCC TCCGCCTTCA TCAGAACCTG CGCCTCAACC TCCATTAGAT
876.4_1h  ATGTTAAGCC TCCGCCTTCA TCAGAACCTG CGCCTCAACC TCCAGTAGAT
 1984_3a  ATGTTAAGCC TCCGCCTTCA TCAGAACCTG CGCCTCAACC TCCATTAGAT
876.4_1e  ATGTTAAGCC TCCGCCTTCA TCAGAACCTG CGCCTCAACC TCCATTAGAT

          201                                                       250
 1966_2h  CCTCTGCCTC CAACAAAACC AGATCCTAAC CCTCCATCAG ATTCTAGCTC
876.4_1h  CCTCTGCCTC CAACAAAACC AGATCCTAAC CCTCCATCAG ATTTTAGCTC
 1984_3a  CCTCTGCCTC CAACAAAACC AGATCCTAAC CCTCCATCAG ATTCTAGCTC
876.4_1e  CCTCTGCCTC CAACAAAACC AGATCCTAAC CCTCCATCAG ATTCTAGCTC

          251                                                       300
 1966_2h  TCAGCAAGAT CCAGATACTT CGCTTTCATC AAATCCAACT TCTACATCTT
876.4_1h  TCAGCAAGAT CCACATACTT CGCTTTATCA AATCCAACTT CATACATCTT
 1984_3a  TCAGCAAGAT CCAGATACTT CGCTTTCATC AAATCCAACT TCTACATCTT
876.4_1e  TCAGCAAGAT CCAGATACTT TGCTTTCATC AAATCCAACT TTTACATCTT

          301                                                       350
 1966_2h  NATCAGAACC .......... .......... .......... ..........
876.4_1h  TATCAAAANC .......... .......... .......... ..........
 1984_3a  TATCAGAACC ACCACCACCG CCACCTGCAC CACCACCACC ACCGGAGCAA
876.4_1e  CATCGGAACC ACCACCACCG TCACCTGCAC CTCCACCAAA ACCACAACCA
```

Figure 5.11 ClustalW sequence alignment of the ≈ 600 bp product from the PCR amplification of rat-derived *Pneumocystis* genomic DNA with PcProt74 and PcProt57. The plasmids from the reactions with rat-derived *Pneumocystis* genomic DNA from the populations, 1966, 1984 and 876.4 are indicated. PcProt74 is highlighted in red and PcProt57 is highlighted in green. Any nucleotide variation is highlighted in blue.

```
          351                                                        400
 1966_2h  .......... .......... ...ACCACCA TCGGCA.... ....TTCGTC
876.4_1h  .......... ...CAACATC AATAACTTCA TCTACATCTA CGACTTCATC
 1984_3a  AAGCCAACAT CAATAACTTC .........A TCTACATCTG CGNCTTCATC
876.4_1e  CCACCACCAC CACCACCACC GGAACAAAAA CCAACATCAA TAACTTTATC

          401                                                        450
 1966_2h  AAGCAAAACT AAAATAT.CA ACCACTCGAA AAGCTTCATC TACTAAAGCT
876.4_1h  AAGCAAAACT AAAATAT.CA ACCACTCGAA AAGCTTCATC TACTAAAGCT
 1984_3a  AAGCAAAACT AAAATAT.CA ACCACTCGAA AAGCTTCATC TACTAAAGCT
876.4_1e  TACATCTACG ATAAAATATC ACCACTCGAA AAGCTTCATC TACTAAAGCT

          451                                                        500
 1966_2h  TCATCTACTA CAAAAACTTC TACACGGCCG TCTCCTACTG AGGGTACTTT
876.4_1h  TCATCTACTA CAAAAACTTC TACACGGCCG TCTCCTACTG AGGGTACTTT
 1984_3a  TCATCTACTA CAAAAACTTC TACACGGCCG TCTCCTACTG AGGGTACTTT
876.4_1e  TCATNTACTA CAAAAACTTC TACACGACCG TCTCCTACTG AGGGTNNNTT

          501                                                        550
 1966_2h  TACTGGATCA AGTGCTTCTC ATCTTTCATT CTTCGAAAAA AGGCATTTGT
876.4_1h  TACTGGATCA AGTGCTTCTC ATCTTTCATT CTTCGAAAAA AGGCATTTGT
 1984_3a  TACTGGATCA AGT.....TC ATCTTTCATT CTTCGAAAAA AGGCATTTGT
876.4_1e  TACTGGATCA AGTGCTTCTC ATCTTTCATT CTTCGAAAAA AGGCATTTGT

          551                                                        600
 1966_2h  TACTTCAAAT GATATTATTG TTATTCTTTT TCTTATTTTT GGGTTACTCT
876.4_1h  TACTTCAAAT GATATTATTG TTATTCTTCC TCTTATTTTT GGGTTACTCT
 1984_3a  TACTTCAAAT GATATTATTG TTATTCTTTT TCTTATTTTT GGGTTACTCT
876.4_1e  TACTTCAAAT GATATTATTG TTATTCTTTT TCTTATTTTT GAGTTACTCT

          601                                630
 1966_2h  TTTTAACCTA TAGGATTTTT AAGAAAACGG
876.4_1h  TTTTAACCTA TAGGATTTTT AAGAAAACGG
 1984_3a  TTTTAACCTA TAGGATTTTT AAGAAAACGG
876.4_1e  TTTTAACCTA TAGAATTTTT AAGAAAACGG
```

Figure 5.11. cont

5.2. Discussion

The aim of this part of the study was to investigate the diversity and number of genomic and expressed copies of *PRT1* and to determine whether *PRT1* gene expression is regulated. It was hoped that length polymorphism of the cDNA and DNA PRR amplification profiles would provide enough variation. Potentially, there could be 30 different *PRT1* alleles per organism (Lugli *et al.,* 1997). One rat-derived *Pneumocystis* population amplified by the four primer sets produced no more than 10 different PRR bands. Genomic rat-derived *Pneumocystis* DNA from diverse rat colonies from different countries extracted at different time points within a twelve year period, did not provide data with enough variation so that the difference between the *PRT1* genes could be compared.

The PRR PCR amplification was not sufficiently reliable and reproducible. The PRR by definition contains much repetitive DNA sequence that can cause mispriming, primer slippage and polymerase jumping within the elongation step (Meyerhans *et al.,* 1990, Paabo *et al.,* 1990). A combination of these artefacts could be responsible for the results seen in section 5.1.8. A series of bands was generated from a template consisting of cloned PRRs of one length. Therefore, a cDNA or DNA PRR profile could contain an unknown proportion of amplification artefacts.

The data obtained from the PRR assay was difficult to interpret. Four primer sets were used to amplify a polymorphic region of a multicopy gene within a non-clonal population. The assay appears not to be sufficiently sensitive to detect the level of variation required. For example, it is difficult to identify a level of variation of 10% between the PRR profiles because this difference will be overwhelmed by the 90% of data that is identical.

The primers used do not appear to amplify every PRR from the genomic DNA extracted from the rat-derived *Pneumocystis* populations. There is not enough published *PRT1* sequence data to be able to design a universal pair of primers that bind to every PRR domain. Only a subgroup of PRRs can be amplified so more than one pair of primers needs to be used, complicating the overall dataset. Also, the primers appear to preferentially amplify the shorter PRRs. It was hoped that one band would consist of many PCR products of the same length but with different sequences. This was shown to be the case with UCS/MSG products (Keely *et al.*, 2003). Cloned products with different sequences were identified with the larger ≈ 600 bp band. Unfortunately, little variation was seen with the cloned products derived from the smaller 440 bp band.

The concentration of the RNA does not seem crucial but an effective DNase treatment is necessary. There are no introns in the PRR to distinguish between cDNA and DNA PRR PCR products. Thus, the RNA has to be free of all contaminating DNA. A DNase step was included in the PRR RT-PCR and ideally; an RT-PCR test using the TS primers should have been performed on every RNA preparation to verify the efficiency of this step. Also, the same amplification artefacts that have been described for the PRR PCR are likely to apply in this case.

The results of the genescan analysis proved to be difficult to interpret. Firstly, data from agarose and acrylamide gels cannot be directly compared. Not all peaks from genescan analysis can be anticipated in terms of number and intensity of bands visible on an agarose gel. The loading of genescan gels was subject to many inconsistencies. The reproducibility of loading aliquots onto a gel could alter the amount of DNA within a sample, changing the peak intensity of a product so it could appear more or less significant than it really was.

The genescan electropherograms were hard to analyse. A DNA product would appear as a double peak so consistency was needed when interpreting them. Often a single DNA product of one size would appear as a series of multiple peaks rather than one single peak. This was probably the result of multiple products because of inexact primer annealing or the plus A artefact, which is the tendency of Taq polymerase to add a non-templated nucleotide, usually adenine, to the 3' end of double-stranded DNA. These multiple peaks were very hard to interpret. The resolution of product size over 800 bp tended to be poor as there were few internal size standards at this range. Internal size standards were crucial to genescan gel reproducibility. Also, the genescan software could not interpret peaks with an intensity larger than 4000 RFU. Overall, the technique was too sensitive to fluctuations between individual samples.

In order to make the data easier to interpret, the PRR profiles could be cloned and sequenced either as plasmids or as PCR products from a PCR screen. However, this strategy would be time consuming and many clones per population would have to be sequenced in order to gain a comprehensive picture of the PRR alleles of the population. Also, as previously described this method is subject to artefacts induced during PCR amplification.

The PRR assay did not provide data to satisfy our initial aim. A successful assay needs to be more reliable and reproducible, and less complex, using universal primers to investigate a better characterised sequence.

6. Characterisation of the Complex Pattern and the Possible Regulation of *PRT1* Gene Expression within the Context of Simple Populations of Rat-Derived *Pneumocystis*

The aim of this part of the study was to develop a second strategy to investigate the diversity and number of genomic and expressed copies of *PRT1* and to determine whether *PRT1* gene expression is regulated. This was attempted with potentially clonal populations of rat-derived *Pneumocystis* obtained via the low dose inoculation method. Like *MSG* genes, one *PRT1* gene could be expressed per *Pneumocystis* organism. Alternatively, a few *PRT1* genes could be expressed per *Pneumocystis* organism or no regulation of *PRT1* genes may occur, resulting in all the *PRT1* genes within the *Pneumocystis* genome being expressed per *Pneumocystis* organism.

6.1.1. Simple Populations of Rat-Derived *Pneumocystis*

Twelve to seventeen week old nude Fischer rats (Pasteur Institute, Lille, France) free of *Pneumocystis* infection were housed in barrier conditions, treated with dexamethasone (1mg L^{-1} in the drinking water) for 5 weeks and were then intratracheally inoculated with approximately 1 to 10 organisms per rat (Keely *et al.,* 2003), (section 2.6.1). The complex population inoculum of 0.25 x 10^6 parasites ml^{-1} was serially diluted four times, 1/25, 1/250, 1/2500, 1/25000, resulting in a dilution equivalent to 1 parasite per 100µl. The same preparation of diluted organisms was used to inoculate each rat. Eight weeks post-inoculation, the animals were sacrificed and organism burden for each rat was enumerated by light microscopy of a small aliquot of lung tissue. The average number of organisms per animal was 8 ± 0.75 x 10^8 (S.D.). In total, five simple populations of rat-derived

Pneumocystis organisms were produced by this method, EA15, EA17, EA19, EA21 and EA24. These animal model experiments were completed in France, in the spring of 2000 by E. Dei-Cas and E.M. Aliouat.

Three of the nude rats acted as sentinels (not inoculated with rat-derived *Pneumocystis* organisms). They were housed under the same barrier conditions in the same unit at the same time as the experimental rats in order to monitor for an external contaminating infection or any latent rat-derived *Pneumocystis* organisms that might develop into infection. These rats were also sacrificed at the conclusion of the experiment and no *Pneumocystis* organisms were identified by light microscopy. The rats were screened for *Pneumocystis* DNA by nested PCR amplification of the mt LSU rRNA locus and no *Pneumocystis* DNA was detected. It was assumed therefore that the five simple populations were not the result of latent or external contaminating infection.

6.1.2. *PRT1* Assay

The catalytic domain was chosen for analysis because it is the most characterised region of the *PRT1* gene (Lugli, 1999a, Palmer, 2000a). In the previous chapter, the repetitive nature of the proline rich region limited characterisation of this domain by PCR amplification and genescan analysis. In contrast, there is much distinguishing sequence variation in the catalytic domain, which can be PCR amplified by the universal primers PcProt1 and PcProt3 (Lugli *et al.,* 1997). To date, a product has been obtained from every *PRT1* clone analysed using these primers and the primer sequences are present in all published *PRT1* gene sequences. The amplification product includes two introns of approximately 50 bp each so it is possible to distinguish between cDNA and DNA products. Both genomic DNA and cDNA were amplified so that the expressed cDNA clones could be correlated back to their genomic counterparts. The RNA generated from the simple populations was used as a template for

RT-PCR amplification. The first strand of cDNA was generated with reverse transcriptase using the primer PcProt1 (section 2.12.3.) and in the second step of the PCR amplification synthesis of double stranded DNA an annealing temperature of 57°C was used (Figure 6.1.). In initial experiments, relevant PCR amplification products were cloned and plasmids containing an insert were extracted and sequenced. Later, a PCR screen (section 2.8.) was found to be a more efficient way of identifying plasmids containing an insert. The amplification products from the PCR screen were then sequenced directly.

6.1.3. Clone Summary

Seven rat-derived *Pneumocystis* populations were investigated, the five simple populations EA15, EA17, EA19, EA21 and EA24 and two complex populations, the initial undiluted inoculum and a complex population, 1984, generated from the rat model of *Pneumocystis* infection (Frenkel *et al.*, 1966). Generally, 12 genomic and 12 cDNA clones were analysed from each population. EA21 was analysed twice, (EA21.1 and EA21.2) in order to validate the reproducibility of the assay. In total 199 clones were sequenced and analysed from the seven different populations (Table 6.1.).

A nomenclature was devised for the clones. The four characters describe from which population the clone is derived, for example, ea15 refers to a specific simple population, inoc refers to the inoculum population while 1984 refers to the control complex population. There are two options for the next character, either g for a genomic clone or c for a cDNA clone. Finally, the last two characters are the specific clone name, for example 7a or 1b. The 199 sequences obtained were then aligned using the ClustalW programme.

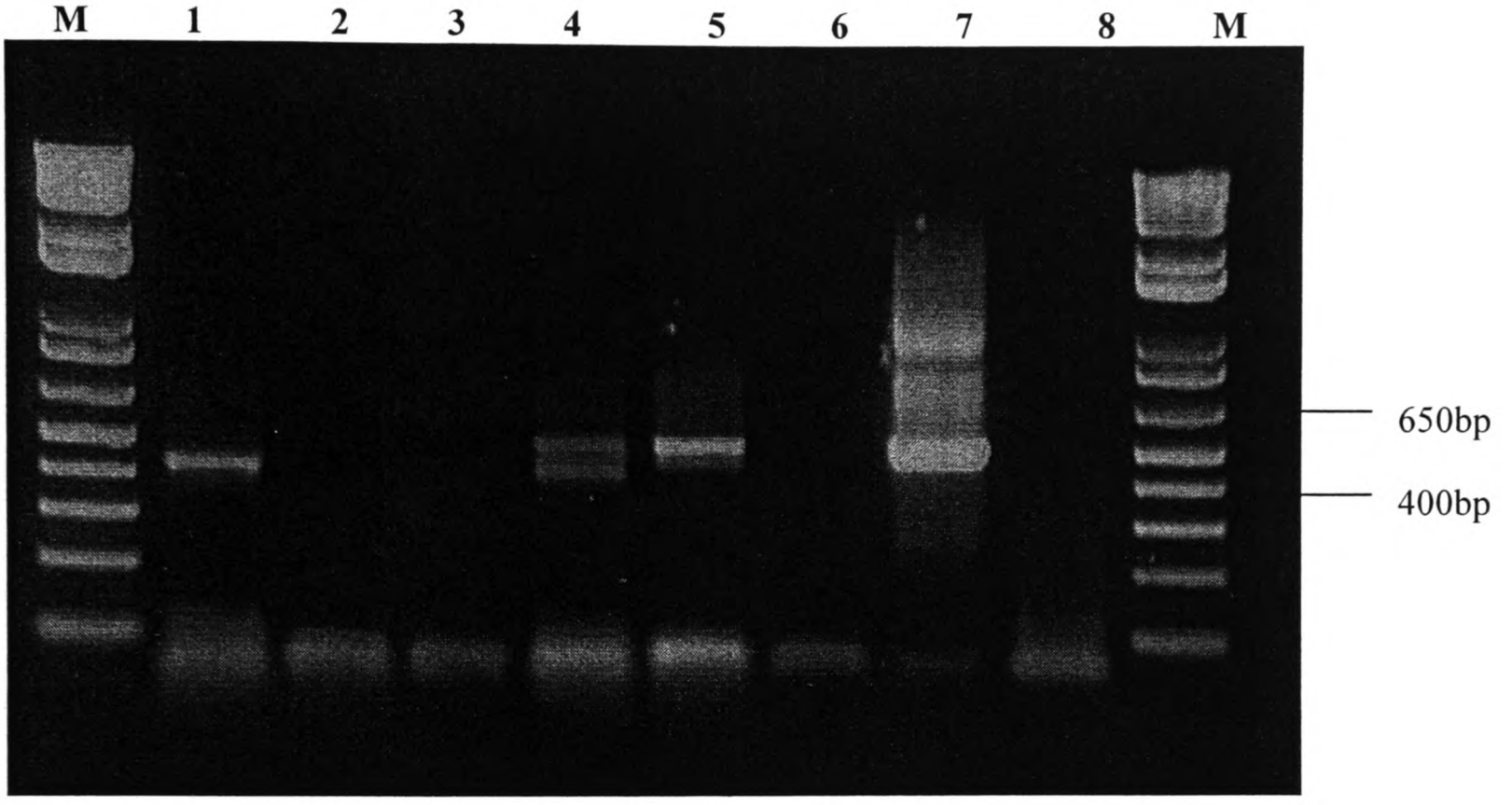


Figure 6.1. Electrophoretic separation of the products of an RT-PCR amplification using the primers PcProt1 and PcProt3. M, 1Kb Plus DNA Ladder. Both RNA and DNA are used from the rat-derived *Pneumocystis* low dose population as indicated. The different templates and conditions are; lane 1: DNase treated EA19 RNA plus RT; lane 2: DNase treated EA19 RNA, minus RT; lane 3: dH_2O in place of a template plus RT; lane 4: untreated EA19 RNA plus RT; lane 5: untreated EA19 RNA minus RT; lane 6: dH_2O in place of a template plus RT; lane 7: genomic EA19 DNA; lane 8 negative control using dH_2O in place of a template. Lane 1 contains the cDNA product with no contaminating DNA. The second lane is the control for contaminating *Pneumocystis* DNA while the third lane is the control for contaminating *Pneumocystis* RNA. The next three lanes are monitoring the efficiency of the DNase treatment. Lane 4 contains both the cDNA and the genomic DNA product, which differ by 2 introns of 50 bp each. The seventh lane contains the genomic DNA product while the eighth lane monitors for *Pneumocystis* DNA contamination during the second strand synthesis.

Table 6.1. Summary of the clones analysed for each population The name of the population is at the top of each box. The left hand column lists the genomic clones while the right hand column lists the cDNA clones. The total number of clones in each list is at the bottom of the column.

EA15	
Genomic	cDNA
ea15g7a	ea15c1b
ea15g7b	ea15c1c
ea15g7c	ea15c1g
ea15g7d	ea15c1h
ea15g7g	ea15c1l
ea15g7h	ea15c1m
ea15g7k	ea15c1s
ea15g7j	ea15c1x
ea15g7n	ea15c1z
ea15g7v	ea15c4a
ea15g7x	ea15c4d
ea15g7y	ea15c4j
	ea15c4m
12	13

EA17	
Genomic	cDNA
ea17g3n	ea17c2b
ea17g3o	ea17c2c
ea17g7a	ea17c2d
ea17g7b	ea17c2e
ea17g7c	ea17c2f
ea17g7e	ea17c2l
ea17g7l	ea17c2o
ea17g8a	ea17c2p
ea17g8b	ea17c2q
ea17g8c	ea17c2cc
ea17g8d	ea17c2ff
ea17g8e	ea17c4h
12	12

EA19	
Genomic	cDNA
ea19g7a	ea19c1b
ea19g7b	ea19c1c
ea19g7c	ea19c1f
ea19g7d	ea19ccd
ea19g7e	ea19cce
ea19g7f	ea19ccf
ea19g7h	ea19ccj
ea19g7i	ea19cck
ea19g7k	ea19ccn
ea19g7l	ea19cco
ea19g7n	ea19ccq
ea19g7o	ea19ccs
ea19g7p	ea19ccu
	ea19ccw
13	14

EA21	
Genomic	cDNA
ea21g7b	ea21c1a
ea21g7c	ea21c1d
ea21g7d	ea21c1e
ea21g7e	ea21c1g
ea21g7j	ea21c1i
ea21g7k	ea21c1j
ea21g7l	ea21c1n
ea21g7m	ea21c1p
ea21g7n	ea21c1v
ea21g7o	ea21c1y
ea21g7p	ea21c1z
	ea21c1ee
	ea21c1jj
11	13

EA21.2	
Genomic	cDNA
ea21.2g7a	ea21.2c21b
ea21.2g7c	ea21.2c21c
ea21.2g7d	ea21.2c21d
ea21.2g7e	ea21.2c21f
ea21.2g7f	ea21.2c21g
ea21.2g7h	ea21.2c21h
ea21.2g7i	ea21.2c21l
ea21.2g7l	ea21.2c21m
ea21.2g7m	ea21.2c21v
ea21.2g7o	ea21.2c24k
ea21.2g7q	ea21.2c24l
ea21.2g7t	
ea21.2g7u	
ea21.2g7x	
14	11

EA24	
Genomic	cDNA
ea24g7c	ea24c4b
ea24g7d	ea24c4c
ea24g7e	ea24c4d
ea24g7f	ea24c4m
ea24g7g	ea24c4t
ea24g7i	ea24c219
ea24g7k	ea24c4117
ea24g7l	ea24c4211
ea24g7s	ea24c4212
ea24g7v	ea24c4223
ea24g7y	ea24c4224
ea24g7ii	
12	11

Inoc	
Genomic	cDNA
inocg7b	inocc1a
inocg7d	inocc1c
inocg7e	inocc1e
inocg7g	inocc1i
inocg7k	inocc1p
inocg7l	inocc1q
inocg7m	inocc1r
inocg7o	inocc1t
inocg7p	inocc4a
inocg7q	inocc4c
inocg7r	inocc4e
inocg7s	inocc4g
	inocc4i
	inocc4j
12	14

1984	
Genomic	cDNA
1984g2a	1984c3b
1984g2b	1984c3c
1984g2c	1984c3d
1984g2d	1984c3e
1984g2e	1984c3f
1984g2f	1984c3h
1984g2h	1984c3i
1984g2i	1984c3j
1984g4d	1984c3p
1984g4f	1984c3r
1984g4g	1984c3s
1984g4h	1984c3t
1984g4i	
13	12

6.1.4. Phylogenetic Trees for Individual Populations

A Neighbour Joining tree (Saitou and Nei, 1987) comparing the catalytic domain sequences was produced for each population from the ClustalW alignment (Figure 6.2 and Figure 6.3). The scale bar represents the branch length, which is equivalent to ten nucleotide changes per one hundred nucleotide residues. In this study, two clones with joining branch lengths shorter than 0.009 were considered to be similar enough to be classed as being identical. This represents less than one nucleotide change per one hundred nucleotide residues between the two clones. The length of the scale bar is different for each tree; therefore, the relatedness of the clones within each tree is different.

The trees demonstrate two main themes. One sequence, the cDNA majority sequence type is present in the populations EA15, EA17, EA19, inoc and 1984. These clones have been highlighted in yellow. A second sequence, the genomic majority sequence type, is present in all eight populations. This sequence has no cDNA equivalent clones. These genomic clones have been highlighted in green.

To summarise the 199 clones, there are 99 genomic clones and 100 cDNA clones. There are 21 copies of the cDNA majority sequence type, 3 genomic clones and 18 cDNA clones while there are 17 copies of the genomic majority sequence type. With the pairs of two identical clones from the individual phylogenetic trees for each population, there are 7 pairs of two cDNA clones, 5 pairs of two genomic clones and 8 pairs of a genomic and a cDNA clone. Finally, from the individual phylogenetic trees, there are 11 clusters of related sequences.

Divergence analysis, using MEGA version 2.1. (Kumar *et al.*, 2001), was used to determine the amount of diversity within a population (section 2.9.3). Here, it reflects the lack of distinction between the sequences of the simple and complex populations. As a control,

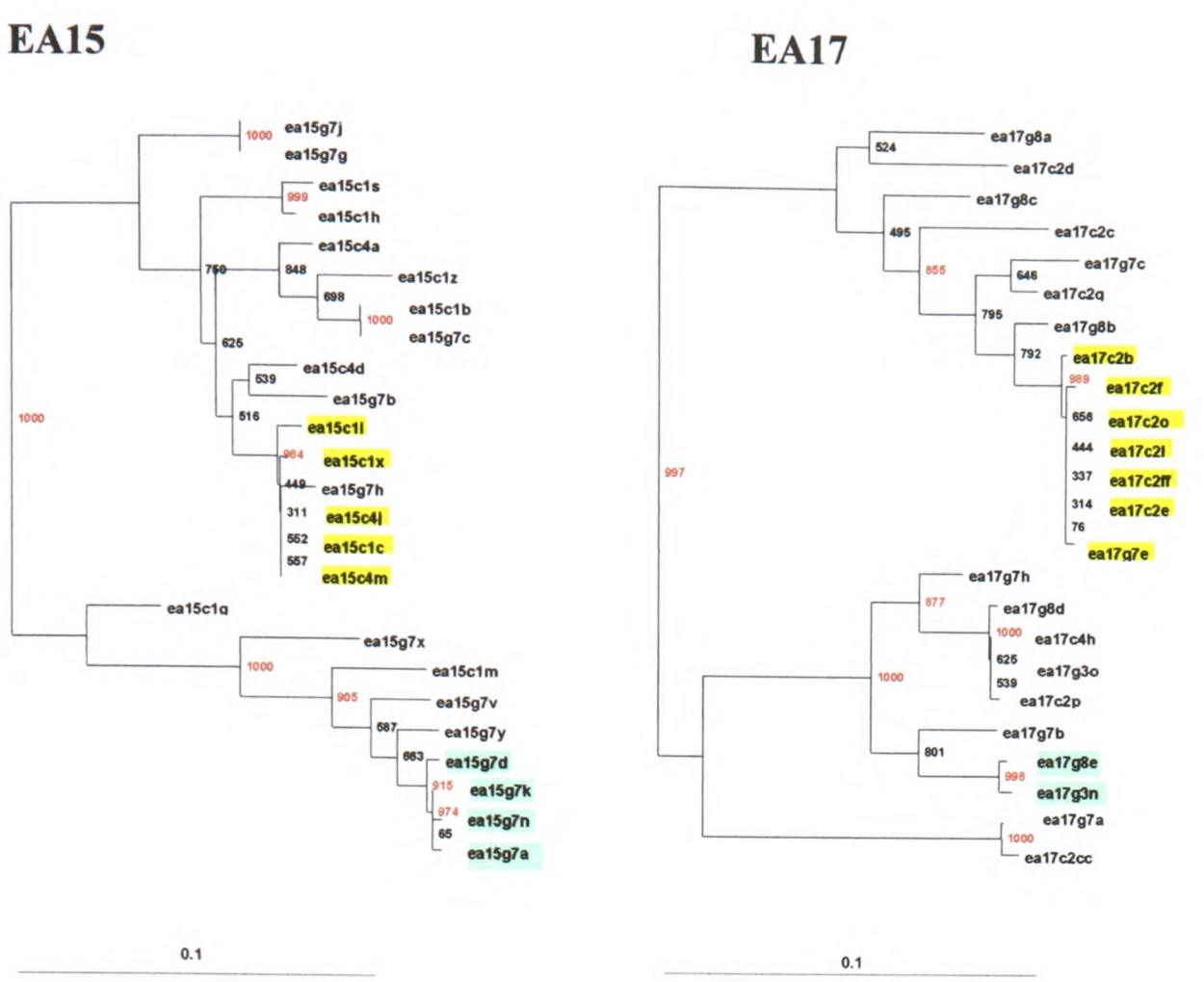
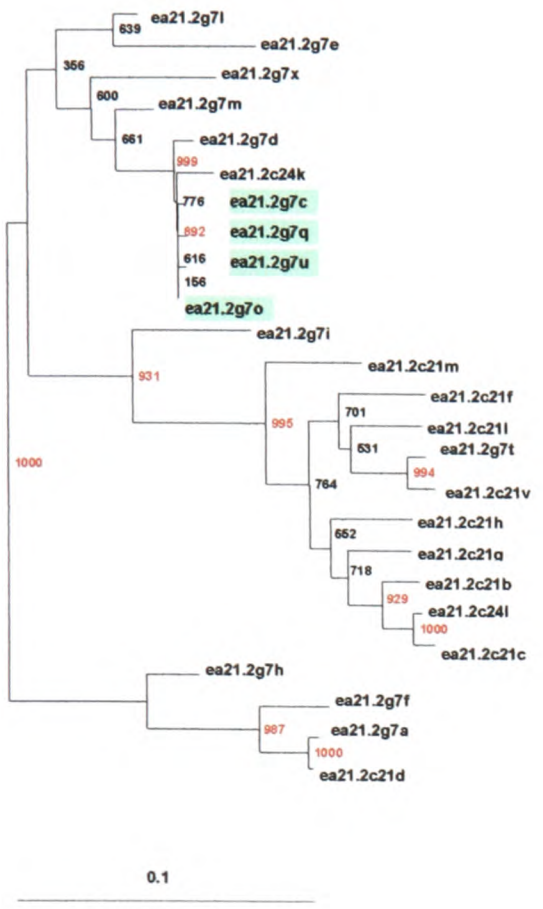
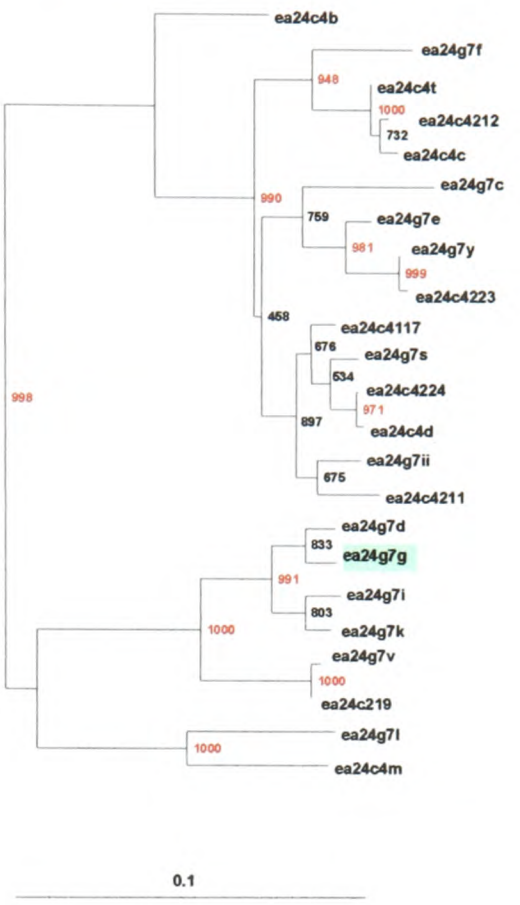
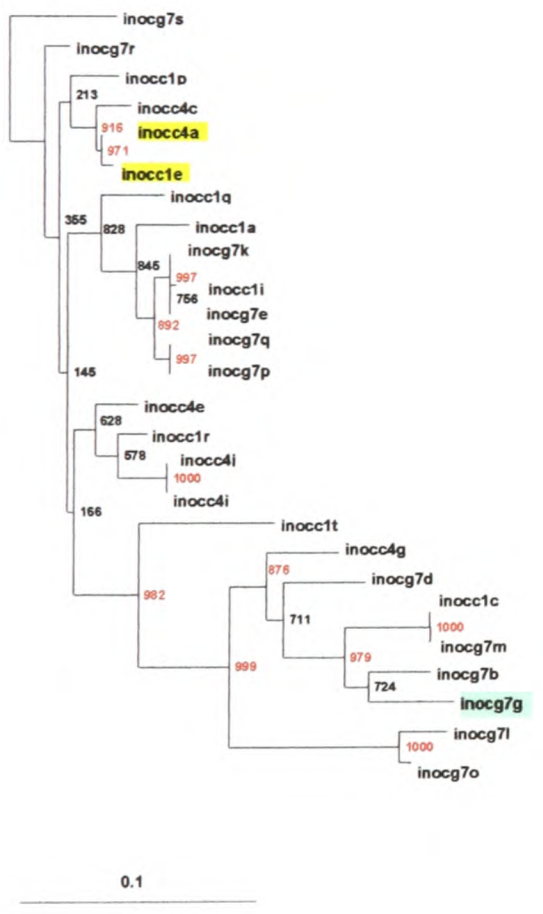


Figure 6.2. Individual phylogenetic trees for *PRT1* sequences from populations EA15, EA17, EA19 and EA21.1. They are derived by the ClustalW Neighbour-Joining Method. cDNA majority sequence type clones are highlighted in yellow and genomic majority sequence type clones are highlighted in green. Bootstrap values are placed at the nodes of the branches and significant values are coloured red. A bootstrap value is significant if it is higher than 850. The scale bar represents the branch length that is equivalent to 10 nucleotide changes per 100 nucleotide residues.







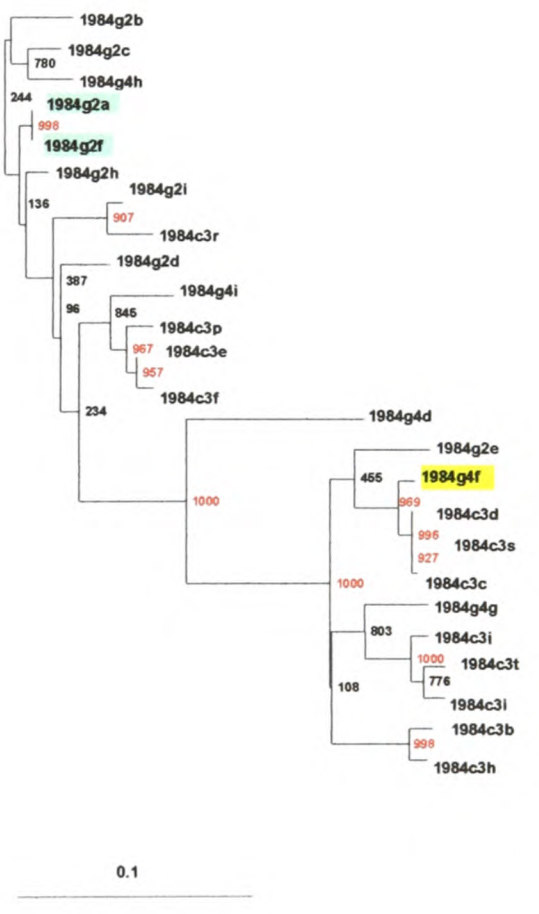


Figure 6.3. Individual phylogenetic trees for *PRT1* sequences from population EA21.2, EA24, inoculum and 1984. They are derived by the ClustalW Neighbour-Joining Method. cDNA majority sequence type are highlighted in yellow and genomic majority sequence type clones are highlighted in green. Bootstrap values are placed at the nodes of the branches and significant values are coloured red. A bootstrap value is significant if it is higher than 850. The scale bar represents the branch length that is equivalent to 10 nucleotide changes per 100 nucleotide residues.

twelve random clones were selected from the total 199 clones. This group of sequences had an overall average divergence of 15.2%. The population 1984 had the same value of 15.2% and EA21.1 had a similar value of 15.7%. This implies that EA21.1 has the same level of complexity as 1984 and the control group. It was only in the populations that contained many copies of the cDNA majority sequence type that there was a lower level of divergence, for example EA15 was 13.5% divergent and EA17 was 12.2% divergent.

6.1.5. The cDNA majority sequence type

There were 21 copies of the cDNA majority sequence type; 18 are cDNA clones and 3 are genomic clones, which were analysed by the construction of a phylogenetic tree (Figure 6.4). They were all identical, being defined as having less than 1 nucleotide change per hundred nucleotide residues between them (as described above). The tree of the translated sequences is also on figure 6.4. and appears to be more divergent than the nucleotide tree. This is because approximately 400 nucleotide residues are being compared with approximately 135 amino acid residues. One nucleotide change represents a much smaller change of 1/400, compared with one amino acid change of 1/135. Most of the nucleotide differences were also present in these translated trees so these changes are non-synonymous, (section 2.9.4). SNAP analysis calculates the ratio of synonymous, (ds) and non-synonymous, (dn) changes i.e. ds/dn. Ds/dn ratios are a way of analysing the amount and type of selection a coding region is subject to. As the sequences are so alike, it is not feasible to use the SNAP programme (Nei and Gojobori, 1986) so the nucleotide and amino acid sequences must be visually compared. There were 3 synonymous changes and 17 non-synonymous changes between the nucleotide sequences in the alignment. This gave a ratio of 3:17 or a ds/dn ratio of 0.176. This implies that any changes that have occurred in the cDNA majority sequence type were the result of positive selection.

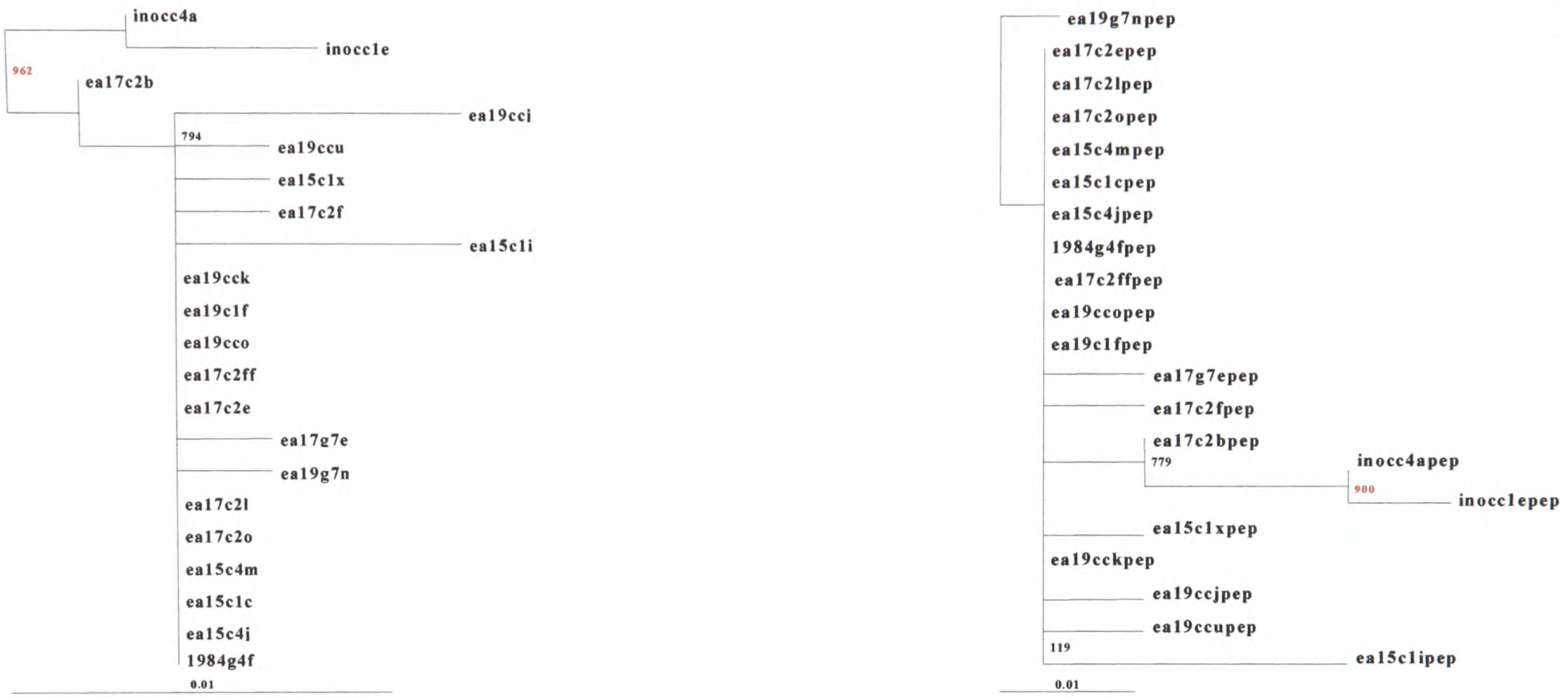


Figure 6.4 Phylogenetic trees for the copies of the *PRT1* cDNA majority sequence type. They are derived by the ClustalW Neighbour-Joining Method. The nucleotide sequence tree is on the left and the amino acid sequence tree is on the right. Bootstrap values are placed at the nodes of the branches and significant values (>850) are coloured red. The scale bar represents the branch length that is equivalent to 1 nucleotide change per 100 nucleotide residues.

6.1.6. Published Catalytic Domain Sequences

The sequences generated within this project were compared with the ten published full length *PRT1* gene sequences (section 1.4.3). The published sequences are from different populations and the majority are from different labs. The phylogenetic tree of the catalytic domain of the published *PRT1* gene sequences is dicotymous (data not shown) with the majority of the sequences being unique, apart from 1b22, *ORF-3* and paga, which are identical because they are all the genomic majority sequence type. None of the published sequences correlated with the cDNA majority sequence type.

6.1.7. Genomic Majority Sequence Types

Seventeen copies of the genomic majority sequence type were identified in this study. A phylogenetic tree was constructed to compare them with three of published sequences, which are the genomic majority sequence type as described above (data not shown). There are no equivalent cDNA copies. As with the cDNA majority sequence type, the clones are all identical. The amino acid phylogenetic tree is less divergent than the nucleotide tree. From a visual comparison of the 13 nucleotide changes of the nucleotide sequences in the alignment, 8 were synonymous changes compared to 5 non-synonymous changes. This gave a ds/dn ratio of 1.6, suggesting that there is no positive selective pressure on the genomic majority sequence type.

6.1.8. Phylogenetic Analysis of EA21.1 and EA21.2 and Different Combinations of Sequences from the 209 Clones

Population EA21 was analysed twice (EA21.1 and EA21.2) to see if the analysis was reproducible. A phylogenetic tree was constructed to compare the two analyses of the population (Figure 6.5.). 46.9% of clones were identified in both analyses, for example EA211c1n and EA212c21d are identical while 53.1% of clones were only detected in one of

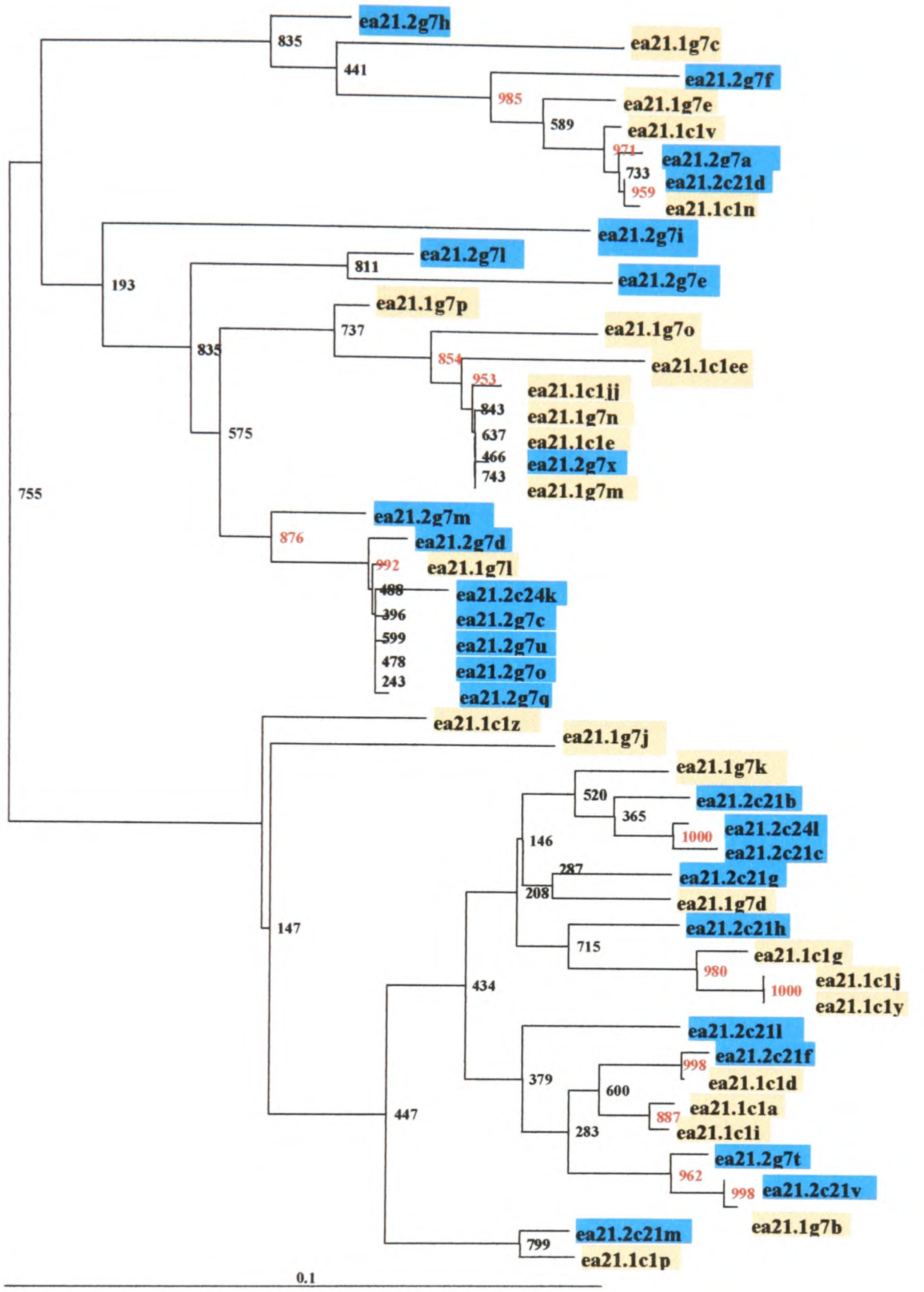


Figure 6.5. Phylogenetic tree comparing *PRT1* sequences from the two assays of populations EA21, EA21.1 and EA21.2. They are derived by the ClustalW Neighbour-Joining Method. EA21.1 clones are marked in beige while EA21.2 clones are marked in blue. Bootstrap values are placed at the nodes of the branches and significant values (>850) are coloured red. The scale bar represents the branch length that is equivalent to 10 nucleotide changes per 100 nucleotide residues.

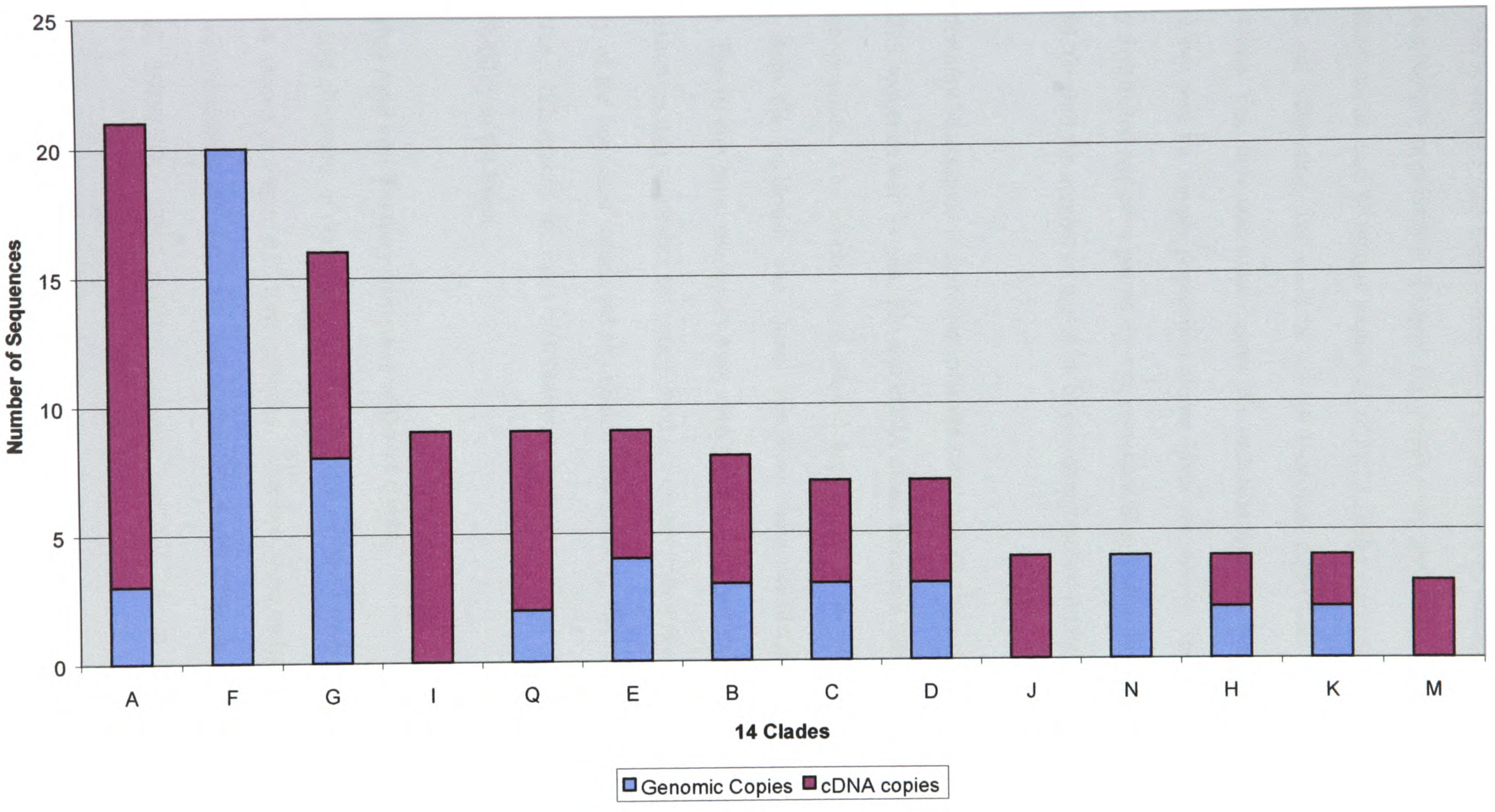
the analyses. Generally different clones were produced in the two experiments, because a different subset of the PRR alleles present in the population was analysed each time.

Nucleotide sequences from all 209 clones, 199 clones generated from this study plus the ten published sequences, were aligned and a phylogenetic tree was drawn, the 209 sequence tree (data not shown). The tree structure consisted of four main groups. There was an uneven distribution of the 209 clones, with the genomic clones being predominant in two of the groups while cDNA clones were mainly present in the other two groups. The seven populations are evenly distributed throughout the tree, demonstrating that the clones are ubiquitous throughout the populations.

Within the tree, there are 14 subgroups or clades (Page and Holmes, 1998). A clade is a group of sequences that are clearly distinguishable from other sequences by the discrete structure of their joining branches within the overall tree. From the scale of the 209 sequence tree, sequences were defined as being part of a clade if there were 0.6 or less nucleotide differences per 100 nucleotides between them. The clades are labelled A-N, in the order in which they descend down the 209 sequence tree (Figure 6.6). Both the cDNA majority sequence type and the genomic majority sequence type form two of these clades, A and F respectively. Some clades only contain cDNA or genomic clones whereas the seven populations are scattered amongst the clades.

The 209 clones were then divided into cDNA and genomic clones and the sequences aligned. The resulting trees (data not shown) both produced a similar pattern to the larger 209 tree of 4 groups. However, the two trees were distinct from each other and did not have the same structure or pattern. The 14 clades could also be found evenly distributed throughout the two trees.

Figure 6.6. Frequency Distribution of PRT1 Catalytic Domain Sequences



Theoretically, all of the clones from the simple populations originated from the inoculum population. So, if all the *PRT1* alleles from the inoculum could be identified they would be present in the simple populations. Twenty six clones were analysed from the inoculum. Assuming there are at least 30 unique sequences of *PRT1* alleles per organism (Lugli *et al.*, 1997), it is not surprising that there is not an inoculum equivalent for every simple population clone. Yet, there was some degree of correlation. In a tree consisting of only the inoculum clones and the simple population clones (data not shown), the inoculum clones were evenly distributed and often paired up with corresponding clones. However, the same is true for 1984. Overall the number of inoculum clones characterised limits the analysis.

The 14 clades may be assumed to represent different subgroups of alleles that form the frame of the big 209 sequence tree. A genomic and cDNA clone sequence were taken from each clade where possible. The clones were aligned together with all 10 published *PRT1* sequences. Both the nucleotide and amino acid trees were calculated (Figure 6.7 and Figure 6.8). The 10 published sequences were evenly distributed; yet U82999, U62910 and 3G5 had distinct enough sequences not to be placed in a cluster with related sequences. Also, the majority of the bootstrap values are significant, strengthening the validity of the trees. These 14 clades had exactly the same distribution in both trees. This is further evidence of the reproducibility of the trees.

6.1.9. Amino Acid and Tertiary Structure of the 14 Clades

The amino acid alignment of the sequences representative of the I4 clades and 10 published sequences is shown in Figure 6.9. Four protease sequences were included, kexin 2 from *Saccaromyces cerevisiae,* human furin and human PC1 and PC3 protease (Steiner, 1998, Fuller *et al.,* 1989). The cDNA majority sequence type (A) has no discernable unique features

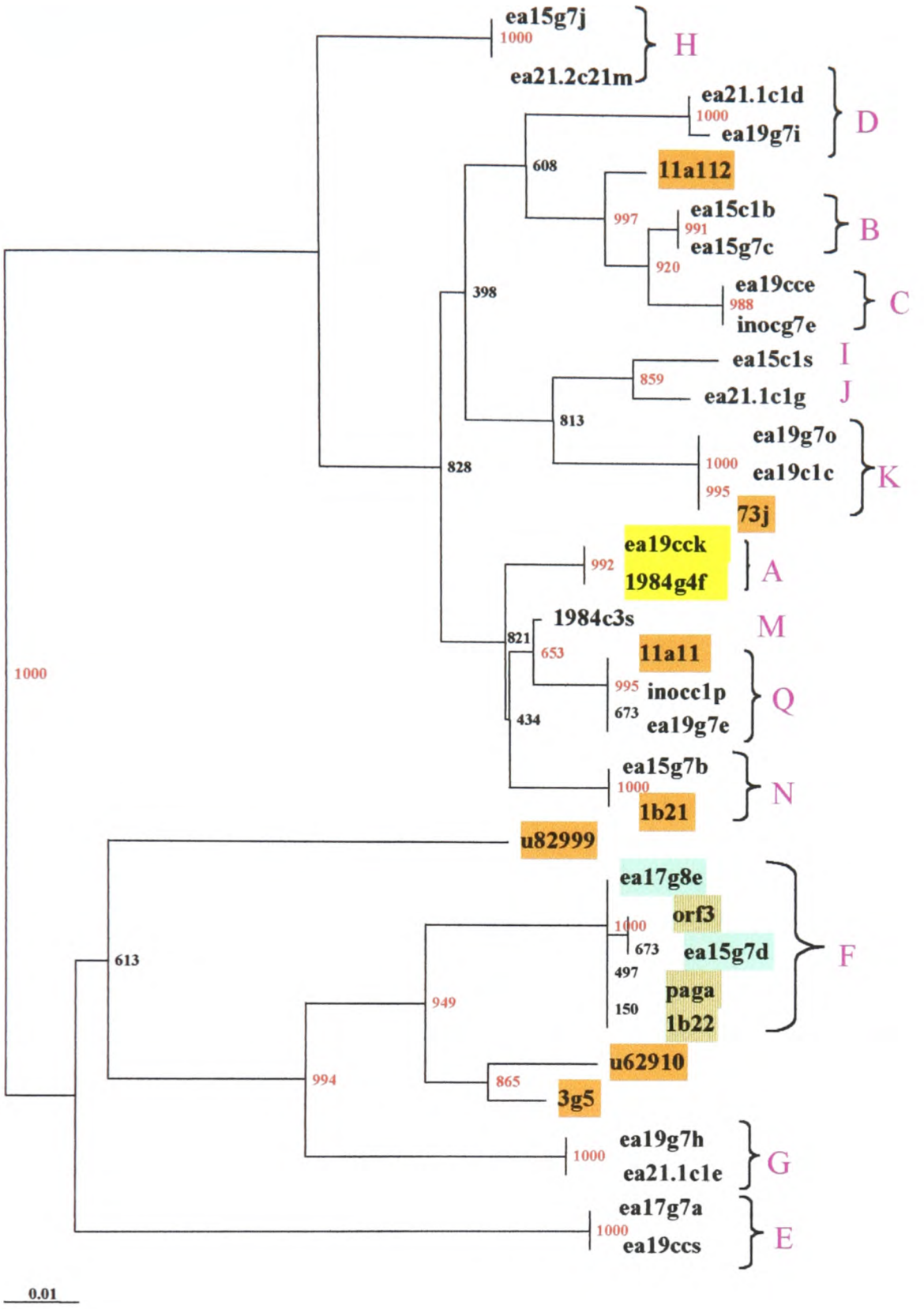


Figure 6.7. Phylogenetic tree from the nucleotide sequence alignment of the 14 clades. It is derived by the ClustalW Neighbour-Joining Method. Published sequences are coloured in orange. cDNA majority sequence type clones are marked in yellow and genomic majority sequence type clones are marked in green. Published sequences that are also genomic majority sequence type clones are striped in both green and orange. The clade letters are in pink by the relevant sequences. Bootstrap values are placed at the nodes of the branches and significant values (>850) are in red. The scale bar represents the branch length equivalent to 1 nucleotide change per 100 nucleotide residues.

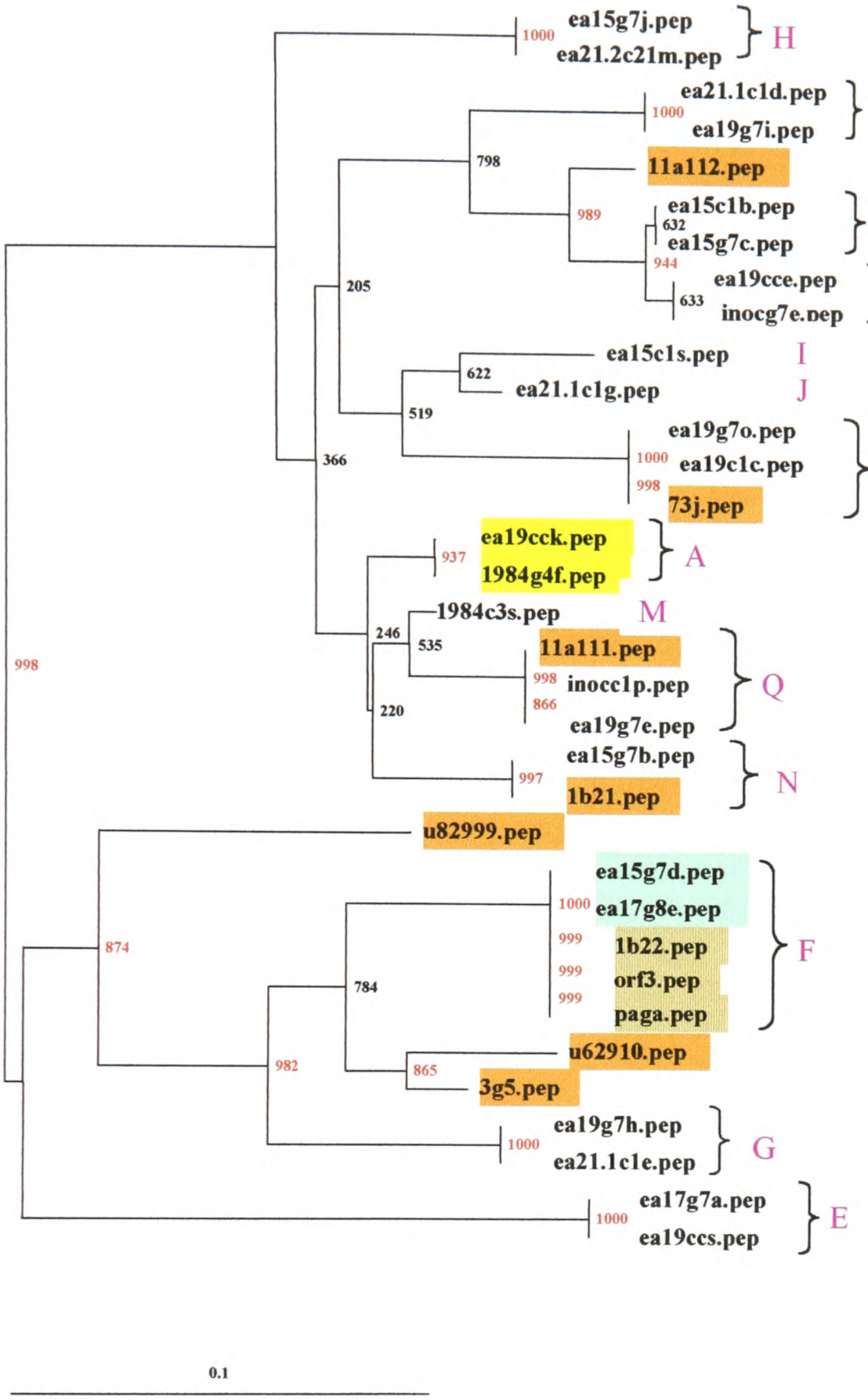


Figure 6.8. Phylogenetic tree from the amino acid sequence alignment of the 14 clades. It is derived by the ClustalW Neighbour-Joining Method. Published sequences are coloured in orange. cDNA majority sequence type clones are marked in yellow and genomic majority sequence type clones are marked in green. Published sequences that are also genomic majority sequence type clones are striped in both green and orange. The clade letters are in pink by the relevant sequences. Bootstrap values are placed at the nodes of the branches and significant values (>850) are in red. The scale bar represents the branch length equivalent to 10 nucleotide changes per 100 nucleotide residues.

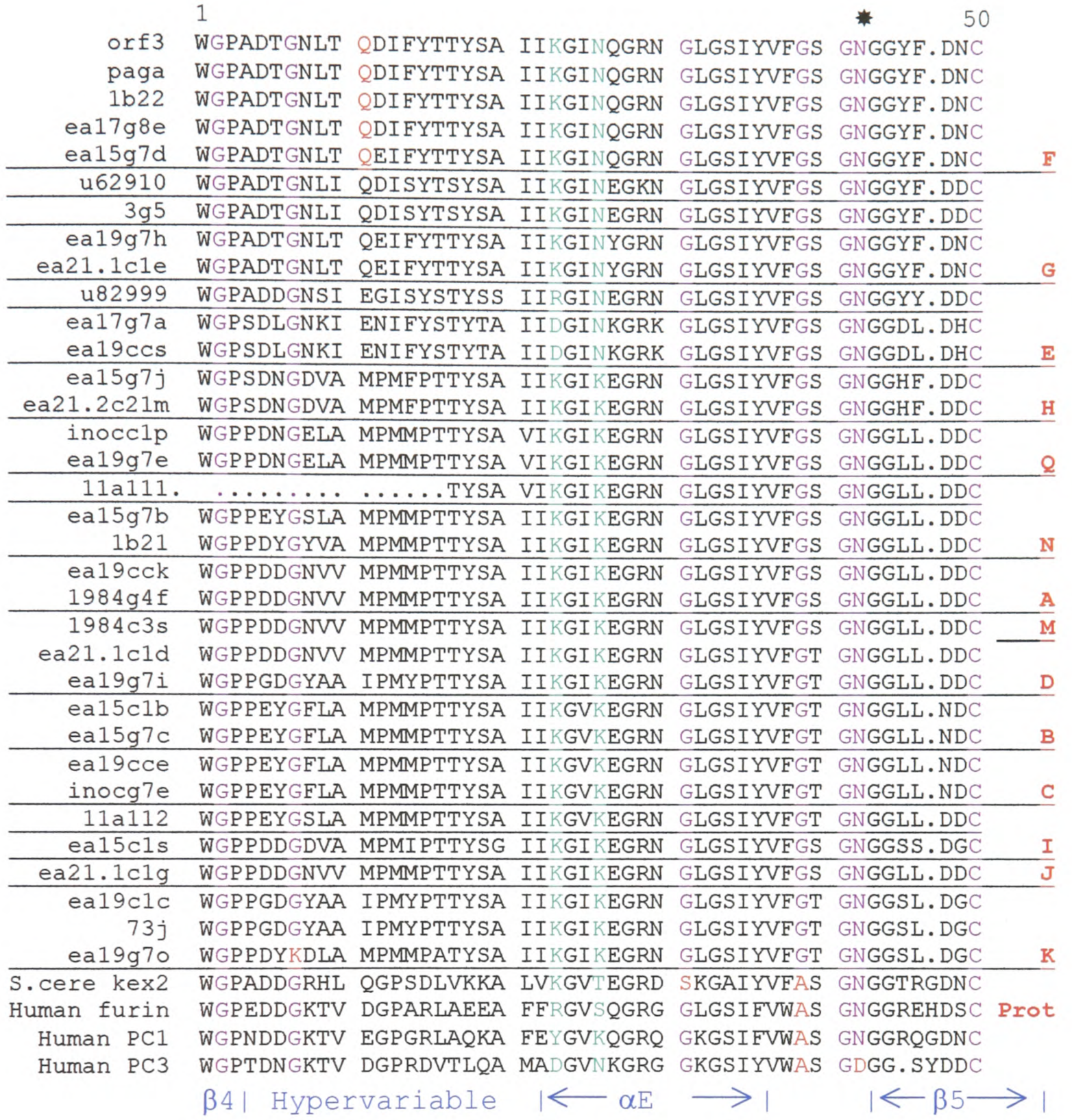


Figure 6.9. Amino acid alignment of different sequences of *PRT1* catalytic domain compared with other protease sequences. The alignment consists of the 10 published protease sequences, the representative sequences of the 14 clades and 4 protease sequences, kexin 2 from *Saccaromyces cerevisiae*, human furin and human PC1 and PC3 protease. The letters on the right label the sequences of each of the 14 clades. The secondary structure is marked in blue at the bottom; β - B-sheet and α - α-helix, (Lugli *et al.*, 1997). The residues marked in purple are conserved residues. N42 forms the oxyanion hole and S111 is part of the catalytic triad, both are marked ✸. Sites that have both acidic and basic residues are marked in green. Residues unique to that particular allele or site are coloured red. Finally there is one unique motif that is indicated in blue and magenta.

```
                                                                  100
      orf3  NYDGYANSPY TITIAAIDAE EKRFIFSEPC PCILASTYSG KR..GA..YI
      paga  NYDGYANSPY TITIAAIDAE EKRFIFSEPC PCILASTYSG KR..GA..YI
      1b22  NYDGYANSPY TITIAAIDAE EKRFIFSEPC PCILASTYSG KR..GA..YI
    ea17g8e NYDGYANSPY TITIAAIDAE EKRFIFSEPC PCILASTYSG KR..GA..YI
    ea15g7d NYDGYANSPY TITIAAIDAE EKRFIFSEPC PCILASTYSG KR..GA..YI  F
      u62910 NYDGYANSPY TITIAAIDAK EERFMFSEPC PCILASTYSG RR..GA..YI
       3g5  NYDGYANSPY TITIAAIDAE EKRFRFSEPC PCILASTYSG KS..GA..YI
    ea19g7h NYDGYANSPY TITIAAIDGD EKRFRFSEPC PCILASTYSG KL..GA..YI
   ea21.1c1e NYDGYANSPY TITIAAIDGD EKRFRFSEPC PCILASTYSG KL..GA..YI  G
      u82999 NYDGYANSPY TITIAAIDAE EKRFIFSEPC PCVLASTYSG GS..NG..RI
    ea17g7a NYDGYANSPY TITIAAINAR ERKFRFSEPC PCILASTYSG GE..SG..RI
    ea19ccs NYDGYANSPY TITIAAINAR ERKFRFSEPC PCILASTYSG GE..SG..RI  E
    ea15g7j NYDGYANSPY TVTIASIDSE DKNLYFSESC PCILASTYSG GE..KG..YI
   ea21.2c21m NYDGYANSPY TVTIASIDSE DKNLYFSESC PCILASTYSG GE..KG..YI  H
    inocc1p NYDGYANSPY TVTIGAIDSE DKNFYFSESC PCILASTYSG GE..NE..SI
    ea19g7e NYDGYANSPY TVTIGAIDSE DKNFYFSESC PCILASTYSG GE..NE..SI  Q
     11a111 NYDGYANSPY TVTIGAIDSE DKNFYFSESC PCILASTYSG GE..NE..SI
    ea15g7b NYDGYANSPY TVTIAAIDAE DKKFDTSESC PCILASTYSG GE..NG..SI
      1b21  NYDGYANSPY TVTIAAIDAE DKKFDTSESC PCILASTYSG GE..NG..SI  N
    ea19cck NYDPYANSPY TITIAAIDSE DKNFYFSESC PCILASTYSG GE..NG..SI
    1984g4f NYDPYANSPY TITIAAIDSE DKNFYFSESC PCILASTYSG GE..NG..SI  A
    1984c3s NYDGYANSPY TVTIGAIDSE DKNFYFSESC PCILASTYSG GE..NG..SI  M
   ea21.1c1d NYDGYANSPY TITIAAINSE DERPYFSESC PCILASTYSG GG..NG..SI
    ea19g7i NYDGYANSPY TITIAAINSE DERPYFSESC PCILASTYSG GG..NG..SI  D
    ea15c1b NYDGYANSPY TITIAAINSE DERPYFSESC PCILASTYSG GG..NG..SI
    ea15g7c NYDGYANSPY TITIAAINSE DERPYFSESC PCILASTYSG GG..NG..SI  B
    ea19cce NYDGYANSPY TITIAAINSE DERPYFSESC PCILASTYSG GG..NG..SI
    inocg7e NYDGYANSPY TITIAAINSE DERPYFSESC PCILASTYSG GG..NG..SI  C
     11a112 NYDGYANSPY TITIGAINSE DERPYFSESC PCILASTYSG GG..NG..SI
    ea15c1s NYDGYANSPY TITIAAIDSE DKSFYFSESC PCILASTYSS GE..NG..SI  I
   ea21.1c1g NYDPYANSPY TITIAAIDSE DKSFYFSESC PCILASTYSS GE..NG..SI  J
    ea19c1c NYDGYANSPY TITIAAIDSE DKNFYFSESC PCILASTYSG GE..NG..SI
       73j  NYDGYANSPY TITIAAIDSE DKNFYFSESC PCILASTYSG GE..NG..SI
    ea19g7o NYDGYANSPY TITIAAIDSE DKNFYFSESC PCILASTYSG GE..NG..SI  K
S.cere kex2 NYDGYTNSIY SITIGAIDHK DLHPPYSEGC SAVMAVTYSS GS..GE..YI
Human furin NCDGYTNSIY TLSISSATQF GNVPWYSEAC SSTLATTYSS GN..QNEKQI Prot
  Human PC1 DCDGYTDSIY TISISSASQQ GLSPWYAEKC SSTLATSYSG GD..YTDQRI
  Human PC3 NCDGYASSMW TISINSAIND GRTALYDESC SSTLASTFSN GRKRNPEAGV
                  ← β6 →|                 | β7 |
```

Figure 6.9. Cont.

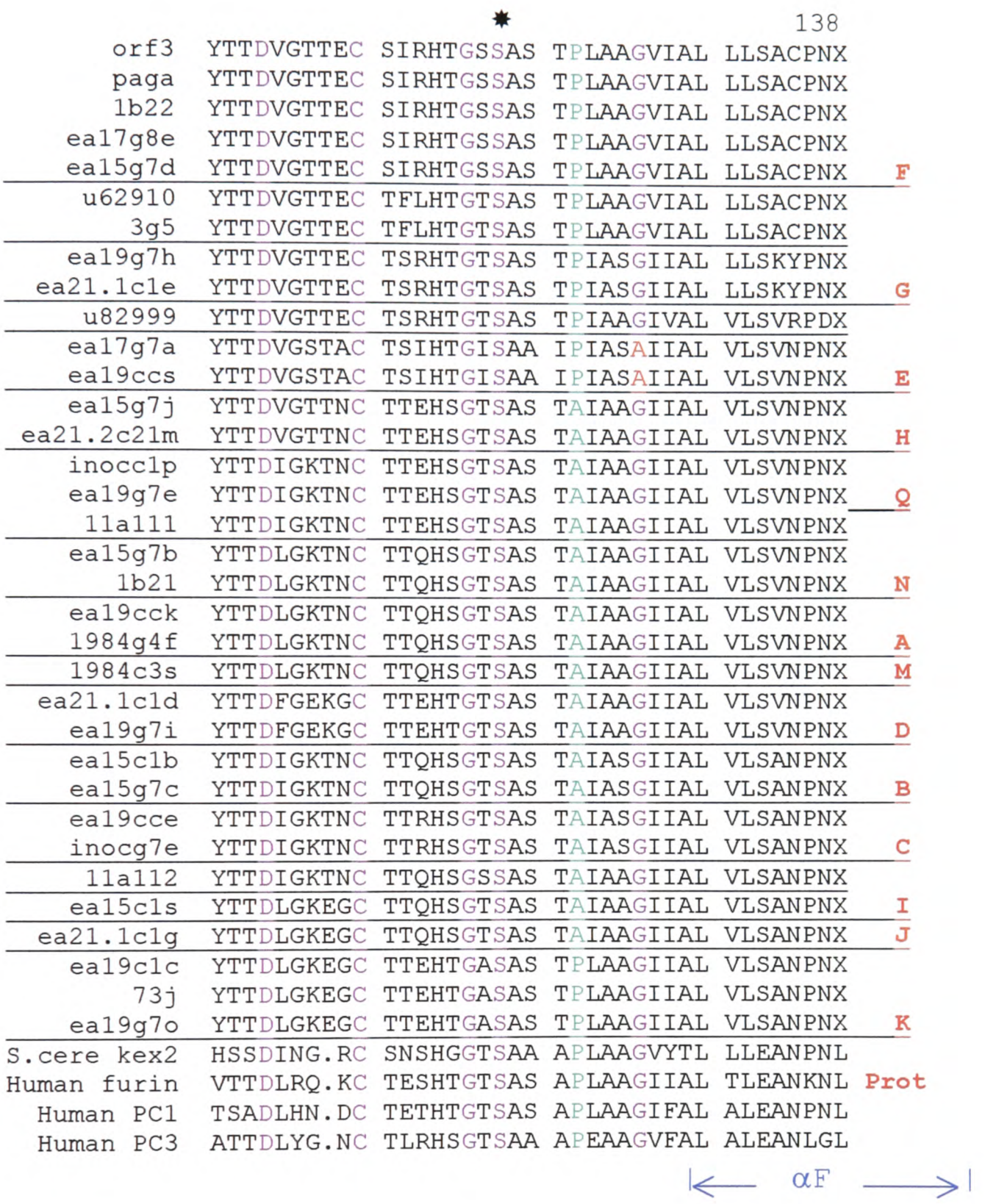


Figure 6.9. Cont.

apart from a proline residue instead of a conserved glycine residue, at residue 54, using the numbering of Figure 6.9. This could have structural implications because proline is an inimo acid rather than an amino acid and this may induce kinks in the polypeptide chain, which could affect the folding of the protein.

The genomic majority sequence type (F) is more unusual than the cDNA majority sequence type. At residue 11, it has glutamine while the majority of other sequences have a methionine. This occurs within the hypervariable region, at the beginning of the amplified region of the catalytic domain. At residue 74, the genomic majority sequence type has an isoleucine while most other sequences have a tyrosine. These two residues have very different influences upon protein folding. One motif, which makes the genomic majority sequence type unique, KRGAY, is described in chapter 4. This motif is in contrast to other combinations at this position for example GENES or GSNGR. GENES is highly acidic while KRGAY is basic. Also, KRGAY contains two dibasic residues, KR, which have the potential to be an autocatalytic site. This could possibly render the gene dysfunctional. The genomic majority sequence type encodes for the KRGAY motif while the cDNA majority sequence type encodes for the acidic GENGS motif.

The majority of the conserved residues (Lugli *et al.,* 1997) were present in all the sequences, apart from G39 and C82 that are A39 and S82 in the non-PRT1 sequences, (apart from kex2 that is A82). One of the three essential residues of the catalytic domain, serine, is present in all the sequences and the asparagine from the oxyanion hole is present in all sequences apart from human PC3. The motif of interest is KRGAY, at residue 90. Kexin 2 has GSGEY that even though it is unique in this alignment, lines up with the PRT1 sequences. Human furin and human PC1 have two additional amino acids inserted at this site while human PC3 has

four, including KR. The dibasic residues do not act as a cleavage site in PC3 (Steiner, 1998), so they might not be cleaved in PRT1.

6.1.10. SNAP analysis

By comparing non-synonymous versus synonymous changes within the sequence alignment of the seven populations, SNAP analysis (Nei and Gojobori, 1986) determined if the populations were under selective pressure (Table 6.2.). The 14 clade alignment was included as a population, which consisted of a full range of *PRT1* sequences. The average ds/dn ratios for the populations ranged from 0.67 to 1.64, showing that some of the populations by the definition described in section 2.9.4. were under selective pressure while others were not. The results from the SNAP analysis can be presented as a scatter graph. The 14 clade population has been shown as an example, (Figure 6.10(a)). Most of the populations have two separate groups within these scatter graphs. An explanation for the two groups would be the division of genomic and cDNA clones. However, when the populations are divided into cDNA and genomic clones, the SNAP analysis still results in two subgroups.

SNAP analysis was calculated for each subgroup to define what was causing the subdivision. Figure 6.10.(b) and (c) show the results for the uppermost and lowermost subgroups respectively of the 14 clade population. The uppermost group is evenly scattered while the lowermost group appeared to have a further subdivision. Divergence analysis carried out on the uppermost group was 10.0% divergent, while the lowermost group was 6.4% divergent. These low numbers reflect the high level of homology within the subgroups. Further, the ds/dn ratios are striking. The whole population has a ds/dn ratio of 0.96, while the uppermost group has a ds/dn ratio of 0.67, which is in contrast to the lowermost group, which has a ds/dn ratio of 1.64. If the division in a subgroup is the result of the clone sequence, this division should be apparent in the phylogenetic trees. Looking at the sequence alignments of

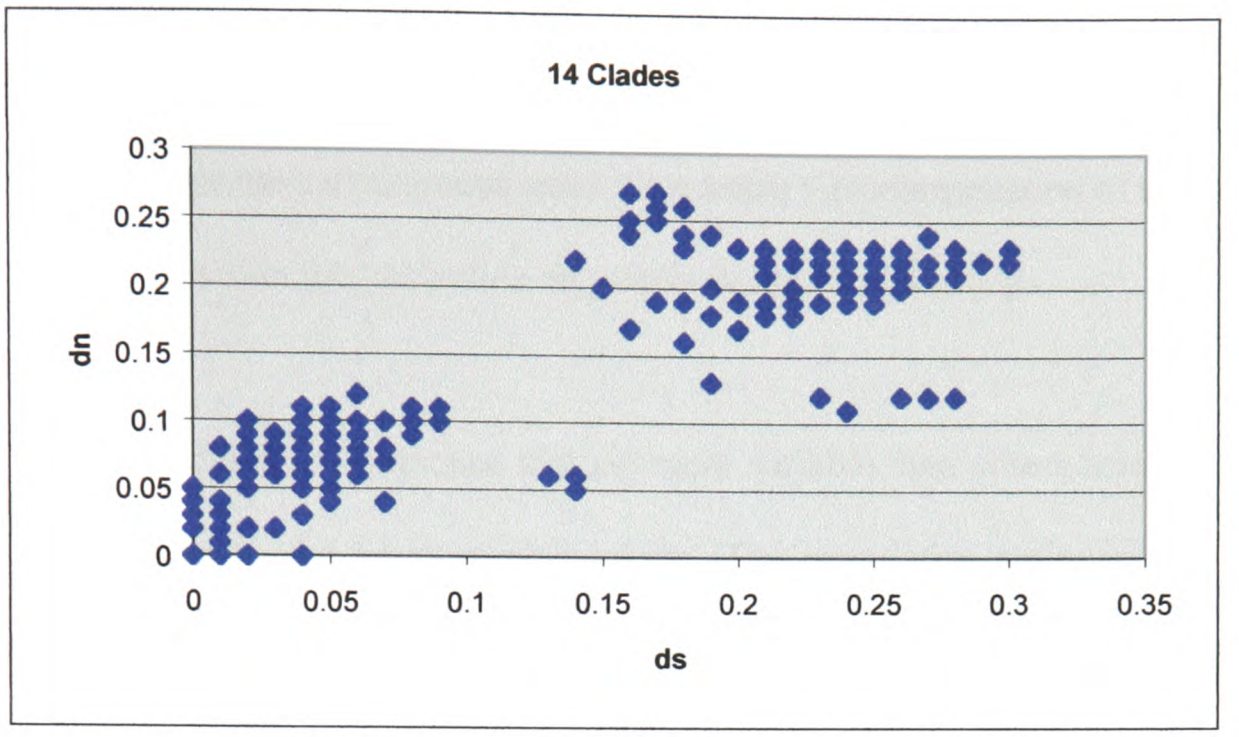
Table 6.2. Averages for ds, dn and ds/dn values and interquartile ranges for ds, dn and ds/dn for the simple populations.

Population	Average			Interquartile Ranges*		
	ds	dn	ds/dn	ds	dn	ds/dn
EA15gc	0.15	0.14	1.12	0.04 - 0.24	0.06 - 0.22	0.72 - 1.23
EA17gc	0.15	0.17	0.91	0.05 - 0.22	0.07 - 0.22	0.66 - 1.07
EA19gc	0.14	0.15	0.93	0.04 - 0.22	0.07 - 0.22	0.60 - 1.12
EA21gc	0.14	0.17	0.87	0.07 - 0.19	0.09 - 0.23	0.65 - 0.99
EA21.2gc	0.17	0.16	1.09	0.08 - 0.23	0.09 - 0.22	0.78 - 1.23
EA24gc	0.14	0.16	0.92	0.05 - 0.23	0.08 - 0.23	0.64 - 1.11
1984gc	0.16	0.15	1.14	0.06 - 0.24	0.07 - 0.22	0.86 - 1.23
inocgc	0.13	0.15	0.86	0.06 - 0.20	0.08 - 0.22	0.60 - 1.12
EA17c	0.14	0.17	0.73	0.00 - 0.20	0.01 - 0.23	0.61 - 1.03
EA17g	0.16	0.16	1.06	0.07 - 0.22	0.07 - 0.23	0.79 - 1.12
EA24c	0.10	0.13	0.91	0.05 - 0.17	0.07 - 0.22	0.64 - 0.83
EA24g	0.17	0.17	1.02	0.06 - 0.24	0.07 - 0.23	0.77 - 1.16
1984g	0.17	0.17	1.02	0.03 - 0.24	0.07 - 0.22	0.70 - 1.21
1984c	0.15	0.15	1.15	0.06 - 0.24	0.07 - 0.23	0.77 - 1.16
14 clades	0.15	0.15	0.96	0.04 - 0.24	0.07 - 0.22	0.62 - 1.18
14 clades uppermost group	0.04	0.08	0.67	0.02 - 0.05	0.06 - 0.09	0.4 - 0.78
14 clades bottom group	0.15	0.11	1.64	0.05 - 0.21	0.04 - 0.17	1.07 - 2.39

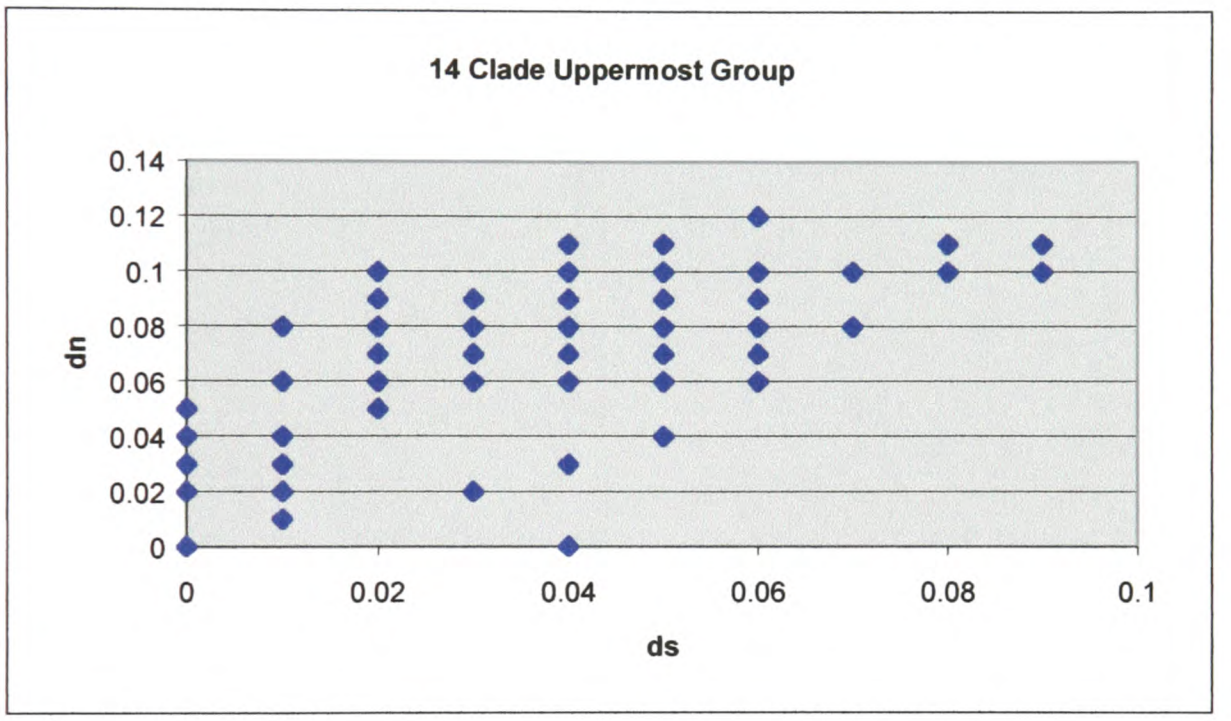
g = genomic clones
c = cDNA clones

* Interquartile ranges = the range of values from the 25th percentile to the 75th percentile

(a)



(b)



(c)

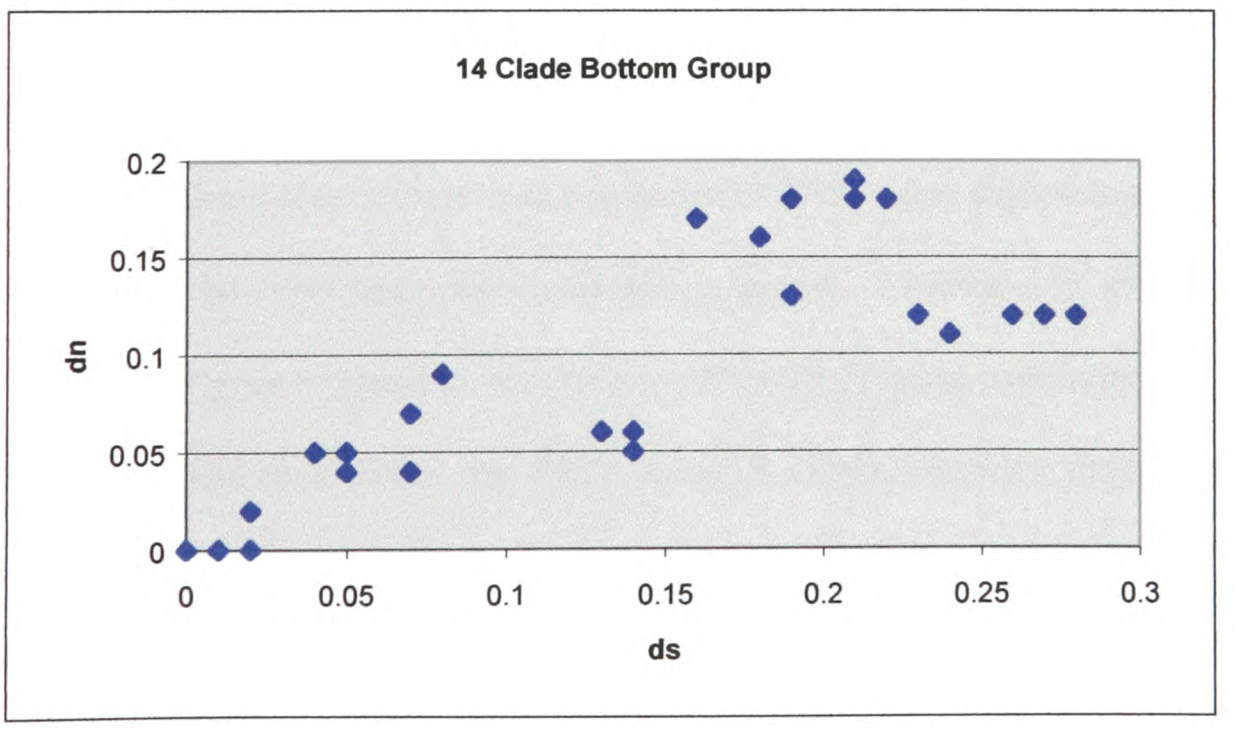


Figure 6.10. Scatter graphs of results of SNAP analysis for (a) all clones from 14 clades, (b) the clones from the uppermost group of the 14 clades, (c) and the clones from the bottom group of the 14 clades.

the whole populations, especially the hypervariable region, there was a natural division of sequences within the alignment that correlated with the subgroups. Therefore populations with two subgroups have dicotymous trees, supporting this interpretation of the data that the actual sequences within the nucleotide alignment define the subgroups.

There are some regions of sequence that are more variable than others across the catalytic domain, for example, the hypervariable region. The cumulative amino acid change was calculated for the whole sequence using the 14 clade alignment (Figure 6.11.). The synonymous changes appear to be fairly constant across the domain. The non-synonymous changes far exceed the synonymous changes in number and they occur at particular regions across the domain. This implies that some motifs undergo positive selection more than others.

6.2. Discussion

6.2.1. Does the Low Dose Inoculation Method Result in Simple Populations?

The aim of this part of the study was to use potentially simple populations of rat-derived *Pneumocystis* obtained from the low dose inoculation method, to characterise the diversity, number and relatedness of the genomic and expressed copies of *PRT1* genes. The resulting data was used to determine if there was regulation of *PRT1* gene expression. One or a few *PRT1* genes might be expressed per *Pneumocystis* organism. Alternatively, there might be no regulation of *PRT1* gene expression, resulting in all the *PRT1* genes within the *Pneumocystis* organism being expressed. From the *PRT1* assay, a cDNA majority sequence type was identified in three of the five simple populations, EA15, EA17 and EA19, while only clusters of expressed *PRT1* genes were detected in the other two simple populations, EA21 and EA24.

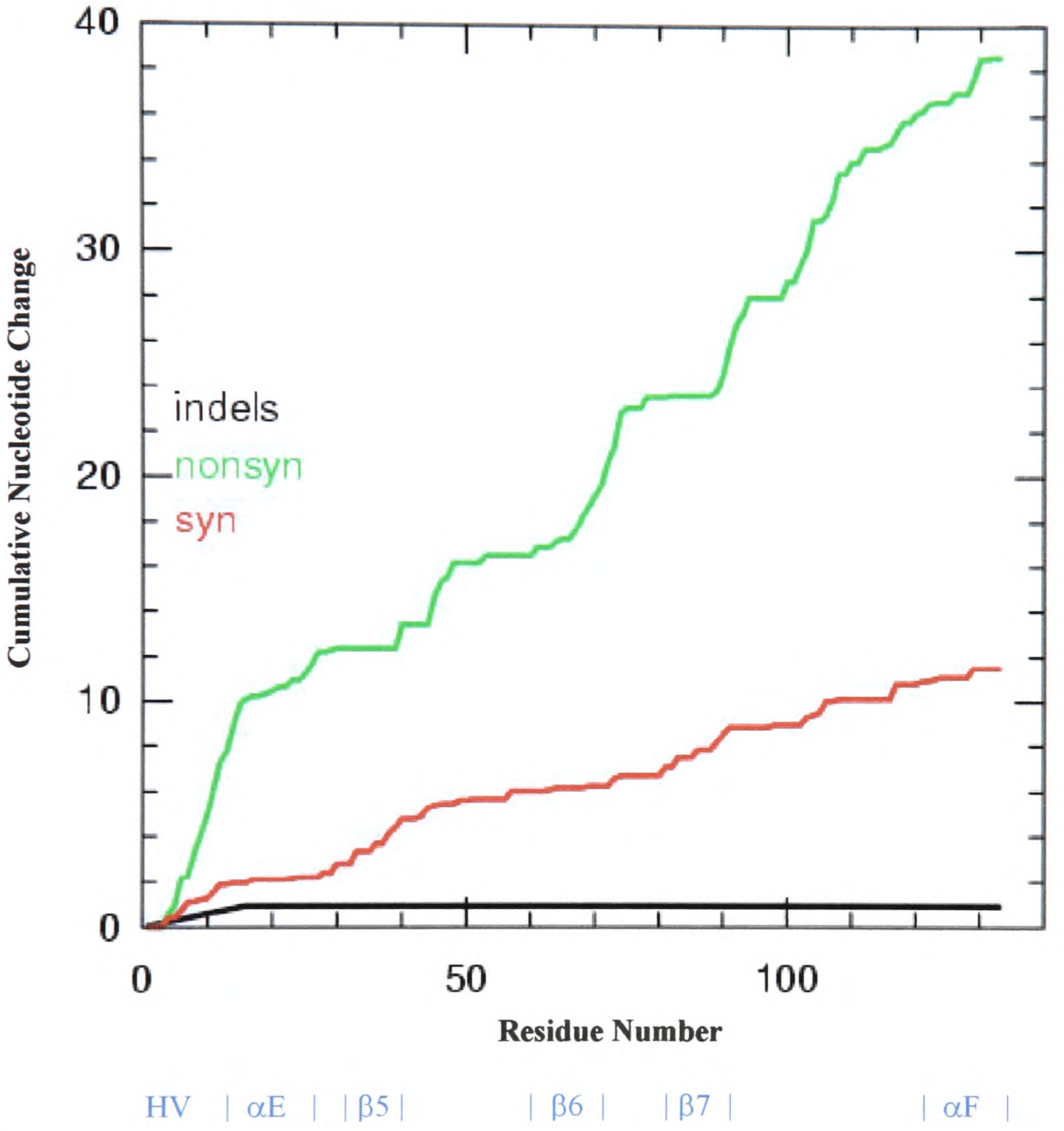


Figure 6.11. Cumulative behaviour of synonymous and non-synonymous changes codon by codon. Secondary structure is marked in blue.

The UCS/MSG junction can be used as an indicator of the clonality of a population (Keely *et al.,* 1999). The UCS has been shown to be present as one copy per *Pneumocystis* genome (Wada *et al.,* 1995), and is necessary for *MSG* gene expression (Sunkin and Stringer, 1997). Different types of organisms within a *Pneumocystis* population may be defined by the sequence of the *MSG* gene attached to the UCS locus. Amplification and sequence analysis of this junction gives an indication of how many different *Pneumocystis* organisms types are present in a rat-derived *Pneumocystis* population.

Stringer and Keely (personal communication) analysed the DNA sequence of *MSG* genes installed at the UCS from the five simple populations (Table 6.3.).

Population	**Type of population**	**No. plasmids sequenced**	**Majority *MSG* type**	**Major %[b]**	**Minority *MSG* type**	**Minor %[c]**
Inoculum[a]	Complex	34	A to X	-	-	-
EA15	Simple	19	*MSG*-A	94.7	*MSG*-a	5.3
EA17	Simple	10	*MSG*-Y	80	*MSG*-b, c	10, 10
EA19	Simple	10	*MSG*-A	80	*MSG*-a, d	10, 10
EA21	Simple	10	*MSG*-A	100	None	-
EA24	Simple	10	*MSG*-Y	80	*MSG*-b, e	10, 10

Table 6.3. Number and types of UCS/MSG sequences found in inoculated nude rats.
[a] This is the population of rat-derived *Pneumocystis* used to inoculate the rats. [b] Percent of plasmids containing the majority *MSG* gene. [c] Percent of plasmids containing the minority *MSG* gene.

Thirty four plasmids were analysed from the inoculum population yielding 24 different UCS-linked *MSG* genes labelled A to X. This number correlates with previous findings (Sunkin and Stringer, 1997). From Table 6.3, all five simple populations showed a tendency towards clonality with respect to the *MSG* gene. Nineteen plasmids were analysed from population EA15, of those 94.7% were the majority *MSG-A* type, suggesting that the population is simple. Ten plasmids were analysed from the remaining four simple

populations. All ten plasmids from population EA21 contained the *MSG-A* sequence, implying that this population is clonal. Eight of the ten plasmids from the other three populations EA17, EA19 and EA24 contained a majority *MSG* type, either *MSG-A* or *MSG-Y*, implying these populations, like EA15, are simple.

Not all of the sequences identified in the simple populations were detected in the initial inoculum. This implies that, either switching of *MSG* gene expression has occurred or, only a subset of the total *MSG* gene alleles was identified from the inoculum population. Switching of *MSG* alleles may also explain the origin of the minority *MSG* types identified in the simple populations. However, the minority *MSG* types may also be a reflection of the heterogeneity of the population.

The UCS/MSG data is the only available measure of *Pneumocystis* population clonality. Therefore, when assessing the potential clonality of the population, the assumption is made that the UCS/MSG data is reliable and reproducible. It is possible that there are different types of *Pneumocystis* within the simple populations that have the same *MSG* allele at the UCS. If this were true the populations would be more heterogeneous than the UCS/MSG data implies.

Based on the assumption that one *PRT1* gene is expressed per *Pneumocystis* organism, the data obtained from the *PRT1* assay does not fully correlate with the UCS/MSG data. From the *PRT1* data, the populations are more heterogeneous than the UCS/MSG data suggests. The identification of the *PRT1* cDNA majority sequence type in populations EA15, EA17 and EA19 suggests these three populations are simple. Here, the *PRT1* data concurs with the UCS/MSG data. However, in populations EA21 and EA24, the UCS/MSG data suggests that EA21 is clonal and that EA24 is simple. Yet, the results of the *PRT1* assay performed on

EA21 suggest that this population is heterogeneous with respect to the *PRT1* gene. The data from population EA24 is inconclusive, as several related sequences were detected, but a predominant sequence type was not identified from the twelve cDNA clones. An alternative explanation is that one copy of *PRT1* could be expressed per *Pneumocystis* organism, yet *PRT1* could undergo a higher level of switching in comparison to the UCS/MSG locus. If this were true, the simple populations would appear to be more heterogeneous when characterised by the *PRT1* assay than when characterised at the UCS/MSG locus.

A recent set of experiments was carried out by Stringer and colleagues (Keely *et al.*, 2003) with pathogen free rats using the low dose inoculation method. Populations were produced from an inoculum of 10^1 *Pneumocystis* organisms in which a single *MSG* sequence resided at the UCS locus in 80-90% of the organisms, confirming that *Pneumocystis* can proliferate within a rat without generating very high levels of UCS/MSG diversity. In this work, an inoculum consisting of 10^4 organisms was also used. It was concluded from the resulting low-level of diversity at the *MSG* locus that possibly only as few as 10 organisms had proliferated from the 10^4 inoculum in order to form the resulting population. This idea was extended to the low dose inoculum method such that only one organism might be required to initiate the population. This group have proposed that, if two organisms had proliferated from the low dose inoculum, there would have been a more equal proportion of the two types, assuming the two types were different. The hypothesis that only one *Pneumocystis* organism is required to produce a simple population can be extended to the 5 populations used in this study. Overall the UCS/MSG data suggests the five populations are simple and they originated from one or two organisms.

In contrast, assuming that one *PRT1* gene is expressed per *Pneumocystis* organism, the *PRT1* data suggests that at least one to ten different types of *Pneumocystis* organisms within the

10^1 inoculum could have produced the five simple populations. It is possible that one or two different *Pneumocystis* types initiated the simple populations EA15, EA17 and EA19 whose lower complexity is reflected by the cDNA majority sequence type. Populations EA21 and EA24 could have been generated from ten different *Pneumocystis* types and this is reflected by the heterogeneity of expressed *PRT1* genes identified in these two populations.

The nude rats used in these experiments are always vulnerable to external contaminating infection with *Pneumocystis* and it is impossible to prove that there are no latent *Pneumocystis* organisms already present in the lungs. Contaminating infection has the potential to complicate interpretation of data and so raises questions about the validity of the results. Pathogen free rats were used in the Stringer study (Keely *et al.*, 2003) and prior to inoculation were screened for rat-derived *Pneumocystis* DNA using PCR amplification of the mt LSU rRNA locus, yet no *Pneumocystis* DNA was detected. So, pathogen-free rats could be used in further experiments using the low-dose inoculation method, instead of nude rats. If the low dose inoculation method is to be used for further experiments, for example, the investigation of allele switching, only a simple population that is clearly defined can be used, then any contaminating infection can be identified.

To summarise, using the UCS/MSG data as a clonality indicator, five simple populations have been generated by the low-dose inoculation method. Yet, from the data obtained from the *PRT1* assay of these simple populations, what is meant by a simple population is not clear. For example, an alternative explanation for the *PRT1* data could be that each organism expresses more than one *PRT1* copy as discussed further in chapter 8. Ultimately, this method has the potential to generate large numbers of clonal *Pneumocystis* organisms, allowing large amounts of clonal DNA and RNA to be extracted for experiments.

6.2.2. Phylogenetic Analysis of Sequences of the Catalytic Domain Clones

The aim in this part of the study was to determine and display the relatedness of the different sequences and not to produce an accurate and precise evolutionary picture. Hence, the neighbour joining, NJ, method was used throughout the study because it combines computational speed with a single representative tree.

Many types of phylogenetic analysis of the sequences of the *PRT1* gene catalytic domain were considered (Page and Holmes, 1998). Neighbour joining is a widely used method for tree building, which combines computational speed with a single representative tree. Yet, it is a clustering algorithm where results to some extent depend on the order in which the sequences are added to the growing tree. Clustering methods do not allow the evaluation of competing hypotheses; they merely produce a single tree when many different trees could explain the data equally well. A second class of algorithms, the optimality criteria (Page and Holmes, 1998) choose among the set of all possible trees. A score or rank is assigned to each tree that is a function of the relationship between tree and data. Hence, the validity of any tree is evaluated. These methods are intensive to compute. Examples of these methods include maximum parsimony and maximum likelihood.

The three methods described, neighbour joining, maximum parsimony and maximum likelihood were employed on the dataset from the population EA15. Only subtly different trees were produced with all three methods. Clades where bootstrap values were high were maintained in all three trees. Differences between the trees arose where the bootstrap values were weak. This reflects the differences between individual sequences rather than the tree drawing method.

6.2.3. The Majority Sequence Types

The 3 published *PRT1* sequences 1B22 (Keely *et al.*, 2001), pAGA (Lugli *et al.*, 1997), and *ORF-3* (Wada and Nakamura, 1994), suggest that the genomic majority sequence type is not an artefact of PCR amplification in the *PRT1* assay. The 3 genes are identical apart from the proline rich region, which due to the intrinsic repetitive nature is a particularly hard region to sequence. Therefore, the lack of identity at this region could be the result of sequence artefacts. The 5'UTR and 3'UTR for 1B22 (Keely *et al.*, 2001) and *ORF-3* (Wada and Nakamura, 1994) have been published. These are also 100% identical for at least 300 bp either side of the gene, implying that all 3 are copies of the same gene and it can be presumed that 1B22 and *ORF-3* are in the same genomic context.

The synonymous and non-synonymous changes present indicate that the genomic majority sequence type is not under selective pressure. No cDNA copies of this sequence type were identified, implying this might be a nonexpressed gene with many copies, possibly a redundant gene. Alternatively, it could be an active gene with a very low level of expression. The presence of the potentially autocatalytic KRGAY motif supports the idea that it is a nonexpressed gene because, if the protein were expressed it might be cleaved and the catalytic site would be inactivated.

In 3 of the 5 simple populations, the cDNA majority sequence type was identified as the predominantly expressed sequence type. From the low-dose inoculum, the same *Pneumocystis* types expressing the cDNA majority sequence type may have proliferated in these 3 populations while other *Pneumocystis* types, not expressing the cDNA majority sequence type may have proliferated in populations EA21 and EA24. It is possible that the cDNA majority sequence type encodes a more efficient isoform of PRT1. PRT1 is a surface expressed protease, which is assumed to interact with the host environment, or possibly

functions to cleave UCS from MSG. The structure of the protein encoded by the cDNA majority sequence type might be better adapted to perform the potential functions of PRT1. The amino acid sequence of the cDNA majority sequence type offers no suggestion as to why this allele is more stable and could function better than any of the others.

Sequences were identified which represented the 14 clades produced during analysis of the sequences from the 209 clones. There are clades other than the cDNA majority sequence type that contain many cDNA alleles of *PRT1*, for example clade I and clade Q. Clade I contains 9 copies of only cDNA alleles, which are expressed in 5 out of the 7 populations. Clade Q consists of 7 cDNA copies and 2 genomic, one of which is 11a111 and it is also present in 5 out of the 7 populations. Neither clade has a motif or residue to suggest it encodes a unique function or modification of PRT1.

There are two main differences when comparing the *PRT1* sequences with the sequences from other proteases, kexin 2, human furin and human PC1 and PC3 protease. The *PRT1* sequences have a glycine at residue 39 while the other proteases have an alanine, and at residue 82 the *PRT1* sequences have a cysteine while the other proteases either have a serine or an alanine. This implies that the *PRT1* sequences may encode for an extra disulphide bond in contrast with the proteases, so PRT1 may have a subtly different structure or function.

It is not immediately clear as to why some *PRT1* copies should be expressed in some populations more than others. There could be a bias in the generation of the *PRT1* data from the populations but this is not sufficient to explain the results. It is apparent that some copies do seem to predominate. Like the cDNA majority sequence type, these copies could be more stable than other copies or the bias could originate from the paucity of *Pneumocystis* types present within the initial inoculum dilution preparation.

6.2.4. SNAP Analysis

The average ds/dn ratios for the populations ranged from 0.67 to 1.64, showing that some of the populations, by the definition described in section 2.9.4, were under selective pressure while others were not. The populations subdivide, not into cDNA and genomic clones, but into two sets of sequences. Populations with two subgroups have dicotymous trees showing the subgroup division is truly a result of the clone sequence. For the 14 clade population, one subgroup has a high ds/dn ratio, 1.64, while the other ds/dn ratio is low, 0.67. This suggests that the population divides into one group of alleles that are under positive selection and another group that is under some form of selective constraint.

The technique of SNAP analysis has limitations. Assumptions are made in the analysis (Page and Holmes, 1998), all nucleotide sites change independently, the substitution rate is constant over time, the base composition is at equilibrium and that the conditional probabilities of nucleotide substitutions are the same for all sites and do not change over time. While these assumptions make the method tractable they are not wholly realistic and the results may not reflect the true biological situation. Selection of alleles is a complex phenomenon and cannot be summarised as one mathematical programme.

6.2.5. Limitations of the *PRT1* Assay

A single copy expression site like the UCS had yet to be identified for the *PRT1* gene, so instead it has to be investigated at an organism and population level. There are approximately 30 different copies of the *PRT1* gene per *Pneumocystis* organism, 1 copy at the end of all the 16 chromosomes except one (Lugli *et al.*, 1997, Cushion *et al.*, 1993a). So a data set generated by the *PRT1* assay of *PRT1* genomic copies could contain approximately 30 different *PRT1* alleles, making it difficult to investigate the function and expression of PRT1 from the genomic copies alone. From the *PRT1* assay, more might be learnt about the

expression of PRT1 from the cDNA copies, for example, assuming that the populations are simple, if only one copy of *PRT1* is expressed per *Pneumocystis* organism, a predominant cDNA majority sequence type would be observed.

Had a single expression site for the *PRT1* gene been discovered, then an equivalent UCS/MSG junction assay could be performed. Yet, the overall heterogeneity of the *PRT1* data implies that several favoured expression sites is probably more likely for the *PRT1* genes than one sole expression site. The number of *PRT1* expression sites may be elucidated by the *Pneumocystis* genome sequencing project (http://www.gene.genetics.uga.edu). Assuming that a single expression site does not exist, more informative data could be obtained based upon a greater number of clones (>12). One population expressing the cDNA majority sequence type could be subjected to a comprehensive analysis in order to discover whether the same proportion of cDNA majority sequence type to heterogeneous *PRT1* sequence types was also obtained from a larger sample size. The reliability and reproducibility of the *PRT1* assay needs to be further characterised. Population EA21 was analysed twice, however, there was little correlation of *PRT1* alleles between the two analyses (Figure 6.5.). The optimum number of analyses (cDNA synthesis, PCR amplification and cloning), of a population needs to be evaluated such that a reliable comprehensive picture of the *PRT1* alleles present in the population has been obtained.

Even though the *PRT1* gene assay used in this study was successful, it does have limitations. Rat-derived *Pneumocystis* organisms from the simple populations were of low abundance and are of limited viability. This limits the quantity and quality of RNA that could be isolated which will be reflected in the subsequent efficiency of RT-PCR amplifications. Within the molecular biology techniques, there might be a bias as to which alleles can be

cloned successfully. This would result in a skewed picture of the different copies of the *PRT1* gene.

6.2.6. Conclusion

The low dose inoculum method used in this study is able to produce simple populations. The UCS/MSG data suggests that all five populations are simple; by contrast the *PRT1* gene data suggests that only 3 populations are. From the *PRT1* gene data, a cDNA majority sequence type and a genomic majority sequence type have been identified and characterised to the amino acid level. The *PRT1* gene data indicates that expression of *PRT1* genes might be subject to some form of regulation as a predominant cDNA majority sequence type has been detected. The simple populations are related to the initial inoculum but analysis of further samples is required to provide a more comprehensive picture and reduce the contrasting data from the UCS/MSG and *PRT1* assays.

7. Further Characterisation of the Complex Pattern and the Possible Regulation of *PRT1* Gene Expression using Simple Populations, Hybridisation and 5' RACE.

The aim of this part of the project is to investigate further the diversity and number of genomic and expressed copies of *PRT1* and to determine whether *PRT1* gene expression is regulated. Whether one, a few or all *PRT1* genes are expressed per *Pneumocystis* organism. The previous results obtained from analysis of *PRT1* genes were extended in three different studies. Using the *PRT1* assay, the aim of the first study was to analyse the rat derived *Pneumocystis* populations generated from the second passage of the first set of simple populations. This will determine whether the same *PRT1* alleles found in the first set of populations are identified in the second passage. The aim of the second study was to characterise the genomic context of the majority sequence types by designing probes specific to the majority sequence types and using them in hybridisation experiments with genomic rat-derived *Pneumocystis* DNA. The third study was to use 5' RACE to characterise the 5'UTR of the *PRT1* gene in order to determine how the *PRT1* gene might be regulated.

7.1. Second Passage Populations

In the autumn of 2001, another set of experiments utilising the low dose inoculum method (section 2.6.1) was begun in order to generate a second set of simple rat-derived *Pneumocystis* populations. Several aspects of this method were investigated within this one set of experiments. Firstly, a separate dilution was prepared from the complex population, which had been used as an inoculum in the first set of experiments. This was done in order to see if a population resulting from a different dilution preparation expressed the same

majority types as the first set of simple populations. Different dosage levels of the inoculum were used, 10^3, 10^2 and 10^1 organisms. Thus, the relationship between the number of organisms in the inoculum and the complexity of the population that resulted could be analysed. Secondly, non-inoculated rats were put into cages with inoculated rats in order to investigate the potential for airborne transmission of simple populations. This might provide a non-invasive method of propagating the simple population model. Thirdly, a new complex population, different from the one used previously, was used as an inoculum in order to examine if the same majority types are expressed as from the previous inoculum. Finally, inocula of up to 10^7 *Pneumocystis* organisms derived from the previous simple population EA21, were used to attempt propagation of clonal organisms. This assumed that a large inoculum would create a prolific infection, which might contain clonal *Pneumocystis* organisms.

As discussed previously, latent infection present in the nude rats may complicate interpretation of the experiments. In this experiment, three nude rats from the same supplier were screened prior to inoculation for the presence of latent rat-derived *Pneumocystis* DNA by nested PCR amplification of the mt LSU rRNA locus. Total DNA was extracted from 2 lung samples from each rat and 3 independent PCR amplifications were carried out on each sample. Rat-derived *Pneumocystis* DNA was detected in 2 of the 3 nude rats, but in only 4 of 18 nested PCR amplifications. Professor Wakefield carried out these control experiments. Rat-derived *Pneumocystis* DNA was present in these control rats but only at a very low level, as PCR amplification of the mt LSU rRNA locus is able to detect a minimum of 1-2 *Pneumocystis* organisms in a biological sample, (Peters *et al.*, 1992a) and not all amplifications were positive.

Of the 20 rats used in the second set of experiments, 3 were sentinels (not inoculated with rat-derived *Pneumocystis* organisms) so identification of any latent *Pneumocystis* organisms that developed into an infection, or any external contaminating *Pneumocystis* infection was possible. At the conclusion of the combined experiment, in 2 of the sentinel rats, low levels of *Pneumocystis* organisms were detected by microscopy and high levels were detected in the third. These levels of infection were far greater than those detected by PCR amplification prior to the inoculation as microscopy is able to detect only down to a minimum of 10^2 *Pneumocystis* organisms, in a biological sample (Peters *et al.,* 1992a) and *Pneumocystis* organisms were present in all 3 rats. Some form of contaminating infection either latent or external had occurred. Therefore, caution is required when interpreting the results from the second passage populations because the infection in the experimental animals was not necessarily from the injected inoculum and might have been the result of latent infection or external contamination. This contrasts to the first round of experiments using the low-dose inoculum where no *Pneumocystis* DNA was detected within the sentinel rats after they were sacrificed at the end of the experiment.

The cDNA copies of *PRT1* genes from 3 populations were investigated using the *PRT1* assay (section 6.1.2.). The first population, 714, was grown from an inoculum of 10^3 *Pneumocystis* organisms from population EA21. It was felt important to discover if a second passage of the simple populations was possible. The other 2 populations, 716 and 723, were selected in order to investigate airborne passaging of simple populations. This was considered relevant because the possibility of airborne transmission might provide a natural means of propagating the simple populations. Population 723 was created from an inoculum of 10^1 organisms from a new serial dilution preparation from the original complex inoculum. This should result in a simple population that may have the same *PRT1* types as the first set of simple populations. Rat 716 was in the same cage as rat 723 but was not inoculated. The

animal developed a rat-derived *Pneumocystis* infection that if the barrier and filter conditions were 100% efficient should only have come from rat 723. If this is true, the *PRT1* clones obtained from the populations 716 and 723 should correlate, yet the caveat of latent or external infection remains.

Twelve cDNA clones were analysed from populations 714 and 716 and 13 cDNA clones were analysed from population 723 (Table 7.1).

Population	**714**	**716**	**723**
Clone Name	714c1d	716c3a	723c7a
	714c1dd	716c3b	723c7b
	714c1h	716c3d	723c7d
	714c1i	716c3g	723c7i
	714c1k	716c3i	723c7k
	714c1v	716c3j	723c7l
	714c4a	716c3k	723c7p
	714c4d	716c3m	723c7q
	714c4g	716c3o	723c7t
	714c4m	716c3p	723c7u
	714c4o	716c3q	723c7x
	714c4r	716c3r	723c8c
			723c8f
Total Number	12	12	13

Table 7.1. cDNA clones from populations 714, 716 and 723.

The *PRT1* sequences from each population were obtained using the *PRT1* assay (section 6.1.2) then aligned using ClustalW and phylogenetic trees were derived by the Neighbour Joining Method. Figure 7.1 shows the phylogenetic tree comparing the cDNA clones from both previous analyses of population EA21 and the clones from population 714. Population 714 was possibly more clonal than EA21, having one cluster of 3 identical cDNA clones and one pair of cDNA clones. Analysis of population EA21 produced only pairs of cDNA clones. The clonality of population 714 is supported by the divergence analysis: EA21

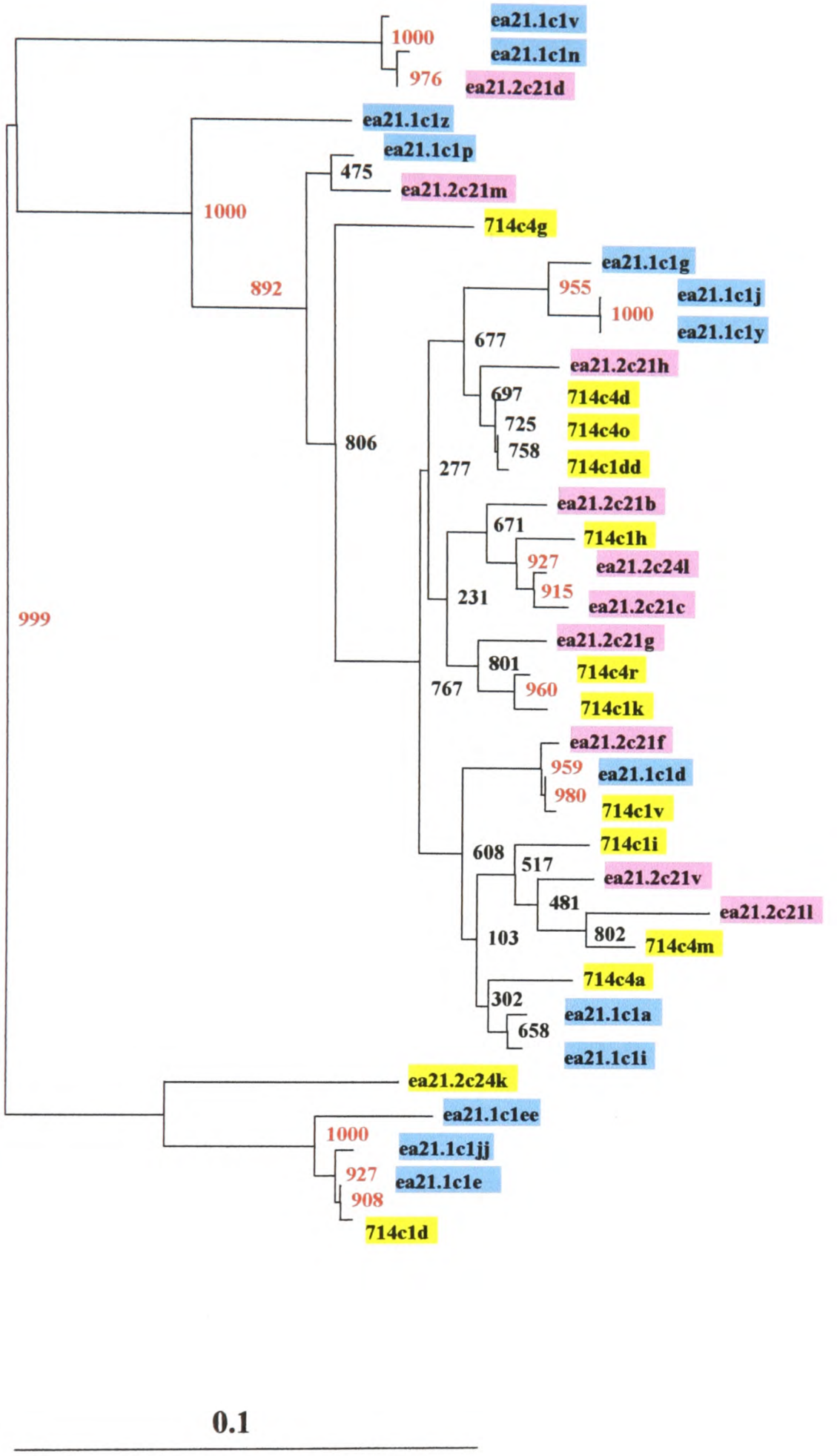


Figure 7.1. Phylogenetic tree comparing the sequence of cDNA copies from populations 714 and EA21. It is derived by the ClustalW Neighbour-Joining Method. Copies of 714 are highlighted in yellow, EA21.1 in blue and EA21.2 in pink. Bootstrap values are placed at the nodes of the branches. Significant bootstrap values (>850) are highlighted in red. The scale bar represents the branch length that is equivalent to 10 nucleotide changes per 100 nucleotide residues.

cDNA clones are 15.7% divergent, while population 714 clones are only 8.7% divergent, and so are more homologous, which implies that population 714 is less complex than population EA21. The relatedness of populations 714 and EA21 was also examined. In the phylogenetic tree containing the clones from both populations (Figure 7.1), there were two clusters of sequences consisting of one clone from population 714 and two clones from EA21. As a control, population 714 was compared with population 1984. In contrast to the comparison of population 714 with EA21, population 714 was less related to 1984 with only 1 pair of sequences in common. These data are consistent with population 714 being the result of the second passage of EA21, and population 714 is thought to be more clonal than EA21.

Figure 7.2 depicts the phylogenetic tree comparing the cDNA clones from populations 723 and 716. Divergence analysis suggests that both populations have comparable levels of complexity. Assuming that population 716 did result from airborne transmission of population 723, the *PRT1* alleles from population 716 would be expected to be a subset of the *PRT1* alleles from population 723. This was not evident from the phylogenetic tree; yet, biologically airborne transmission is the most likely origin of population 716.

Population 723 grew from an inoculum of 10 rat-derived *Pneumocystis* organisms. The inoculum was a different diluted preparation derived from the same stock of organisms used as the complex inoculum population in the first set of experiments. The sequences obtained from the inoculum clones from the first set of experiments were compared with the sequences from populations 723 and 716, in order to determine if there was any correlation. As a control, populations 723 and 716 were also compared with 1984. The clusters are summarised in Table 7.2.

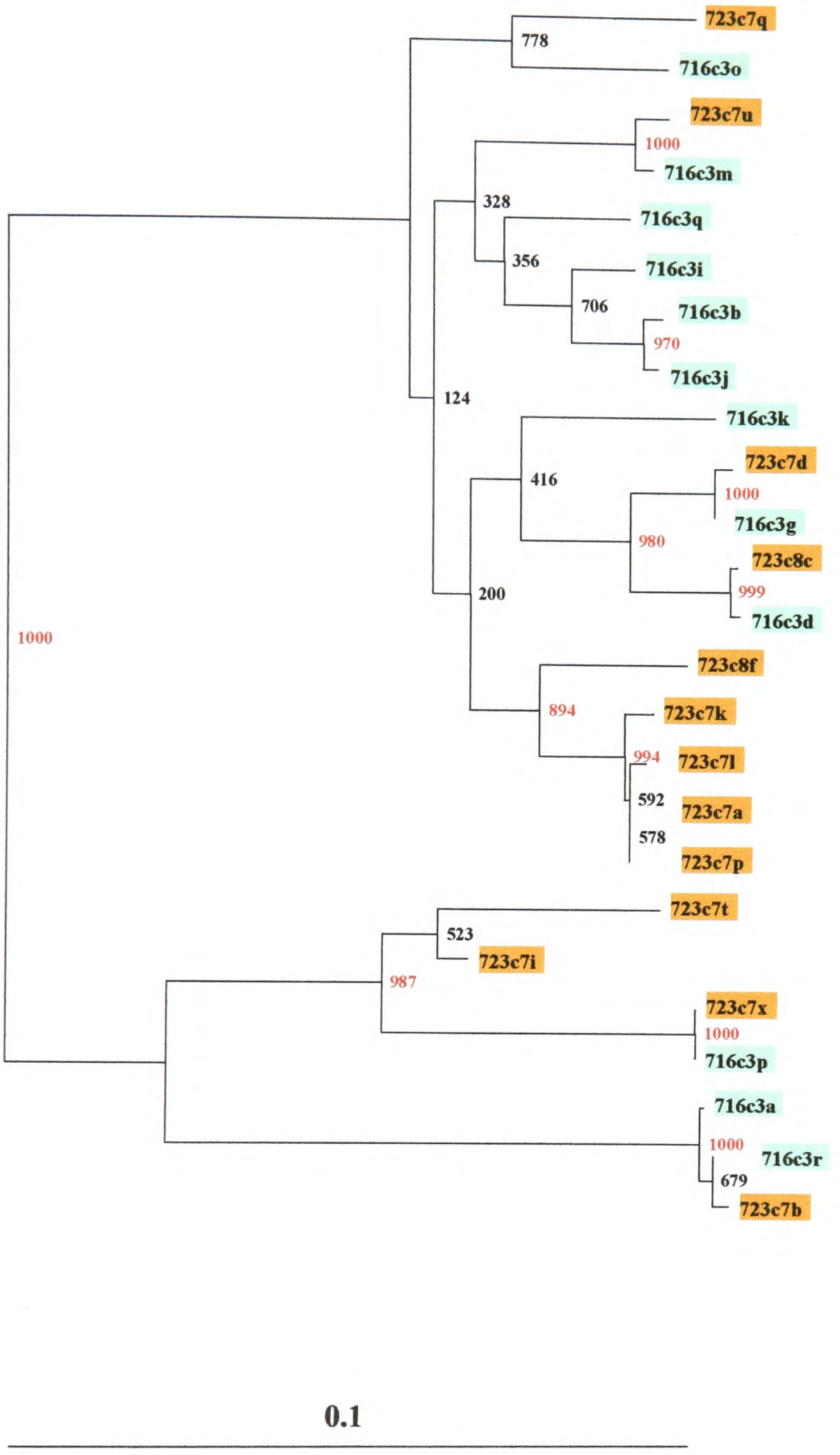


Figure 7.2. Phylogenetic tree comparing the sequence of cDNA copies of populations 723 and 716. It is derived by the ClustalW Neighbour-Joining Method. 723 copies are highlighted in orange while 716 copies are highlighted in green. Bootstrap values are placed at the nodes of the branches. Significant bootstrap values (>850) are highlighted in red. The scale bar represents the branch length that is equivalent to 10 nucleotide changes per 100 nucleotide residues.

(i) Inoculum Clusters	(ii) 1984 Clusters
1x inocg/ 2x 716c/ 1v 723c	1x 1984g/ 1x 716c/ 1x 723c
4x inocg/ 1x inocc/ 1x 716c/ 1x 723c	1x 1984g/ 4x 723c
1x inocc/ 1x 716c	1x 1984c/ 1x 716c/ 1x 723c
2x inocc/ 1x 716c/ 1x 723c	1x 1984c/ 1x 716c/ 1x 723c
1x inocc/1x 723c	
2x inocc/ 4x 723c	

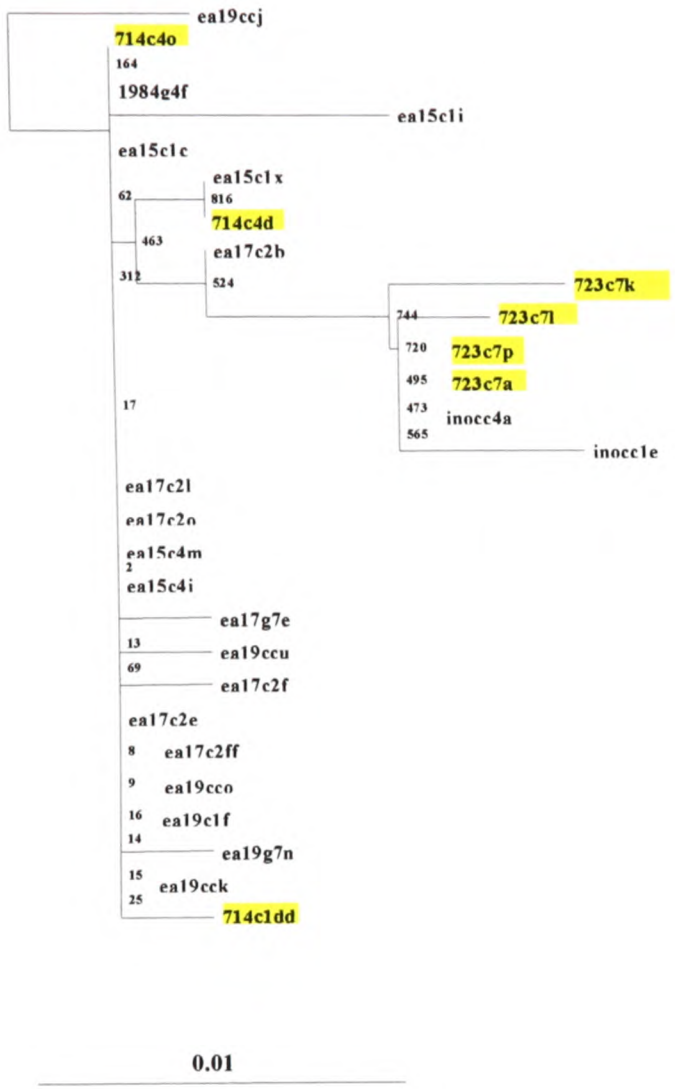
Table 7.2. Comparison of the clusters found in the two phylogenetic trees.

(i) inoculum compared with populations 716 and 723 (ii) 1984 compared with populations 716 and 723. g = genomic clone, c = cDNA clone, inoc = inoculum.

There is very little difference between the relatedness of populations 723 and 716 to each other, and with either the inoculum, or with 1984. Both comparative trees contain a similar number of clusters, implying the sequences of populations 723 and 716 are equally related to the inoculum and also to 1984, thus this data suggests that population 723 may have originated from either EA21 or 1984. Yet, only a subset of *PRT1* alleles was characterised from each population so a fully comprehensive picture cannot be determined.

A copy of the cDNA majority sequence type was compared with all the cDNA clones obtained from the second passage populations. From population 723, 4 copies were identical with the cDNA majority sequence type and from population 714, 3 copies were identical, however, none of the sequence types obtained from population 716 were identical to the cDNA majority sequence type. The cDNA majority type sequences from the first and second passage populations were assembled into a phylogenetic tree (Figure 7.3). Like Figure 6.4, the structure of the amino acid tree is identical to that of the nucleotide tree, because most of the nucleotide changes are non-synonymous. The sequences were visually compared and 37 nucleotide changes were identified in the sequence alignment; 4 are synonymous, and 33 are

Nucleotide



Amino Acid



Figure 7.3. Phylogenetic trees comparing nucleotide and amino acid sequences of the cDNA majority sequence type with equivalent clones from the second passage populations 714 and 723. They are derived by the ClustalW Neighbour-Joining Method. Clones from the second passage populations are highlighted in yellow. Bootstrap values are placed at the nodes of the branches. The scale bars represent the branch length that is equivalent to 1 nucleotide change per 100 nucleotide residues.

non-synonymous, which results in a ds/dn ratio of 0.121. This further supports the idea that the cDNA majority sequence type is under positive selection.

The second passage populations were phylogenetically compared with the 10 previously published *PRT1* sequences (section 1.4.3) (data not shown). 3G5 is identical with a cDNA copy from populations 716 and 723; 11A11 is identical with a cDNA copy from populations 714 and 723, while 73j is identical with 3 cDNA copies from population 716. This last observation is somewhat unusual, because *PRT1* gene copies identical to 73j are rarely identified.

The results from the SNAP analysis of the cDNA clones from the 3 second passage populations are shown in Table 7.3.

Population	Average			Interquartile Range		
	ds	dn	ds/dn	ds	dn	ds/dn
714	0.07	0.10	0.63	0.02 – 0.06	0.06 – 0.10	0.40 – 1.00
716	0.10	0.15	0.67	0.03 – 0.15	0.06 – 0.25	0.50 – 0.79
723	0.13	0.16	0.79	0.05 – 0.19	0.07 – 0.23	0.62 – 1.03

Table 7.3. Average values for ds, dn and ds/dn ratios for the populations 714, 716 and 723 and the corresponding interquartile ranges.

Interquartile ranges are the range of values from the 25th to the 75th percentile.

7.2. Discussion of Second Passage Populations

714 is a simple population resulting from the second passage of EA21. It appears to be more clonal than EA21, thus implying that passaging may decrease the number of *Pneumocystis* organism types within a simple population, even though an inoculum of 10^3 organisms was used. Airborne transmission of *Pneumocystis* organisms appears to have occurred between the rats from which populations 723 to 716 were extracted, yet, there is little evidence from

the *PRT1* assay that population 723 is derived from the inoculum rather than the complex population, 1984. Throughout all of this analysis, a comprehensive picture cannot be obtained as only a subset of *PRT1* alleles were characterised from each population.

The cDNA majority sequence type was detected in 2 of the second passage populations, 723 and 714, strengthening the argument for it being one of the dominant *PRT1* cDNA forms, especially as it was detected in population 723, which was generated from a low-dose inoculum using a separate dilution of the original complex inoculum. It is of interest that a different dilution of 10 *Pneumocystis* organisms resulted in the same *PRT1* phenotype, because these ten organisms might be different from the original ten organisms that generated the first five simple populations, assuming that the original complex inoculum consisted of more than ten different types of *Pneumocystis*. More than ten different types of *Pneumocystis* would be feasible if the original complex inoculum was generated from a pool of *Pneumocystis* organisms but may not be so valid if the original complex inoculum was extracted from one infected rat. The true test would be to take a novel complex population as an inoculum, for example population 1966 (from the same rat colony as population 1984) and determine if the cDNA majority sequence type is present.

Population 714 is the second passage of EA21. It is possible that the cDNA majority sequence type was expressed in EA21 but was not identified because only a subset of the total number of *PRT1* alleles present were sampled. Alternatively, the cDNA majority sequence type was not expressed in EA21 but was expressed in population 714 because the influence of different environmental factors caused each population to generate a different overall phenotype.

cDNA copies were analysed from these 3 populations, therefore the SNAP analysis is exclusive to expressed *PRT1* alleles. All 3 of the second passage populations have ds/dn ratios of less than 1, indicating by definition (section 2.9.4) that *PRT1* gene expression is under strong positive selection. This suggests that expression of *PRT1* alleles in these populations was under some form of external pressure. As in the first set of simple populations (section 6.1.10) all 3 populations subdivide into 2 groups and, as before one group has a small ds/dn ratio while the other ratio is larger.

Within this second set of simple populations other aspects could be investigated. An important experiment would be to see if the new complex inoculum, which was used to inoculate some of the rats in the second set of experiments, produced the same cDNA majority sequence types and also whether or not the genomic majority sequence type was present. Rat-derived *Pneumocystis* populations that result from inocula with different amounts of *Pneumocystis* organisms could be analysed. This latter experiment might determine how the number of organisms in the inoculum affects the level of complexity in a population. The large dose inocula of 10^7 *Pneumocystis* organisms, from the simple population EA2I, might produce simple populations consisting of an abundance of potentially clonal rat derived *Pneumocystis* organisms. Sufficient DNA and RNA could be extracted to characterise the genomic and expressed context of the genomic majority sequence type and the cDNA majority sequence type in these organisms.

7.3. Results of DNA Hybridisation Experiments to Characterise the Majority Sequence Types.

For the second study in this section of the project, DNA hybridisation experiments were planned to examine the genomic context of both the cDNA majority sequence type and the genomic majority sequence type using specific oligonucleotides.

7.3.1. Probe Design

Alignment of the DNA sequences of *PRT1* alleles indicated that the first 50 bp of sequence was the most variable region (Figure 7.4). Forty sample sequences were chosen from the 209 nucleotide sequence alignment to represent all of the major groups of clones from the *PRT1* assay. For both the cDNA majority sequence type and the genomic majority sequence type, a probe of approximately 20 bp was chosen for that majority sequence type from the region that contained the most unique residues. A third probe was designed that contained the most unique residues of any sequence to bind in that region. This probe may act as a control in order to determine the specificity of hybridisation. The probe specific to the genomic majority sequence type is abbreviated to geno or HA1 when it is used a primer; the cDNA majority sequence type specific probe was abbreviated to cDNA or HA2 when it is used as a primer, and the third probe for determining specificity is abbreviated to MID or HA3 when it is used as a primer.

The three probes had an outstanding level of specificity. They are almost 100% specific to these sequences. However, they are also homologous to a few clones that are not identical to the majority sequence types at other parts of the catalytic domain sequence. There are only a few nucleotides that are unique to the majority sequence types. To fully take advantage of

```
          1                                                  50
ea211g7j  TGGGGACCTG CCGATACTGG GAATTTAATT CAAGATATTT CTTATACTAG
  inog7e  TGGGGACCTC CTGAATACGG ATTTTTAGCT ATGCCAATGA TGCCTACTAC
 ea19g7i  TGGGGACCTC CTGGTGATGG ATATGCAGCT ATCCCAATGT ATCCTACTAC
 1984g4g  TGGGGACCTC CTGATTATAA AGATTTAGCT ATGCCAATGA TGCCTGCTAC
 ea19g7o  TGGGGACCTC CTGATTATAA AGATTTAGCT ATGCCAATGA TGCCTGCTAC
 inocg7r  TGGGGACCTC CTGATTATGG ATATGTAGCT ATGCCAATGA TGCCTACTAC
 inocg7s  TGGGGACCTC CTGAATACGG ATCTTTAGCT ATGCCAATGA TGCCTACTAC
 ea19g7e  TGGGGACCTC CTGATAATGG AGAATTAGCT ATGCCAATGA TGCCTACTAC
 ea15c4j  TGGGGACCTC CTGATGATGG AAATGTAGTT ATGCCAATGA TGCCTACTAC
 ea19cco  TGGGGACCTC CTGATGATGG AAATGTAGTT ATGCCAATGA TGCCTACTAC
 ea15c1x  TGGGGACCTC CTGATGATGG AAATGTAGTT ATGCCAATGA TGCCTACTAC
 1984g4f  TGGGGACCTC CTGATGATGG AAATGTAGTT ATGCCAATGA TGCCTACTAC
 inocc4c  TGGGGACCTC CTGGTGATGG ATATGCAGCT ATCCCAATGT ATCCTACTAC
 ea15c4d  TGGGGACTTC CTGATAATGG AGAATTAGCT ATGCCAATGA TGCCTACTAC
 1984c3c  TGGGGACCTC CTGATGATGG AAATGTAGTT ATGCCAATGA TGCCTACTAC
 1984c3h  TGGGGACCTT CTGATGATGG AGATGTAGCT ATGCCAATGT TTCCTACTAC
 inocc4i  TGGGGACCTC CTGATGATGG AGATGTAGCT ATGCCAATGT TTCCTACTAC
 ea19c1b  TGGGGACCTC CTGATGATGG AGATGTAGCT ATGCCAATGT TTCCTACTAC
 inocc1a  TGGGGACCTC CTGATTATAA AGATTTAGCT ATGCCAATGA TGCCTGATAC
 ea19g7d  TGGGGACCTT TTGATTTAGG AAATAAGATT GAAAACATTT TTTATTCTAC
 ea19ccs  TGGGGACCTT CTGATTTAGG AAATAAGATT GAAAACATTT TTTATTCTAC
 ea19g7k  TGGGGACCTC CTGAATACGG ATTTTTAGCT ATGCCAATGA TGCCTACTAC
 ea17c2d  TGGGGACCTC CTGAATACGG ATTTTTAGCT ATGCCAATGA TGCCTACTAC
 inocc1t  TGGGGACCTC CTGATGATGG AGATGTAGCT ATGCCGATGT TTCCTACTAC
 ea15g7j  TGGGGACCTT CTGATAATGG AGATGTAGCT ATGCCAATGT TTCCTACTAC
 ea15g7k  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGATATTT TTTATACTAC
 ea19g7a  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGATATTT TTTATACTAC
ea211g7l  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGATATTT TTTATACTAC
 ea19g7b  TGGGGACCTG CCGATACTGG GAATTTAATT CAAGATATTT CTTATACTAG
 ea24g7i  TGGGGACCTG CCGATACTGG GAATTTAATT CAAGATATTT CTTATACTAG
 1984g4h  TGGGGACCTG CCGATACTGG GAATTTAATT CAAGATATTT CTTATACTAG
 ea19g7f  TGGGGACCTG CCGATACTGG GAATTTAATT CAAGATATTT CTTATACTAG
 ea17g3o  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGAGATTT TTTATACTAC
 ea19g7h  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGAGATTT TTTATACTAC
ea24c219  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGAGATTT TTTATACTAC
ea211c1e  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGAGATTT TTTATACTAC
 ea15g7v  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGAGATTT TTTATACTAC
ea211g7o  TGGGGACCTG CCGATACTGG GAATTTAACT CAAGAGATTT TTTATACTAC
 ea19g7l  TGGGGACCTT CCGATACTGG GAATTTAACT CAACAGATTT TTTATACTAC
ea211c1z  TGGGGACCTG CCGATACTGG GAATTTAATT CAAGATATTT CTTATACTAG
```

cDNA majority sequence type

MID probe

Genomic majority sequence type

Figure 7.4. ClustalW alignment of the first 50 bp of a representative group of 40 *PRT1* catalytic domain sequences. The specfic probe sequences are highlighted in pink and are labelled on the right in bold. The cDNA majority sequence type and genomic majority sequence type are indicated.

these unique nucleotides, hybridisation conditions would have to be established that could distinguish between a difference of 1 nucleotide. Such fine-tuning of hybridisation conditions was not considered to be appropriate for this project, as it would be likely to involve many months of careful experimentation. A series of probes could be designed, whereby each one bound to a subset of clones that would include the majority sequence type, plus other clones. The majority sequence clones would be identified as the alleles that hybridised to all of the probes. Here, the first probes of a potential series were designed and characterised by experiment.

7.3.2. Dot Blots to Test the Specific Probes

A dot blot, comprised of 48 clones from the *PRT1* assay, was devised to test the specificity of the three probes, geno, cDNA and MID. To examine any difference dependent upon the source of the DNA, both PCR products and plasmids were used in equal amounts as templates for hybridisation. Figure 7.5(a) shows the results of a test hybridisation of the 48 clones with the universal probe PcProt3. Plasmid templates were more efficient at binding PcProt3 than the PCR products. Column 6 is comprised solely of PCR products. Each of these 8 PCR products had an equivalent plasmid clone elsewhere on the blot. PcProt3 bound more efficiently to all 8 plasmids than to the PCR products. There is one exception, clone F1. PcProt3 bound poorly to this plasmid because the DNA was not extracted correctly using the DNA extraction kit (section 2.8).

Different levels of stringency of washing were examined with constant hybridisation conditions. Under the standard wash conditions (section 2.11.) the stringency was taken to 2 x SSC and 0.2% SDS at a temperature of 45°C. These conditions were not sufficiently stringent to attain the level of specificity required for the probe. The temperature and length of washes were kept constant but the salt concentration was lowered to 0.5 x SSC and then to

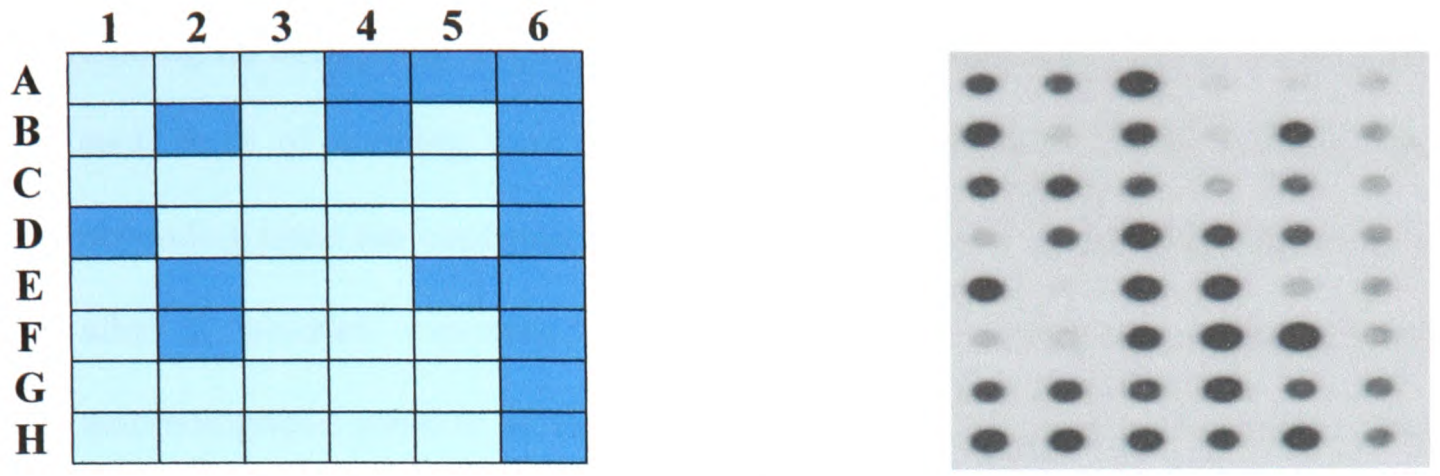


Figure 7.5. (a) Dot blot hybridisation demonstrating the different efficiencies of hybridisation between PCR products and plasmid DNA. The table on the left is a schematic representation of the positions of the *PRT1* catalytic domain clones. The PCR products are highlighted in dark blue while the plasmid DNA is highlighted in light blue. The dot blot hybridisation with the universal probe PcProt3 is shown on the right. Plasmid DNA produces more intense spots than PCR products, so is a better hybridisation target. This is true for all the clones apart from F1. This plasmid DNA was of poor quality.

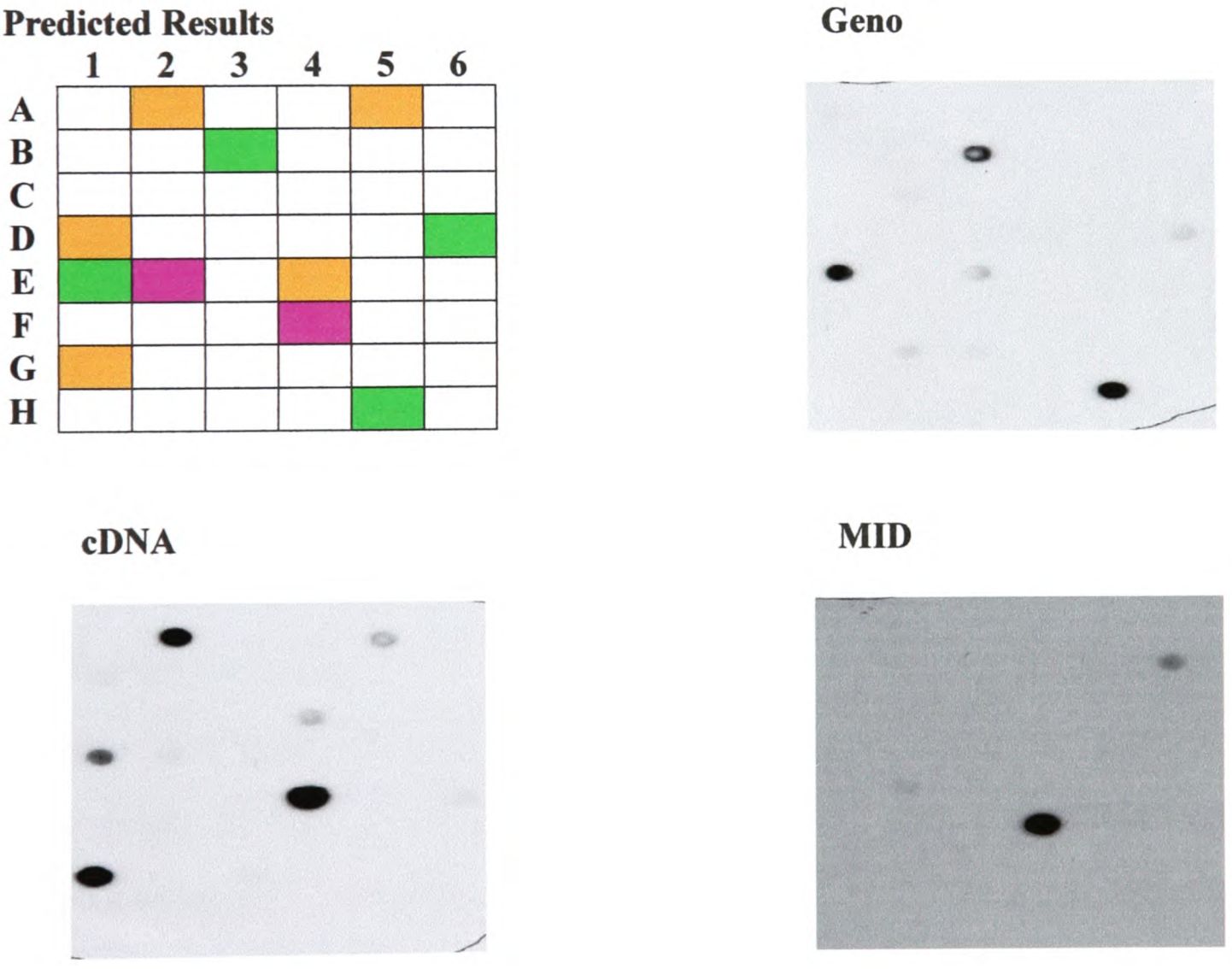


Figure 7.5. (b) Dot blot hybridisation for *PRT1* catalytic domain clones using specific probes. The table (top left) is a diagrammatic representation demonstrating which clones each of the specific probes will bind to. Homologous clones are highlighted in: geno probe green; cDNA probe orange and the MID probe pink. The actual dot blot hybridisations are shown and are labelled accordingly.

0.125 x SSC in order to achieve the level of stringency needed to gain the specificity of binding for each probe. Sequential washes where an autoradiograph was developed between each stage of washing were relatively unsuccessful. The film development time varied depending upon the experiment but generally an initial autoradiograph film was developed after 5 minutes exposure at room temperature. Exposure times for subsequent autoradiographs could be adjusted dependent upon the result of the 5 minutes exposure.

The dot blot hybridisation results for the catalytic domain clones are shown in Figure 7.5(b). The salt concentration was taken to 0.5 x SSC and the film was developed after 10 minutes exposure at room temperature. Shown in Figure 7.5.(b) are the clones to which each of the specific probes should bind. Overall, an exceptional level of specificity was achieved, showing that the probes bind as predicted. The probes hybridised to four *PRT1* clones that were not 100% specific to the clones (Table 7.4).

Probe	Square	Reason
Geno	E4	Clone sequence differs from geno probe by 2 bp
Geno	G2	An anomaly
cDNA	C4	Clone sequence differs from cDNA probe by 2 bp
MID	B6	Clone sequence differs from MID probe by 1 bp

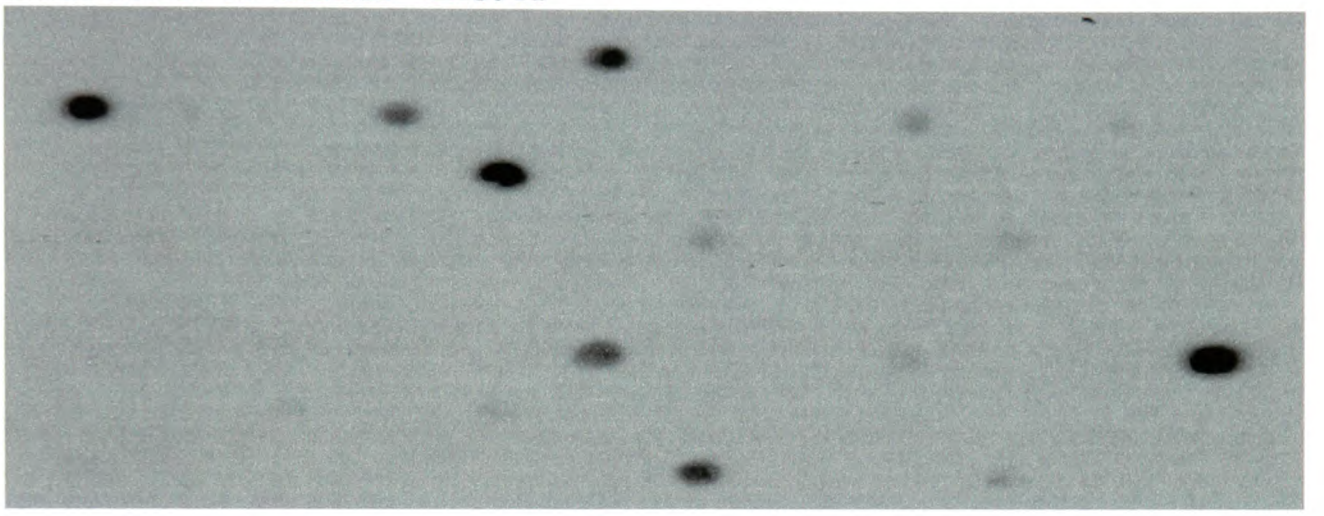
Table 7.4. The explanation for binding of the different specific probes to semi-homologous clones.

To prove the specificity of the probes, 160 DNA clones from the *PRT1* assay, both plasmid and PCR products, were used as templates. Two membranes containing 80 clones each were prepared. The same variables were considered for hybridisation, DNA quality, PCR or plasmid template, stringency conditions and autoradiograph development time. Figure 7.6 shows the results for the cDNA specific probe, which was studied in detail. Again, the probe bound to all the homologous clones. The clones G3, G5 and H1 are PCR products so

Predicted Results

	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

First Set of 80 Clones - Blot A



Second Set of 80 Clones - Blot B

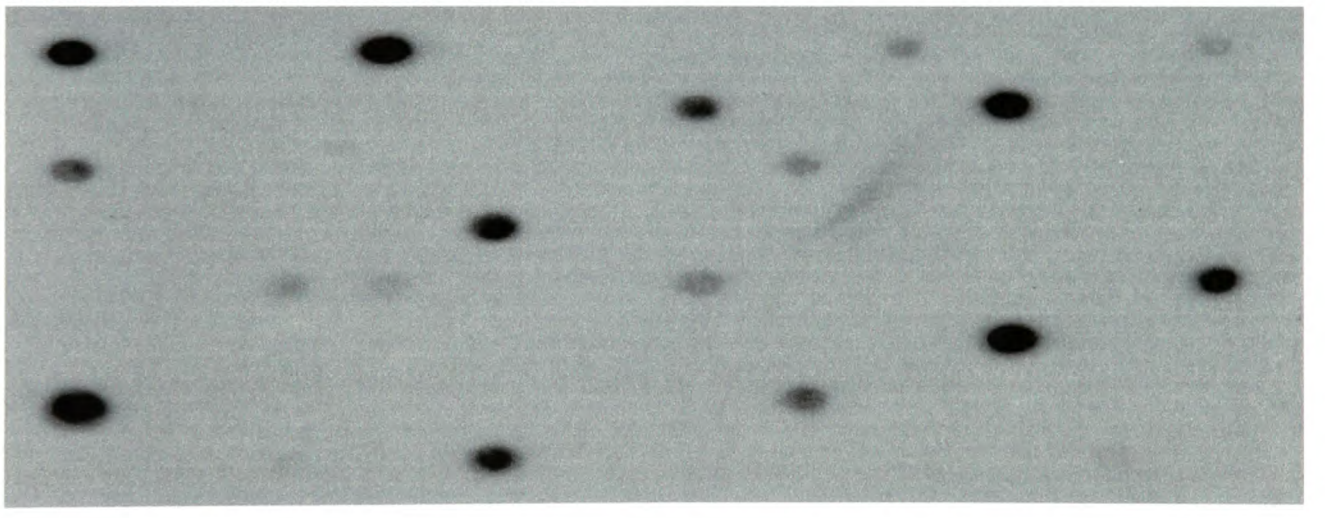


Figure 7.6. Dot blot hybridisation for 160 *PRT1* catalytic domain clones using cDNA specific probe The table is a schematic representation showing which clones the cDNA specific probe should bind to. Clones that the probe should bind to are highlighted in: first set of 80 clones, orange, second set of 80 clones, yellow. The hybridisation results are shown, blot A is the first set of 80 clones and blot B is the second set of 80 clones. The blots were washed to a salt concentration of 0.25 x SSC and the autoradiograph was developed after 30 minutes exposure at room temperature.

hybridise less strongly in blot a. The probe also bound to 7 clones with incomplete homology.

A mini-dot blot was prepared using the 7 clones with incomplete homology. Homologous clones and non-homologous clones were used as controls. Again, the same 7 clones were found to bind to the cDNA probe, even when washed to a high stringency of 0.125 x SSC. All 7 clones were plasmids, rather than PCR products, so the DNA should be of good quality. The semi-homologous clones were resequenced in order to verify that the sequence difference between the probe and the clone was correct. F10 from blot B was found to have the same sequence as the cDNA probe whereas the remaining clones were correctly sequenced initially. Overall, it can be concluded that an outstanding level of almost 100% binding to homologous clones was achieved with the cDNA probe, only 6/160 semi-homologous clones (3.8%) bound the probe, with the clone sequence being different from the probe by only 1 bp to 2 bp. The cDNA probe is specific for its subgroup of *PRT1* assay clones. Experiments using the geno and MID probes (section 7.3.1) produced similar results to the cDNA probe where almost 100% of hybridisation positives were due to the probe binding to homologous clones.

7.3.3. Using the Specific Probes as Primers in PCR amplification.

The 3 specific oligonucleotide probes can be used as primers for PCR amplification. This provides another test for the specificity of the probes; only the *PRT1* alleles from homologous clones should be amplified in the reaction. To test the DNA preparations to be used as templates, the preparations were first amplified using the universal primers PcProt1 and PcProt3. A product was generated from each template (data not shown). Different populations of complex rat-derived *Pneumocystis* DNA were amplified with the specific probes (geno-HA1, cDNA-HA2, and MID-HA3) paired with PcProt3. Figure 7.7a

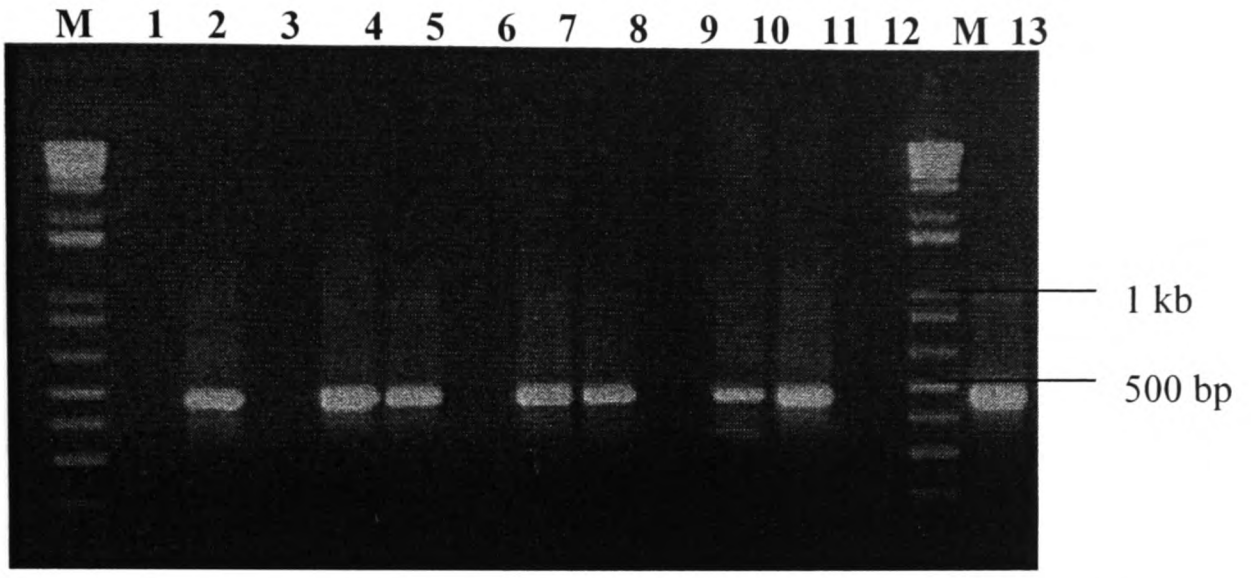


Figure 7.7.(a) Electrophoretic separation of the products of PCR amplification using genomic rat-derived *Pneumocystis* DNA and the primers HA1 and PcProt3. M: 1Kb plus DNA Ladder. The templates are; lanes 1 and 2: rat-derived *Pneumocystis* genomic DNA extracted from population 848.3, undiluted and x5 dilution respectively; lanes 4 and 5: rat-derived *Pneumocystis* genomic DNA extracted from population 876.4, undiluted and x5 dilution respectively; lanes 7 and 8: rat-derived *Pneumocystis* genomic DNA extracted from population 901.4, undiluted and x5 dilution respectively; lanes 10 and 11: rat-derived *Pneumocystis* genomic DNA extracted from population 910.4, undiluted and x5 dilution respectively; lanes 3, 6, 9 and 12: negative control using dH_2O in place of a template. Each population contained a copy of the allele.

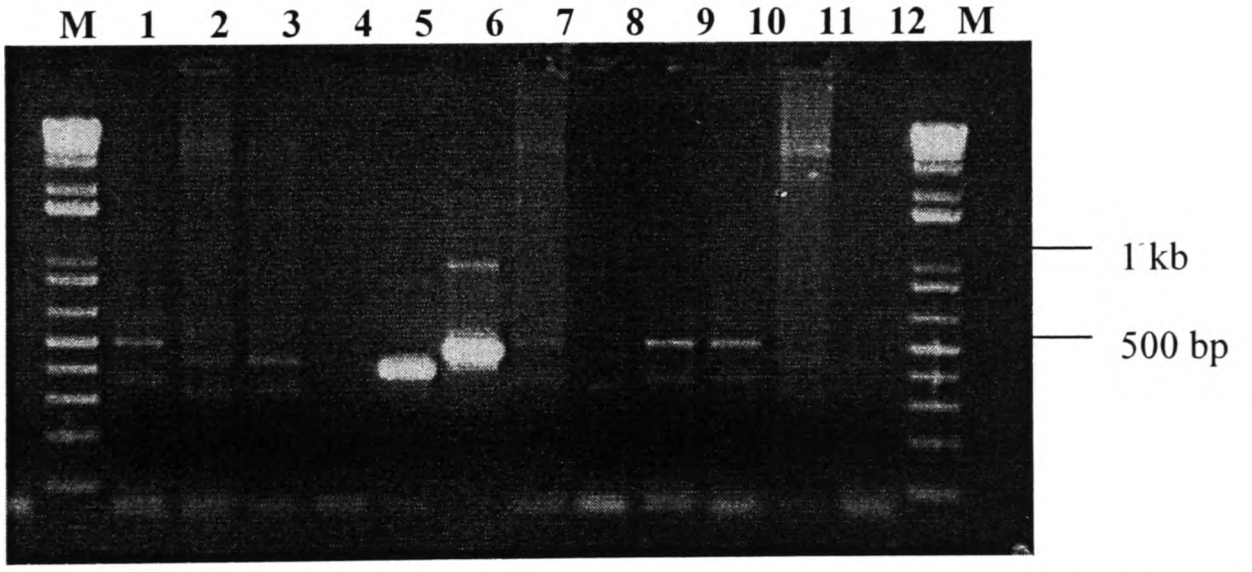


Figure 7.7.(b) Electrophoretic separation of the products of a PCR using bacterial clones of the *PRT1* catalytic domain as templates and the primers HA3 and PcProt3. M: 1Kb Plus DNA Ladder. The following templates were used; lane 1: clone ea19g7a (specific to the geno probe); lane 2: clone ea19cco (specific to the cDNA probe); lane 3: clone 1984c3c (specific to the cDNA probe); lane 5: clone ea19ccs (specific to the MID probe); lane 6: clone ea19g7d (specific to the MID probe); lane 7: inocg7d (borderline specificity to the geno probe); lane 9: ea15g7d (borderline specificity to the geno probe); lane 10: ea24g7s (borderline specificity to the cDNA probe); lane 11: ea17c2f (borderline specificity to the cDNA probe); lanes 4, 8 and 12: negative control using dH_2O in place of a template. HA3 and PcProt3 only amplified the clones containing inserts specific to the MID probe.

shows the results for HA1 and PcProt3. Alleles specific to the geno probe are present in each of the complex populations. A product was not generated from the undiluted sample of population 848.3. The other two specific primers produced the same result. Therefore, if the four populations, 848.3, 876.4, 901.3 and 910.4 were used as templates in a Southern hybridisation, at least one copy of the relevant allele is present as a target for each of the specific probes.

The genomic DNA from the 7 populations from the first set of simple populations, EA15, EA17, EA19, EA21, EA24, inoc and 1984 were then analysed. From the *PRT1* assay, the genomic majority sequence type was identified in each population but genomic copies of the cDNA majority sequence type were only identified in EA17, EA19 and 1984. A product was amplified from each population with all 3 specific primers, demonstrating that the majority sequence types are present as genomic *PRT1* copies in all 7 populations. Plasmid clones was analysed, each with different levels of homology to the probes (Table 7.5.).

Clone	PcProt1 and PcProt3	HA1 and PcProt3	HA2 and PcProt3	HA3 and PcProt3
EA19g7a (G+)	✔	✔	-	-
EA19cco (C+)	✔	-	✔	-
1984c3c (C+)	✔	-	✔	-
EA19ccs (M+)	✔	-	-	✔
EA19g7d (M+)	✔	-	-	✔
Inocg7d (Gih)	✔	-	-	-
EA15g7d (Gih)	✔	✔	-	-
EA24g7s (Cih)	✔	-	✔	-
EA17c2f (Cih)	✔	-	✔	-

Table 7.5. PCR amplification of plasmid clones amplified by PcProt3 with PcProt1, HA1, HA2 or HA3.

G+, homologous to the geno probe, C+, homologous to the cDNA probe, M+, homologous to the MID probe, Gih, incomplete homology to the geno probe and Cih, incomplete homology to the cDNA probe. ✔ PCR was successful, - the DNA was not successfully amplified.

Figure 7.7.(b) shows the results for primers HA3 and PcProt3. The two clones that are homologous to HA3 are successfully amplified while the other clones only result in misprimed products. HA1 and HA2 produce similar results (data not shown). HA1 amplifies the HA1 specific clone and one of the HA1 clones with incomplete homology while HA2 amplifies both specific clones and both clones with incomplete homology. The probes are specific enough to be used as primers in a PCR screen (section 2.8.) in order to detect the presence of majority sequence types.

7.3.4. Hybridisation of Specific Probes to Complex Rat-Derived *Pneumocystis* DNA.

There are many parameters to consider when preparing complex genomic DNA for hybridisation experiments. The quality and amount of DNA has to be considered. Rat derived *Pneumocystis* DNA mainly consists of two special forms, *Pneumocystis carinii* and *Pneumocystis ratti*. *P. ratti* genomic DNA has been shown by previous studies in this laboratory to interfere with the hybridisation of probes that have been designed to bind specifically to *P. carinii* DNA. Therefore, only pure *P. carinii* DNA preparations should be used, examples are 848.3, 876.4, 901.3 and 910.4. Also, it is crucial that a sufficient amount of DNA is used in experiments.

Digesting the DNA involves many factors; which enzyme and at what concentration, length of time for digestion, and whether the genomic DNA should remain embedded in agarose or be ethanol precipitated first. Each of these variables was investigated and the optimal procedure is described in (section. 2.11). *Dra*1 was found to be the most effective enzyme. It has an AT rich restriction site, TTT/AAA. The genome of rat-derived *Pneumocystis* DNA is very AT rich so *Dra*1 restriction sites are plentiful, resulting in a good size distribution of fragments when digested. Initially, the DNA was ethanol precipitated and then digested, however, this lead to variable digestion. More consistent digestion was obtained when the

agarose block containing the DNA was equilibriated in digestion buffer and the DNA was digested while embedded in the agarose. The agarose was then melted before the DNA was run on a gel.

Oligonucleotides were used as hybridisation probes in preference to double stranded DNA. More specificity could be achieved with an oligonucleotide even though they do not hybridise as effectively as double stranded DNA. Terminal transferase was used for radiolabelling the oligonucleotides. For a high level of stringency of washing, in order to maximise specificity of the probes, a salt concentration of 2 x SSC was appropriate. Figure 7.8 shows the gel and blot for a typical *PRT1* Southern analysis. The oligonucleotide probe binds to more copies of *PRT1* genes than the double stranded DNA probe because it has homology to more *PRT1* copies than the double stranded probe.

Figure 7.9, Figure 7.10, and Figure 7.11 show the results for the specific probes, geno, cDNA and MID (section 7.3.1) hybridising to complex rat-derived *P. carinii* DNA. Three individual gels were prepared, each loaded with the same *Dra*1 digested genomic DNA samples from four different complex rat-derived *P. carinii* populations, one of which has two preparations on the agarose gels, 848.3 (x2), 876.4, 901.3 and 910.4. 3 pairs of samples consisting of a plasmid and a PCR product, each 100% homologous to one of the specific probes were also loaded, one pair per gel. They act as positive controls for the hybridisation making each gel specific for one probe. For each gel, the rat-derived *Pneumocystis* DNA was concentrated and fully digested. The positive controls diluted 1/10, were still visible on the agarose gels. Each of the 3 individual membranes was first hybridised to the relevant specific probe. In each case the specific probes bound to the positive controls, proving that hybridisation had been successful. The cDNA probe bound strongly to a series of *PRT1* alleles within population 910.4 and weakly to the same series in 901.3. The geno probe also

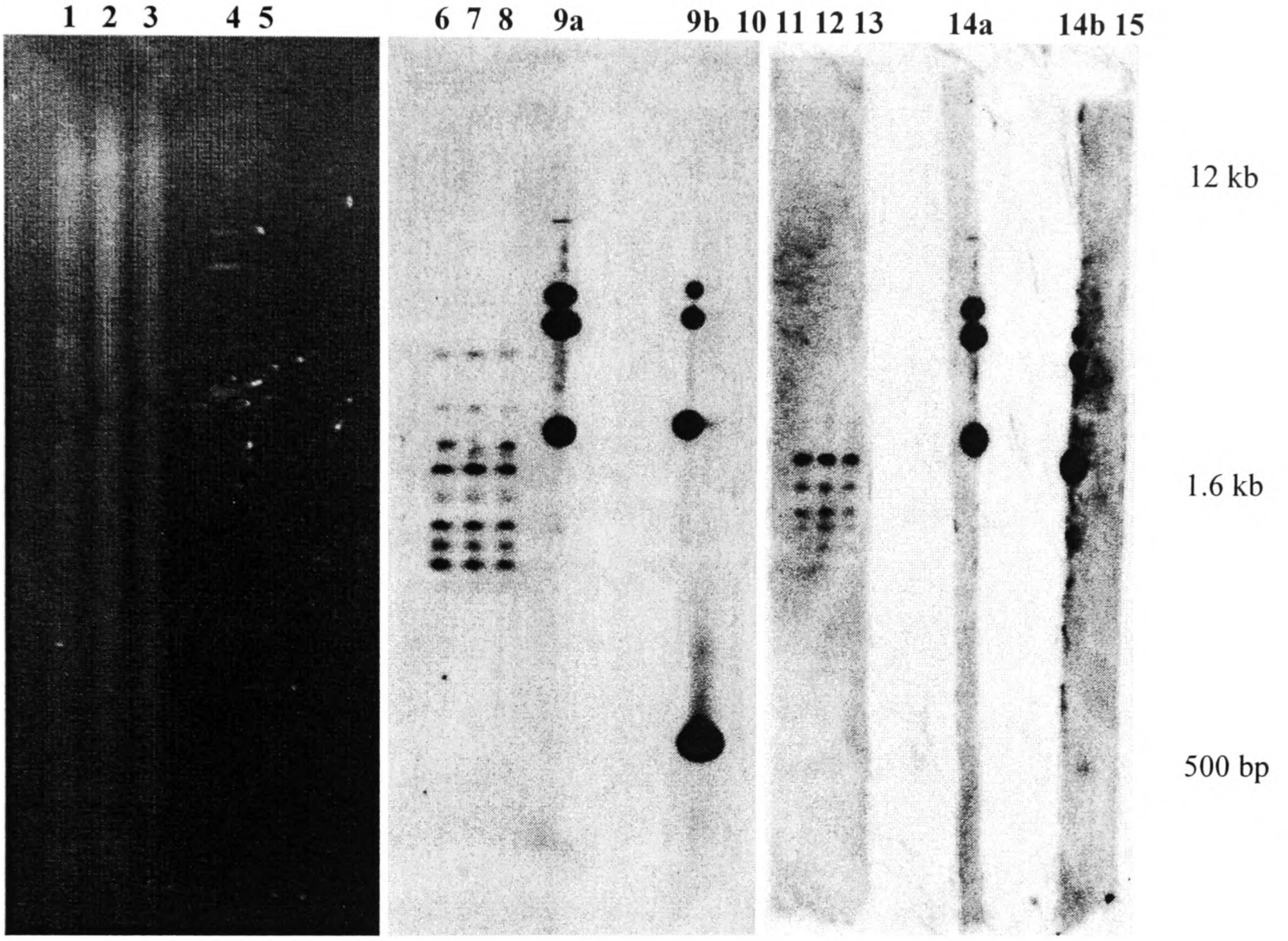


Figure 7.8. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with an oligonucleotide and a PCR amplication product. Lanes 1, 2, 3, 6, 7, 8, 11, 12, and 13: rat-derived *Pneumocystis* genomic DNA, 848.3 digested with *Dra*1; lanes 4, 9a, 9b, 14a and 14b: plasmid EA15c1m whose insert contains a catalytic domain generated by the primers PcProt1 and PcProt3; lanes 5, 10 and 15 catalytic domain PCR product EA15c4m, generated by the primers PcProt1 and PcProt3. Lane 1, 2, 3, 4 and 5:ethidium bromide stained gel; lanes 6, 7, 8, 9a, 9b and 10: Southern Blot probed with oligonucleotide PcProt3; lanes 11, 12, 13, 14a, 14b and 15 Southern Blot probed with catalytic domain PCR product EA19g7c generated by the primers PcProt1 and PcProt3. The membranes were washed to a stringency of 2 x SSC and the autoradiographs were developed after exposure for 3 days at – 80°C.

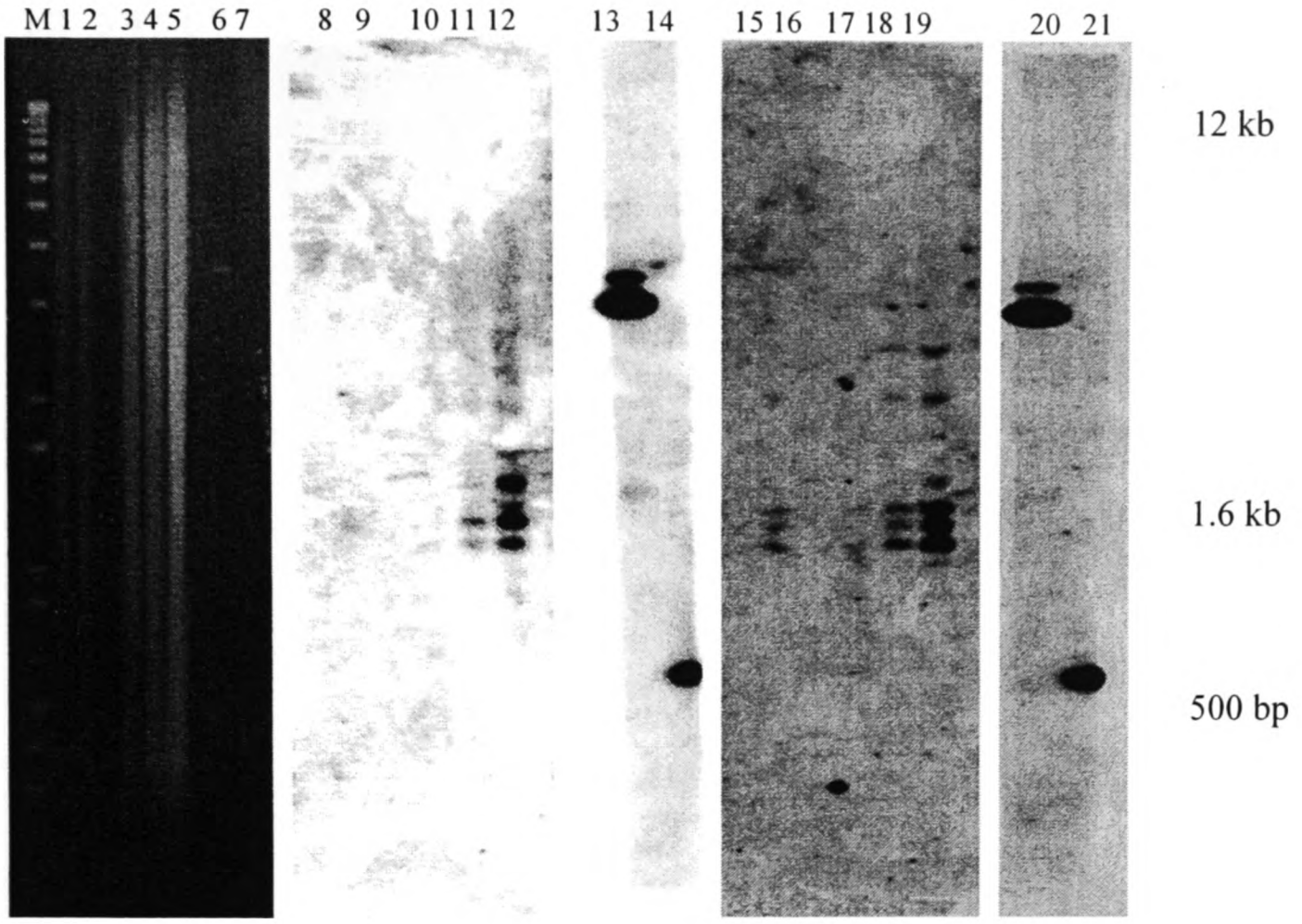


Figure 7.9. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with geno and PcProt3. M, 1 kb plus DNA marker ladder. Lanes 1, 8 and 15: rat-derived *Pneumocystis* genomic DNA 848.3/2; lanes 2, 9 and 16: rat-derived *Pneumocystis* genomic DNA 848.3/3; lanes 3, 10 and 17: rat-derived *Pneumocystis* genomic DNA 876.4/1; lanes 4, 11 and 18: rat-derived *Pneumocystis* genomic DNA 901.3/1; lanes 5, 12, 19: rat-derived *Pneumocystis* genomic DNA 910.4/1. All of the genomic DNA has been digested with *Dra*1. Lanes 6, 13 and 20: plasmid ea19g7a, diluted 1 in 10, homologous to the geno probe; lanes 7, 14 and 21: PCR product ea19g7a, diluted 1 in 10, homologous to the geno probe. Lanes 1–7: ethidium bromide stained gel; lanes 8–14: Southern Blot probed with the specific geno probe; lanes 15–21 Southern Blot probed with the universal probe PcProt3. The blots were washed to 2 x SSC

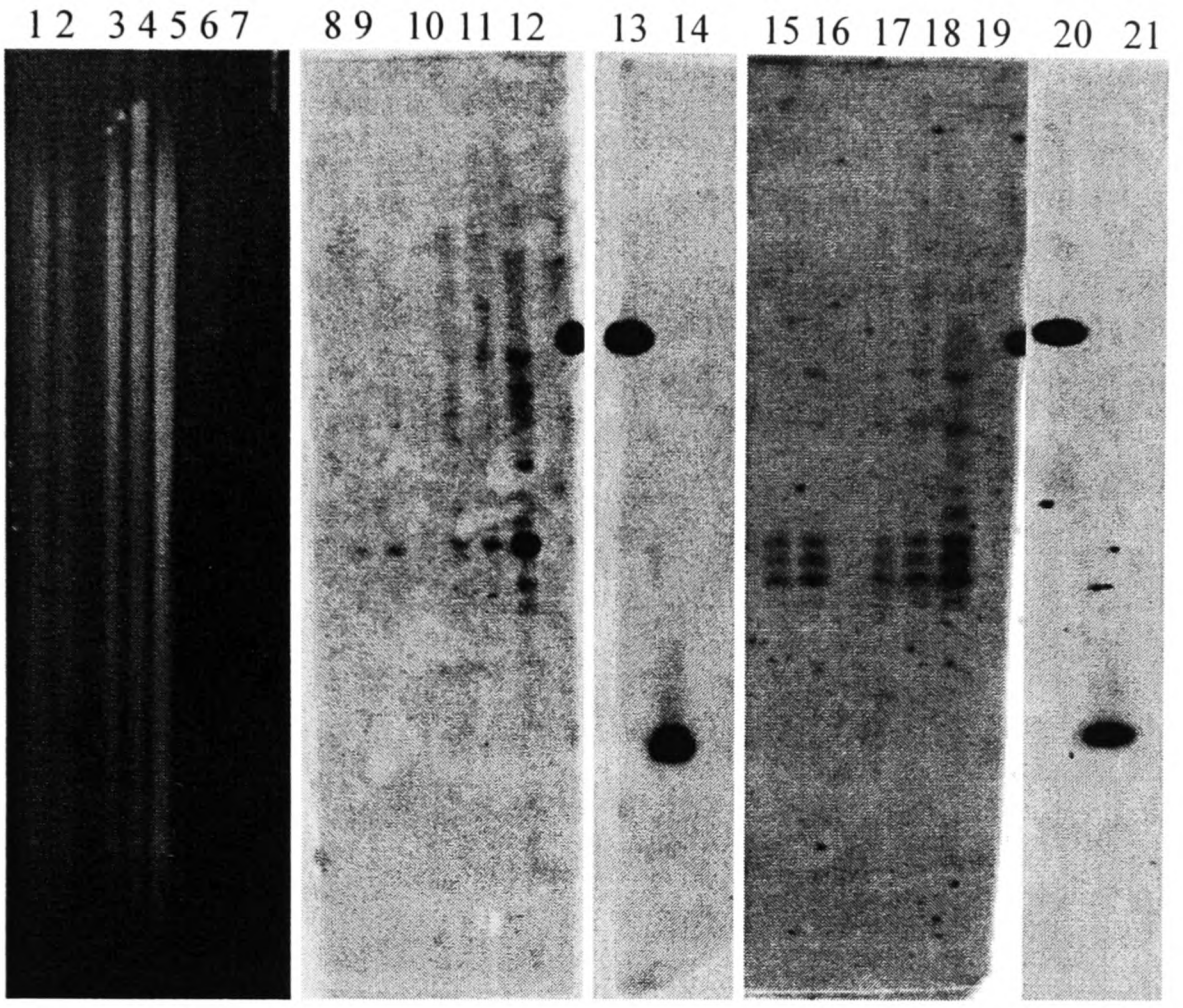


Figure 7.10. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with cDNA and PcProt3. Lanes 1, 8 and 15: rat-derived *Pneumocystis* genomic DNA 848.3/2; lanes 2, 9 and 16: rat-derived *Pneumocystis* genomic DNA 848.3/3; lanes 3, 10 and 17: rat-derived *Pneumocystis* genomic DNA 876.4/1; lanes 4, 11 and 18: rat-derived *Pneumocystis* genomic DNA 901.3/1; lanes 5, 12, 19: rat-derived *Pneumocystis* genomic DNA 910.4/1. All of the genomic DNA has been digested with *Dra*1. Lanes 6, 13 and 20: plasmid 1984c3c, diluted 1 in 10, homologous to the cDNA probe; lanes 7, 14 and 21: PCR product ea19cco,diluted 1/10 homologous to the cDNA probe. Lanes 1–7: ethidium bromide stained gel; lanes 8–14: Southern Blot probed with the specific cDNA probe; lanes 15–21 Southern Blot probed with the universal probe PcProt3. The blots were washed to 2 x SSC.

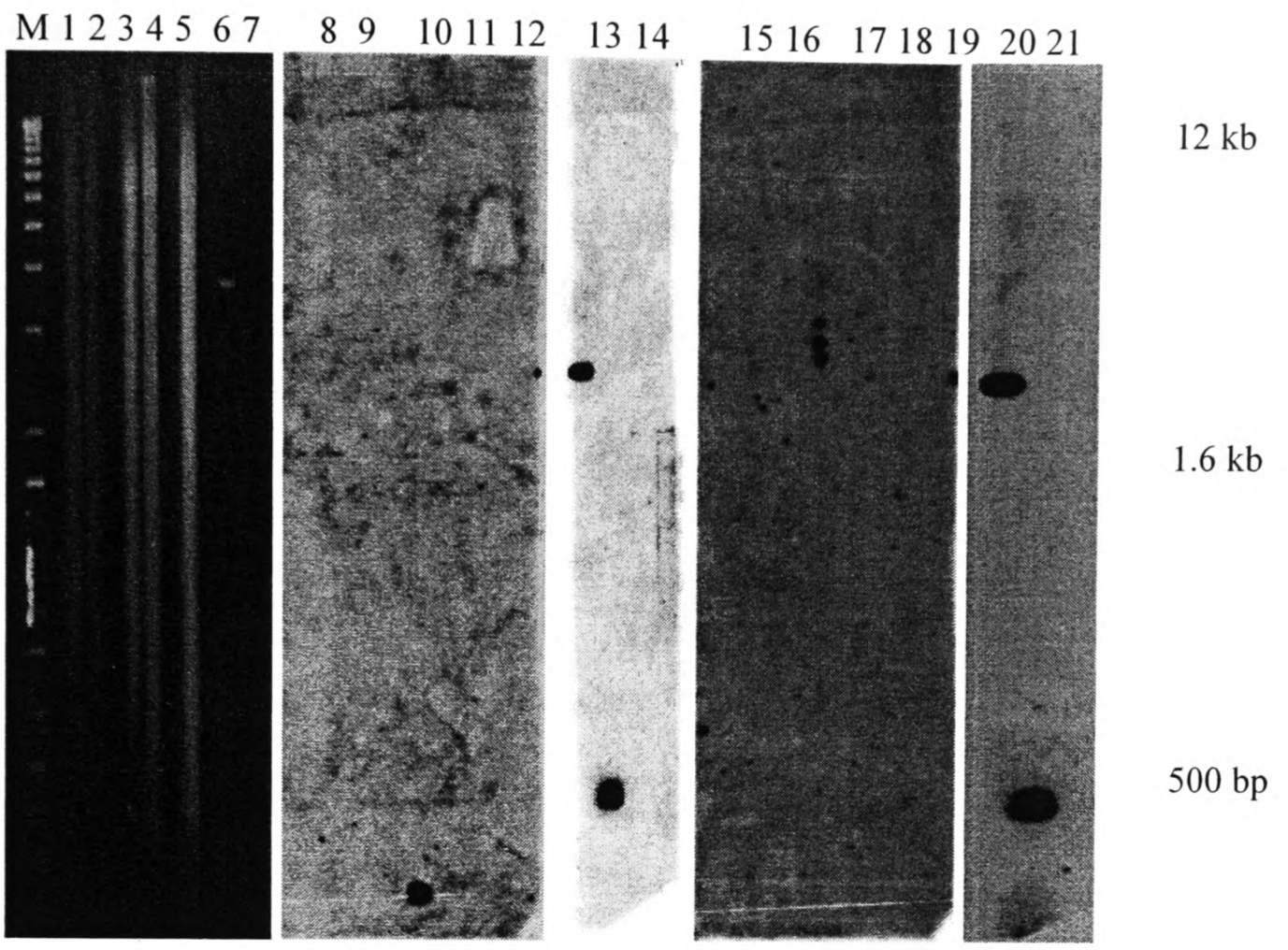


Figure 7.11. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with MID and PcProt3. M, 1 kb plus DNA marker ladder. Lanes 1, 8 and 15: rat-derived *Pneumocystis* genomic DNA 848.3/2; lanes 2, 9 and 16: rat-derived *Pneumocystis* genomic DNA 848.3/3; lanes 3, 10 and 17: rat-derived *Pneumocystis* genomic DNA 876.4/1; lanes 4, 11 and 18: rat-derived *Pneumocystis* genomic DNA 901.3/1; lanes 5, 12, 19: rat-derived *Pneumocystis* genomic DNA 910.4/1. All of the genomic DNA has been digested with *Dra*1. Lanes 6, 13 and 20: plasmid ea19g7d, diluted 1 in 10, homologous to the MID probe; lanes 7, 14 and 21: PCR product ea19gccs, diluted 1 in 10, homologous to the MID probe. Lanes 1–7: ethidium bromide stained gel; lanes 8–14: Southern Blot probed with the specific MID probe; lanes 15–21 Southern Blot probed with the universal probe PcProt3. The blots were washed to 2 x SSC.

bound with the same relative strength with regard to the DNA populations, to a series of *PRT1* alleles but this was a different series to the cDNA probe. This clearly demonstrates that the two probes are specific to different *PRT1* alleles within complex rat-derived *P. carinii* populations.

The membranes were then stripped and re-hybridised with the universal probe PcProt3. The universal probe bound to the same series of *PRT1* alleles on both the cDNA and geno membranes. Theoretically, this series should incorporate all *PRT1* alleles. More bands were seen on the autoradiographs with the universal probe than with the specific probes, confirming that the specific probes bind to subgroups of *PRT1* alleles.

PcProt3 bound to the positive controls on the membrane specific for the MID probe, indicating that hybridisation was successful. However, it did not bind to the complex rat derived *P. carinii* DNA, implying that the DNA was not of a high enough quality for successful hybridisation or was poorly transferred to the membrane. The MID probe did not bind at all to the complex genomic DNA but did to the controls.

7.4. Discussion of DNA Hybridisation Experiments to Characterise the Majority Sequence Types.

Three specific probes have been successfully designed, one to bind predominantly to the genomic majority sequence type, another to bind predominantly to the cDNA majority sequence type and a third probe to bind to a specific sequence to monitor specificity of hybridisation. The specificity of these probes has been proven by dot blot and PCR amplification. The geno specific probe and the cDNA specific probe each bind to their own series of *PRT1* alleles when hybridised to complex rat-derived *Pneumocystis* genomic DNA.

DNA quality appeared to be the main limiting factor of the many parameters and steps involved in these experiments. The probes have been designed to bind to *P. carinii* DNA. The preparations of *P. carinii* are potentially contaminated by protein and cellular tissue from both the *Pneumocystis* and the rat and from rat DNA. The oligonucleotides bound more efficiently to the plasmid DNA rather than the PCR products even though equal amounts of both were used. Efficiency of transfer may have been greater for the larger sized plasmids.

Dot blot hybridisation confirmed that the probes are almost 100% specific for the majority sequence types however they are also specific for alleles which are not the majority sequence types (section 7.3.1). Thus, it is possible that other alleles in the complex *Pneumocystis* genomic DNA, similar to the majority sequence types might have been identified as well as the majority sequence types. However, although the experiments do not offer conclusive proof, it is likely that the majority sequences types have been identified from the Southern blots.

More probes specific to particular sequences or groups of sequences need to be designed and tested in order to create a series of probes from which the relevant sequence types can be identified under experiment. The hybridisation conditions need to be further optimised so that genomic DNA from the simple populations can be reliably and reproducibly studied. The full gene sequences of the majority sequence types could be characterised using the specific probes as primers for PCR amplification in combination with universal primers both upstream and downstream of the catalytic domain.

The specific probes bound to a series of *PRT1* alleles rather than one or two copies. There are two possible explanations for this. Many copies of the majority sequence types might be present within a *Pneumocystis* organism. Alternatively, it is possible that the single copy of

the majority sequence types occupies a dynamic position and lies in a subtly different genomic context within each organism so many different copies appear to be identified within a population.

7.5. Characterisation of 5'UTR of *PRT1* Genes using 5' RACE

The third study in this part of the project examines how *PRT1* genes might be regulated. For the *MSG* gene to be expressed, it has to be located next to the UCS expression site. The UCS was identified by 5' RACE (Wada *et al.,* 1995). 5' RACE, using primers specific to *PRT1* genes, was carried out to determine if *PRT1* genes had a single expression site equivalent to the UCS locus.

7.5.1. Sequence Data for the 5'UTR of *PRT1* Genes

In initial experiments, control RNA (GibcoBRL) and control primers (GibcoBRL) were used to optimise the 5' RACE conditions (section 2.12.4). Figure 7.12 shows the results for the *PRT1* 5' RACE reaction using *Pneumocystis* RNA and primers specific to the 5' region of the *PRT1* genes. The 5' RACE reaction for the *PRT1* genes produced a smear of products ranging over the expected size of 300 bp to 400 bp. PcProt85, one of the nested primers, binds approximately 300 bp downstream of the start codon. Therefore, any PCR product larger than 300 bp should contain a 5'UTR. The control reactions for the 5' RACE are clear apart from those minus TdT, which generated a product of approximately 700 bp. This is most likely the result of mispriming rather than contamination because the negative controls, where dH_2O is used in place of a template, are clear.

The 5' RACE products were cloned. A PCR screen (section 2.8) using the primers AUAP and PcProt85 identified 10 positive bacterial clones. Direct sequencing was attempted on the

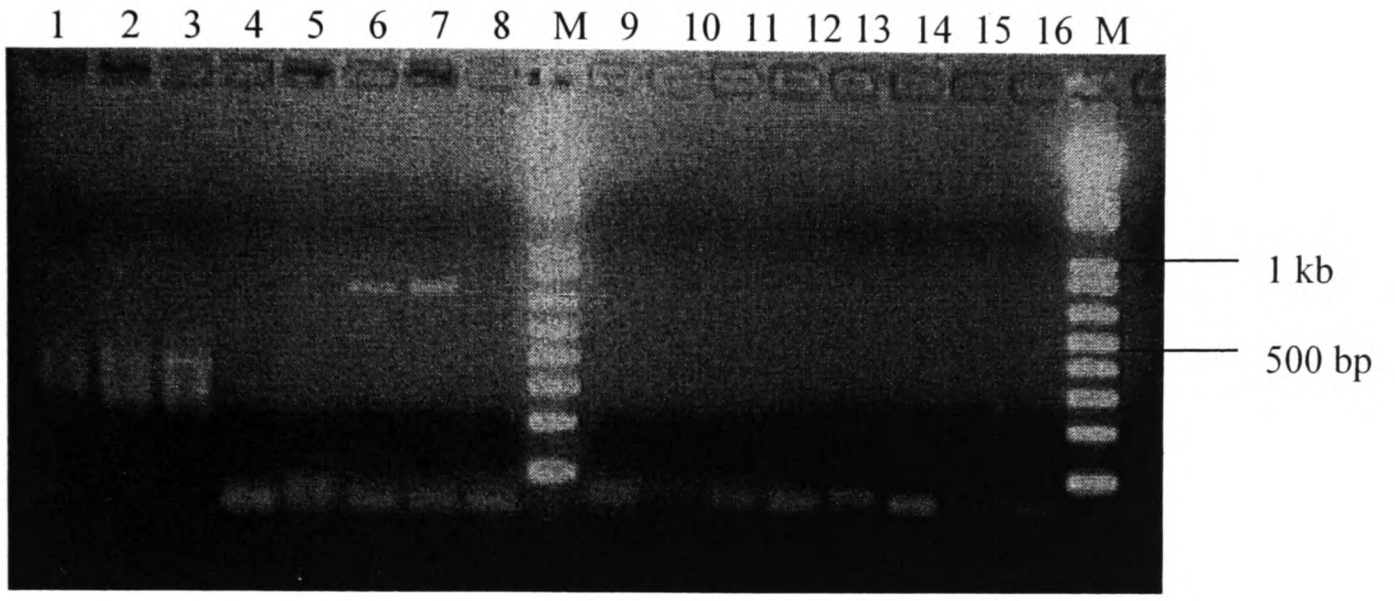


Figure 7.12. Electrophoretic separation of 5' RACE products. 5' RACE second round amplification using the primers PcProt85 and AUAP and the products from the first round amplification. M, 1 kb DNA ladder. All experimental samples contain rat-derived *Pneumocystis* RNA. The different conditions are; lanes 1, 2 and 3: plus RT plus TdT with the rat-derived *Pneumocystis* RNA undiluted, diluted 1 in 10 and diluted 1 in 100 respectively; lanes 5, 6, and 7: plus RT minus TdT with the rat-derived *Pneumocystis* RNA undiluted, diluted 1 in 10 and diluted 1 in 100 respectively; lanes 9, 10, and 11: minus RT plus TdT with the rat-derived *Pneumocystis* RNA undiluted, diluted 1 in 10 and diluted 1 in 100 respectively; lanes 13, 14 and 15: minus RT minus TdT with the rat-derived *Pneumocystis* RNA undiluted, diluted 1 in 10 and diluted 1 in 100 respectively; lanes 4, 8, 12 and 19; negative controls using dH_2O in place of a template. Lanes 6 and 7 contain a PCR product of ≈ 700 bp. The product is the result of mispriming. Lanes 1 –3 contain products from the 5' RACE. There are two predominant bands, one of ≈ 300 bp, which is equivalent to cDNA without a 5'UTR. The second is a smear of products ≈ 400 bp long. This correlates to a range of cDNA sequences with a 5'UTR region.

PCR products but this was only successful in one direction for 7 PCR products because the secondary structure of the polydCTP tail inhibited the sequencing reaction. The plasmids corresponding to the PCR products were purified and 6 were successfully sequenced in one direction using PcProt85 but again the polydCTP tail inhibited sequencing reactions using the primer AUAP. The sequence data from both the PCR and plasmid clones was combined to provide a consensus sequence for each 5'UTR clone. However, this sequence data is only from one strand when usually sequence data is obtained from both strands of the DNA being analysed (Tsolaki *et al.*, 1999).

In order, to generate more 5'UTR sequence data, EST cDNA clones (Smulian, 2001a) (section 2.6.3) with 5'UTR regions were analysed. EST clones coding for complete copies of *PRT1* genes had been identified in this laboratory. Plasmid DNA was purified from these clones and was digested with restriction enzymes to confirm the insert size. Ten EST clones contained full-length *PRT1* sequences. They were sequenced using the vector sequence primer T3 and PcProt58, which binds in the 5' region of the *PRT1* genes. Five of the published *PRT1* sequences have 5'UTR sequence data included (section 1.4.3.). These are the cosmid copies of *PRT1* genes (except 11A111 because the cosmid clone starts in the middle of this copy of a *PRT1* gene) and the Wada genomic *PRT1* clone, *ORF-3*.

Figure 7.13 is a ClustalW sequence alignment of the seven 5'UTR region of the 5' RACE products, the ten EST clones and five published *PRT1* sequences. The vector sequences of the EST clones and the 5' RACE products have been highlighted in orange. The alignment contains approximately 100 bp of sequence and there appears to be a high level of homology between them, yet, there is much individual nucleotide variation.

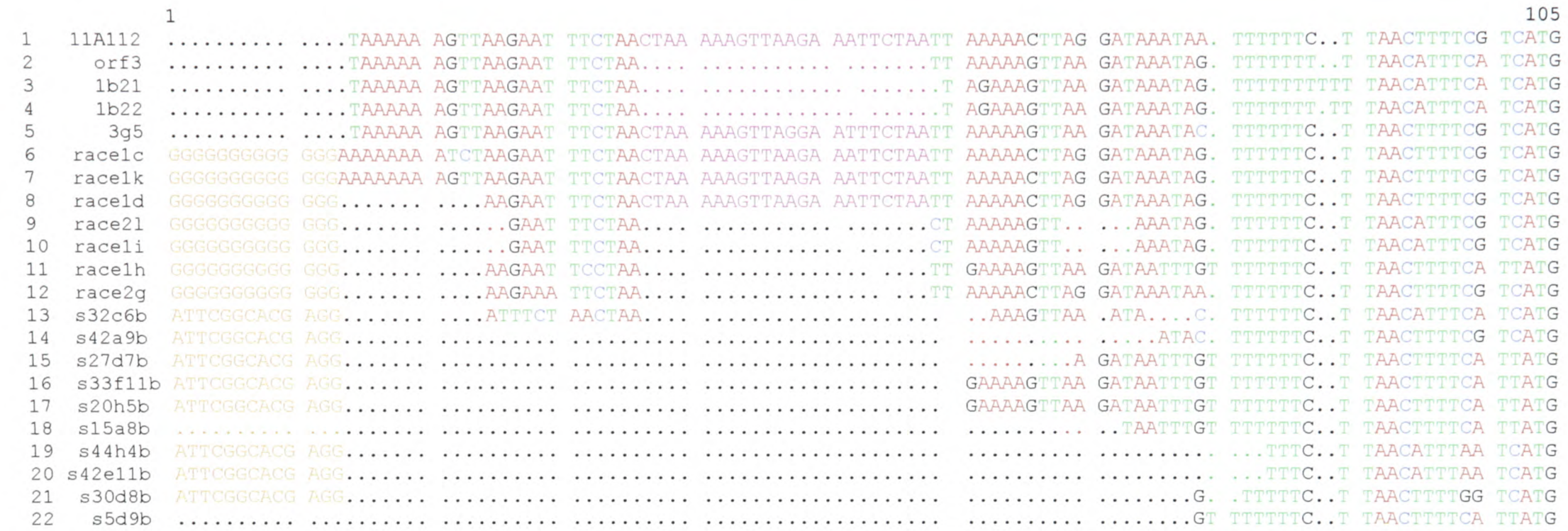


Figure 7.13. ClustalW sequence alignment of 5'UTR of published *PRT1* sequences, 5' RACE products and full length EST sequences. Vector sequence is highlighted in orange and the specific probe HA5 is highlighted in pink. The dNTPs are highlighted in dATP, red; dTTP, green, dCTP, blue and dGTP black. Sequences 1-5 are published *PRT*1 sequences, sequences 6-12 are 5' RACE products and sequences 13-22 are EST clones

The sequence alignment in Figure 7.13 shows no outstanding motif that distinguishes the cDNA 5' RACE and EST data from the published genomic data. There does not appear to be a single expression site for *PRT1* genes. Also, the 5'UTR does not maintain an open reading frame when transcribed. This contrasts with the UCS, which is transcribed and translated along with the *MSG* gene and is cleaved before the protein is surface expressed (Sunkin and Stringer, 1996). From the *PRT1* 5'UTR data, a stop codon is present in all three reading frames. The same observation was made with the 5'UTR of the gene encoding the mouse kexin, KEX1 (Lee *et al.*, 2000).

Apart from the individual nucleotide differences, there is an insertion of 22 bp from nucleotides 37 to 59 in both the genomic 5'UTR sequences, 11A112, 3G5, and the cDNA sequences race1c, race1k and race1d. The same insertion was observed in the intergenic region, between an *MSG* gene and a *PRT1* gene, of the genomic lambda (λ) DNA clones, in previous studies in this laboratory (Palmer, 2000a),where sequence analysis of 12 λ genomic DNA clones generated an alignment of roughly 350 bp of 5'UTR sequence and identified a further insertion of about 40 bp, which was found 200 bp upstream of the start codon.

The sequence consensus from Figure 7.13 was searched using the BLAST algorithm for potential transcription factor binding sites, using the database TRANSFAC® 6.0 with the MATCH™ search engine (Altschul *et al.*, 1997). The search was extended to include the intergenic region, of approximately 1000 bp, between the *PRT1* allele and the upstream *MSG* gene on the 3G5 cosmid. Thirteen potential transcription sites were found, one of which was the *Schizosaccharomyces pombe* transcription factor mat1-Mc. This sequence encodes for a binding site for a high mobility group box protein, for example the transcription factor Ste11. It is possible that an equivalent fungal transcription factor exists in rat-derived *Pneumocystis* and has some role in determining the expression of *PRT1* genes. Thus, a

5'UTR region for *PRT1* genes has been identified. There is no strong evidence to suggest that this sequence alignment is equivalent to the UCS, however, there are potential features that denote it to be a promoter region or expression site. Alternatively there is a promoter further upstream of the analysed sequence.

7.5.2. Hybridisation to Characterise the 5'UTR of *PRT1* Genes

Hybridisation experiments were designed to characterise the genomic context of the 5'UTR of the *PRT1* genes. It is possible that each *PRT1* genomic copy has the same 5'UTR. Alternatively, there may be subgroups of 5'UTR sequences just as there are subgroups of *PRT1* alleles, or less likely, there are unique 5'UTR sequences that have been cloned many times.

Two specific probes were designed to bind to subgroups of the 5'UTR sequences. The first, HA4, is a 23 bp oligonucleotide homologous to the insertion observed 400 bp upstream of the *PRT1* start codon. The second, HA5, is a 22 bp oligonucleotide homologous to the insertion between nucleotides 37 to 59 in the 5'UTR sequence alignment and is indicated on Figure 7.13.

Dot blots were prepared to determine the specificity of the probes. To generate PCR products specific to the probes, DNA prepared from the characterised λ clones were used as templates for PCR amplification using the primers PcProt59 and PcProt58. PcProt59 binds to the intergenic region between an *MSG* gene and a *PRT1* gene and PcProt58 binds downstream of the *PRT1* start codon, so the relevant region of DNA should be amplified. Of the 7 PCR products generated, 6 should have a region homologous to HA4 and 1 should be homologous to HA5. The 6 plasmids and 7 PCR products from the 5' RACE cloning were

also used in equimolar amounts as targets on the dot blots. All the clones were used in equal amounts and are summarised in Table 7.6.

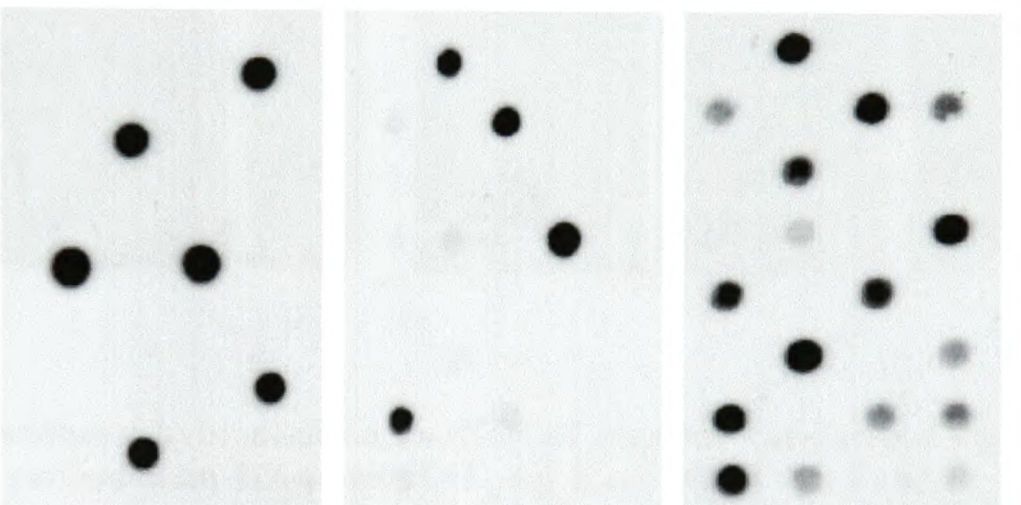
The results for the dot blots are shown in Figure 7.14. A schematic diagram has been included to show the clones that the specific probes were predicted to bind to. HA4 hybridised strongly to all 6 homologous clones. HA5 bound strongly to all homologous plasmid clones and the λ6.16 PCR product, which appeared on an agarose gel to be concentrated good quality DNA. The 5' RACE PCR products were of poor quality and hybridised weakly. The difference between the hybridisation of PCR and plasmid DNA is discussed in section 7.4. Similar results were produced for the universal primer PcProt58. The probe hybridised strongly to all the 5' RACE plasmids, except race1i, which did not hybridise at all. Strong hybridisation was observed with the λ PCR products, which like λ6.16 consisted of good quality DNA. The 5' RACE PCR products hybridised weakly as expected because of the poor quality of the DNA. PcProt58 is assumed to be a universal probe, however, it did not bind to the plasmid race1i. As this plasmid is good quality DNA, the most likely explanation is that the plasmid is not 100% homologous to PcProt58.

Having determined the specificity of the 5'UTR probes, the same probes were hybridised to rat-derived *Pneumocystis* genomic DNA. Difficulties were encountered in the efficient transfer of DNA as discussed in section 7.4. The control templates were not present on the same gel as the genomic *Pneumocystis* DNA samples. Figure 7.15 shows the results for the Southern hybridisation of *Dra*1 digested *Pneumocystis* genomic DNA probed with HA4 and HA5. HA5 produced results that were easiest to explain. The bands are comparable with hybridisations using the universal primer PcProt3. This implies that most alleles of *PRT1* genes have a 5'UTR containing the insertion of 22 bp, equivalent to HA5, upstream of the start codon. This insertion represents a common motif. HA4 has produced the unexpected

Number	Clone	PCR/Plasmid	HA4	HA5	PcProt58
1	λ2.12	PCR	X		X
2	λ3.1	PCR	X		X
3	λ3.3	PCR	X		X
4	λ4.3	PCR	X		X
5	λ5.3	PCR	X		X
6	λ6.5	PCR	X		X
7	λ6.16	PCR		X	X
8	Race1c	PCR		X	X
9	Race1c	Plasmid		X	X
10	Race 1k	PCR		X	X
11	Race 1k	Plasmid		X	X
12	Race 1d	PCR		X	X
13	Race 1d	Plasmid		X	X
14	Race1i	PCR			X
15	Race1i	Plasmid			X
16	Race1h	PCR			X
17	Race1h	Plasmid			X
18	Race2g	PCR			X
19	Race2g	Plasmid			X
20	Race2l	PCR			X

Table 7.6. Clones used in the 5'UTR dotblots. λ – PCR product generated from a lambda genomic DNA clone. HA4 and HA5 are probes specific to sequences in the 5'UTR and PcProt58 is a probe universal to the 5' region of *PRT1*. X – the probe is predicted to bind to that clone.

	1	2	3	4	5	6	7	8	9	10	11	12
A		13		14		13		14		13		14
B	8		11	4	8		11	4	8		11	4
C		1	20			1	20			1	20	
D	15	10		7	15	10		7	15	10		7
E	3		5		3		5		3		5	
F		17		16		17		16		17		16
G	9		12	2	9		12	2	9		12	2
H	19	6		18	19	6		18	19	6		18



HA4 HA5 PcProt58

Figure 7.14. Dot blots after hybridisation with probes specific to the *PRT1* 5'UTR
The schematic diagram (top) highlights the clones that the probes should bind to. Clones specific to HA4 are highlighted in yellow, HA5 in green and PcProt58 in blue with the subdivisions of PCR products highlighted in dark blue while plasmid clones are in light blue. The numbering of the clones is as table 7.7. The dot blots are shown below the schematic diagram. The blots were washed to a stringency of 0.25 x SSC and were developed for 1 hr at room temperature.

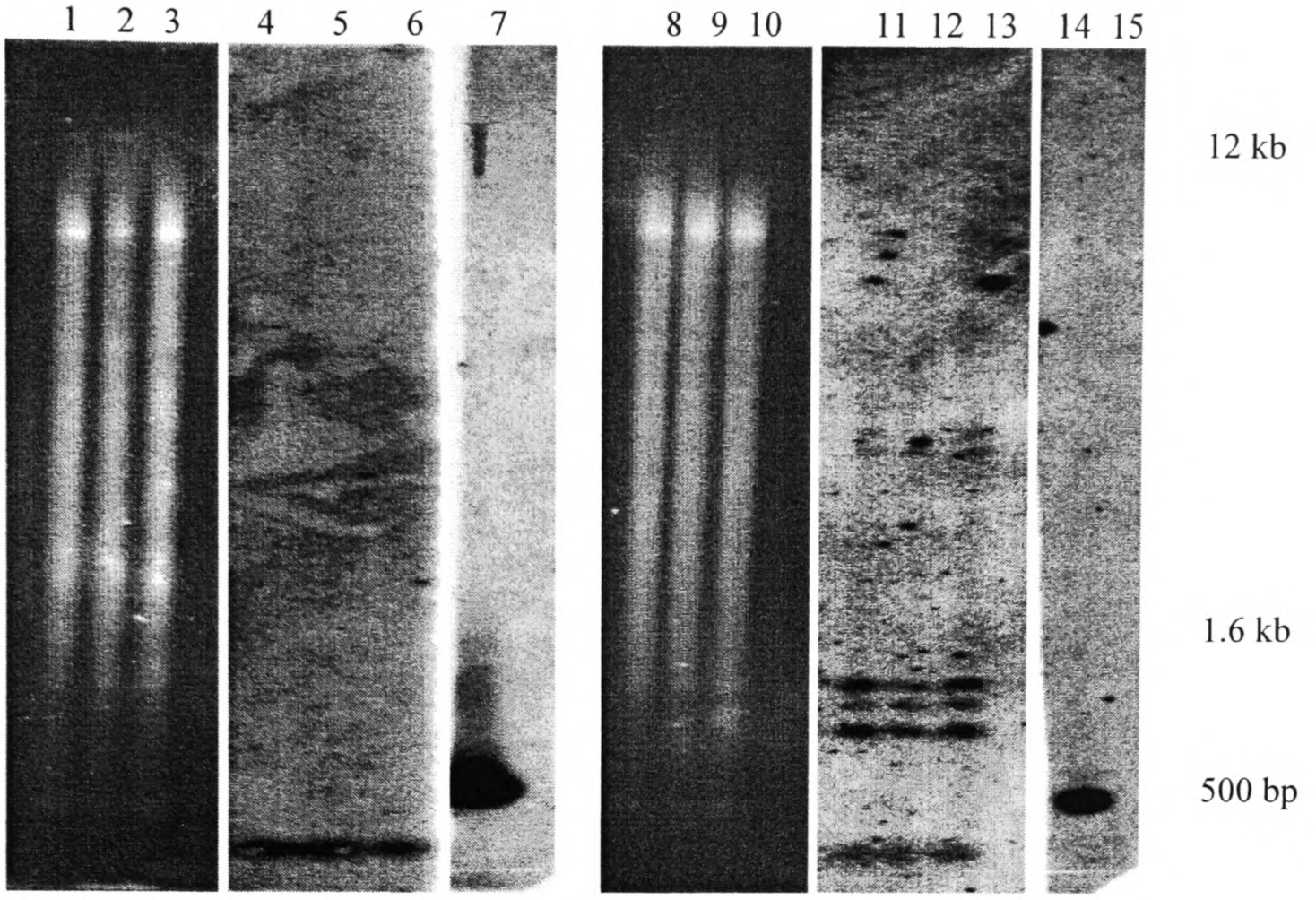


Figure 7.15. Southern hybridisation analysis of digested rat-derived *Pneumocystis* genomic DNA probed with HA4 and HA5. All lanes except 7, 14 and 15: rat-derived *Pneumocystis* genomic DNA 848.3 digested with *Dra*1; Lane 7: PCR product λ4.3 diluted 1 in 10; lane 14: plasmid race1d diluted 1 in 10; lane 15: PCR product race1d diluted 1 in 10. Lanes 1-3 and 8-10: ethidium bromide stained gels; lanes 4-7: Southern Blot probed with specific 5'UTR probe HA4; lanes 11-15: Southern Blot probed with specific 5'UTR probe HA5. The control agarose gels were run separately to the genomic DNA agarose gels. The blots were washed in 2 x SSC and autoradiographs were developed after 5 days exposure at -80°C.

result of a single band. This band correlates to a fragment of approximately 500 bp. Fragments this small from a genomic digest are normally generated from internal parts of the chromosomes where coding sequence is common and more restriction sites are present (section 4.1.1). This implies that HA4 could bind to the 5'UTR of a non-subtelomeric *PRT1* allele.

7.6. Discussion of Characterisation of 5'UTR of *PRT1* Genes using 5' RACE

The 5'UTR sequence was successfully obtained from *PRT1* genes. There was little sequence variation between both the genomic and cDNA 5'UTR clones, implying that most *PRT1* genes have a common 5'UTR sequence. It has yet to be proven that this common 5'UTR sequence contains an expression site for the transcription of *PRT1* genes. Thus, it is probable that *PRT1* genes are regulated by transcription factors, promoters or silencers that are further upstream than the sequence produced here which would be typical of eukaryotic transcription regulation. *PRT1* genes are not regulated in the same way as *MSG* genes, presumably indicating that the regulation of *PRT1* genes is not as restricted as *MSG* gene expression. Switching of the *PRT1* phenotype and multiple expression of *PRT1* genes would then be more common than switching of the *MSG* phenotype if more than one expression site was involved. HA4 bound to one single band of approximately 500 bp in the genomic hybridisation experiments. This is an intriguing result, as the band could correlate to an internal region of the chromosome. This could represent a different subgroup of *PRT1* alleles or a single copy of a *PRT1* gene that may not be surface expressed and the protein it encodes could have a different function to PRT1, such as the potential progenitor kexin gene (section 4.2).

Further experiments are required. Firstly, more extended 5'UTR sequence data is needed. More 5' RACE products could be analysed but the limitation of reverse sequencing would need to be solved. Also, the sequence corresponding to HA4 has only been shown to be homologous to genomic DNA, and has yet to be detected in cDNA (assuming that the 5'UTR extends that far). The genomic hybridisation experiments should be repeated using different and good quality rat-derived *Pneumocystis* DNA. To verify the quality of the DNA, a universal probe; either PcProt3 or PcProt58 could be used. PFGE blots would show the chromosomal positions at which HA4 and HA5 bind. This would clarify how many chromosomes containing *PRT1* genes the probes were binding to. The probes could also be used in Northern analysis to confirm if the 5'UTR really was part of the mRNA transcript.

5'UTR sequence data has been successfully generated from *PRT1* genes. There was little sequence variation between the 5'UTR clones, implying that most *PRT1* genes have a common 5'UTR sequence. This common region may contain promoter elements but more evidence is required to clarify this. The results from the genomic hybridisation experiments strongly suggest that a single expression site for a copy of the protease may exist located on a chromosomal internal region.

7.7. Summary of the Further Characterisation of *PRT1* Gene Expression

Three studies to further investigate the regulation of *PRT1* gene expression were undertaken. Using the *PRT1* assay, the first study analysed the rat-derived *Pneumocystis* populations generated from the second passage of the first set of simple populations. The results demonstrated that in rat-derived *Pneumocystis* simple populations passaging was possible, airborne transmission occurred and the cDNA majority sequence type was detected in the second set of simple populations. The second study characterised the genomic context of the

majority sequence types by utilising probes specific to the majority sequence types and using them in hybridisation experiments with genomic rat-derived *Pneumocystis* DNA. The majority sequence types appeared to be present as subgroups of the total *PRT1* alleles present in the genomic rat-derived *Pneumocystis* DNA. The hybridisation methodology will be useful for tracking *PRT1* alleles in future studies. The third study employed 5' RACE to characterise the 5'UTR of the *PRT1* gene. Substantial amounts of sequence data were produced and possible transcription factor binding sites were identified. Two specific probes were designed that bound to insertions within the 5'UTR. When bound to genomic rat derived *Pneumocystis* DNA, one probe generated a hybridisation pattern similar to a universal probe while the other appeared to bind to a single site. The combined data provides more data regarding the genomic context of expressed *PRT1* alleles.

8. Final Discussion

8.1. Epidemiology of *Pneumocystis jiroveci*

The epidemiology of *Pneumocystis* infection was investigated in three separate genotyping studies. The first investigated the possible transmission of *P. jiroveci* organisms between a mother and her infant who had PCP contemporaneously. Genotyping at three independent loci demonstrated that the infant had the same single *P. jiroveci* type as the majority *P. jiroveci* type infecting the mother. This implies that the infection was transmitted from the mother to the infant. Airborne transmission is most likely but other forms of transmission cannot be ignored. The results from the mother and infant samples are the first clear evidence of direct transmission between two PCP patients. The data was substantiated because three loci were genotyped rather than one.

The aim of the second study was to investigate asymptomatic carriage of *P. jiroveci* in HCW, the majority of whom worked in close contact with PCP patients. *P. jiroveci* DNA was detected in 11 HCW samples and of those it was possible to genotype 5 samples at the ITS locus. BAL samples from patients with PCP who were hospitalised at the same time were also genotyped at the ITS locus. It was not clear from the ITS genotyping results whether direct transmission of *P. jiroveci* had occurred between the patients and the HCW, however some form of transmission must have occurred for *P. jiroveci* DNA to be present in the lungs of the HCW. Thus, asymptomatic carriage of *P. jiroveci* was detected in the HCW who had occupational contact with PCP patients. This study was based upon a unique set of samples, because they were collected in the early 1990s when prevention of *P. jiroveci* transmission was limited. Thus, there was a higher likelihood, compared with other studies of this nature that direct transmission of *P. jiroveci* was occurring between PCP patients and

HCW but like most *P. jiroveci* transmission studies, the evidence for direct transmission between the two groups is incomplete.

The aim of the third study was to validate the assumption that respiratory samples are representative of the distribution of *P. jiroveci* genotypes within a human lung. *P. jiroveci* DNA present in an infected whole lung was genotyped and the distribution of the genotypes was determined. Mixed infection of *P. jiroveci* was demonstrated in the whole human lung tissue and the data from the ITS genotyping suggested that there was an uneven distribution of two ITS genotypes within the lung that may have originated from two foci of infection. Most *P. jiroveci* studies are limited to respiratory samples because whole human PCP infected lungs are a very rare resource. Therefore, the third study was a unique opportunity to look at the distribution and diversity of PCP infection within the whole lung. It was encouraging that two genotypes were identified in the lung tissue because these were the same number and type of genotypes as detected in respiratory samples. Unlike the majority of PCP samples, the lung tissue was from a Chilean child, immunosuppressed by cytotoxic drugs rather than by HIV infection, yet, *P. jiroveci* ITS genotypes equivalent to those from Northern hemisphere HIV adults were identified.

ITS genotyping was used in all three studies. The ITS locus is highly polymorphic and can be used to distinguish between many *P. jiroveci* strains (Lee *et al.,* 1998). Yet, the high degree of polymorphism suggests that this locus is evolving faster than other genes (Nimri *et al.,* 2002). For this reason, the ITS locus is useful for investigations into person-to-person transmission and recurrent infection but perhaps not for other genotyping applications involving more diverse strains. For this purpose epidemiological studies would require less sequence divergence and thus more genetically stable loci. Therefore, the ITS genotyping was useful for the mother and infant study and the whole lung study, yet may have had less

utility for the HCW study. The rate of sequence change at the ITS locus is unknown and if it was evolving during the time course of the study this would influence the results obtained. Due to the polymorphic nature of the multiple ITS sequences in most samples, PCR products amplified from the ITS locus have to be cloned prior to sequencing. In comparison, the other genotyping loci are less polymorphic so the PCR products can be sequenced directly and if a mixed infection is identified two nucleotides are detected at the polymorphic site. Therefore, ITS genotyping has the disadvantage of being more technically demanding and time consuming than genotyping at other loci but the ITS locus is one of the most polymorphic loci identified in *P. jiroveci* to date.

Multilocus genotyping involves genotyping two or more loci from one sample and combining the results to produce one overall genotype. Multilocus genotyping is a more powerful tool than genotyping at a single locus when used to describe differences between isolates of an organism because there is a greater sensitivity for detecting differences when several loci are examined concurrently. Multilocus genotyping when used in the mother and infant study strengthened the evidence for direct transmission. The samples from the other two studies could be genotyped at a range of other loci to determine if the same conclusions could be drawn. Multilocus genotyping should be used for all future *P. jiroveci* epidemiological studies involving respiratory samples as it is proving to be the most sensitive and powerful means of detecting differences between samples.

Multilocus genotyping has one disadvantage that it only detects the type of *P. jiroveci* DNA present. It does not give any information as to the viability of the organisms. Viability tests have been designed which use HSP70 RT-PCR assays (Latouche *et al.*, 2001, Maher *et al.*, 2001). Multilocus genotyping could be combined with the viability assays to generate a

more comprehensive picture of the genotypes of the organisms present and of the infectious nature of the organisms.

It is possible that different types of *P. jiroveci* have adapted to grow in the different microniches of the lung. The whole PCP infected lung contained two foci of infection. The two different genotypes detected could represent two different types of *P. jiroveci* organisms that have adapted to grow in different environmental microniches of the lung. This data suggests some caution when using respiratory samples, different types of *P. jiroveci* organisms may be detected in non-invasive and in invasive respiratory samples as they are derived from different lung environments. This would explain the difference seen between the HCW and PCP patient samples because the HCW samples were non-invasive induced sputum samples while the PCP patient samples were all invasive BAL samples. If this hypothesis were proved to be true fresh interpretation of results would be needed on all epidemiological studies using respiratory samples and future studies should use samples obtained in an equivalent manner.

Recent data generated from the amplification of *P. jiroveci* DNA detected in non-invasive respiratory samples has demonstrated the existence of limited asymptomatic carriage of *P. jiroveci* (Wakefield *et al.*, 2003). Asymptomatic carriage is now recognised as a relatively common feature of *P. jiroveci* and colonised individuals could act as a potential reservoir of organisms. This is further evidence in support of reinfection rather than latency as the model of infection leading to the development of disease, as there has to be circulation of *P. jiroveci* organisms for limited asymptomatic carriage of *P. jiroveci* organisms to occur. *P. jiroveci* is now thought to be transmitted from an immunocompromised PCP patient to either a susceptible immunocompromised host or to an immunocompetent close contact.

Alternatively, *P. jiroveci* could be transmitted from asymptomatic carriers to immunocompromised patients or an exogenous source of *P. jiroveci* might exist.

Potentially, the balance of the life cycle of *P. jiroveci* has altered since the advent of the AIDS pandemic. Pre 1981, *P. jiroveci* existed as a mild opportunistic infection, that mainly infected infants asymptomatically. This mild infection within infants appeared to be adequate for the replication and survival of the organism. Post 1981, there has been a huge increase in the number of immunocompromised patients. A direct effect of this is that *P. jiroveci* is now observed as a common potentially fatal opportunistic infection that proliferates within the lungs of these immunocompromised patients. One hypothesis for the changing balance of the *P. jiroveci* life cycle is that due to the increased number of potential immunocompromised hosts, the numbers of the organism have substantially increased. There is now a higher frequency of transmission between hosts of diverse susceptibility levels involving both carriage and infection of *P. jiroveci.* In fact, airborne *P. jiroveci* infection could possibly consist of a large spectrum of asymptomatic presentations and most likely, what is clinically recognised at present is only the tip of the iceberg (Dei-Cas, 2000).

8.2. Protease1

There were two main aims related to the investigation of Protease1. The first aim was to identify and characterise the original rat-derived *Pneumocystis* single copy *PRT1* gene that encodes the potential progenitor kexin-like protease and the second aim of the protease part of the project was to investigate the diversity and number of genomic and expressed copies of *PRT1* and to determine whether *PRT1* gene expression is regulated. If a progenitor kexin-like protease did exist, it is likely to be constitutively expressed and could complicate data from studies investigating the regulation of *PRT1* gene expression. Sequence data was

produced that may have originated from a progenitor kexin-like protease and *PRT1* gene expression was demonstrated to have some degree of regulation.

In order to investigate the progenitor kexin-like protease, non-telomeric copies of *PRT1* were successfully identified from Southern analysis. One non-telomeric copy, which bound to a probe specific to the *PRT1* catalytic domain but did not bind to a probe specific to the *PRT1* C-terminus domain, was characterised further because it was postulated that the C-terminus would have been altered to accommodate the change from a transmembrane intracellular protein to an extracellular protein bound to the cell surface by a GPI anchor. Sequence data was obtained from the catalytic domain and the 5' domain of the non-telomeric copy but the sequence data was not sufficiently divergent or unique to suggest that it was from the progenitor kexin-like protease. Therefore, not enough data had been obtained to conclude that the non-telomeric *PRT1* copy was the progenitor kexin-like protease. From the *PRT1* assay, there was no evidence of a constitutively expressed *PRT1* gene present in every population. Thus if the progenitor kexin-like protease does exist, it has yet to be shown to complicate studies on *PRT1* expression.

The *Pneumocystis* genome sequencing project may identify the progenitor kexin-like protease, if it does exist. Also recently two *PRT1* clones that encode a transmembrane C-terminus rather than encode a C-terminus with a GPI anchor site have been characterised (Russian *et al.,* 1999). The transmembrane domain is characteristic of an intracellular protein that binds to the Golgi. These clones could be characterised further. Single copy kexin-like genes have been detected in both mouse-derived *Pneumocystis* and *P. jiroveci*. The mouse derived *Pneumocystis* single copy kexin-like protease, KEX1, was identified by screening a cDNA library with two monoclonal antibodies that could recognise antigens other than mouse-derived *Pneumocystis* MSG and were shown to cross-react with *Pneumocystis* from

different hosts. These cross-reactive antibodies could be used to screen a *P. carinii* cDNA library to detect the *P. carinii* progenitor kexin-like protease. The *P. jiroveci* single copy kexin-like protease gene, *kex1*, was identified by a combination of techniques including a two-step PCR with magnetic beads, inverse PCR and genomic library screening. These techniques could be applied to the fragments of DNA obtained in this study to try to obtain their full length gene sequence and identify the progenitor kexin-like protease.

An alternative explanation is that the progenitor kexin-like protease may no longer exist in *P. carinii*. After the divergence of mouse-derived *Pneumocystis* and *P. carinii,* it is possible that within the genome of *P. carinii* a gene encoding a single copy kexin-like protease was translocated to the telomeres and then duplicated many times resulting in the *PRT1* multigene family. After the translocation event, a protein with a complementary function performed the function of the single copy kexin-like protease.

8.2.1. *PRT1* Gene Expression is Regulated

Two strategies were used to investigate whether *PRT1* gene expression is regulated. The first was based on the *MSG* gene PCR amplification assay of Sunkin (Sunkin and Stringer, 1997) and involved distinguishing between genomic and cDNA *PRT1* copies from the same population, by comparing their length polymorphisms within the proline rich region, PRR. The PRR assay was discarded because it lacked reliability and reproducibility. The second strategy involved using potentially clonal populations of rat-derived *Pneumocystis* generated from the low dose inoculation method to characterise genomic and expressed copies of *PRT1* genes by analysing the catalytic domain of *PRT1* alleles. The diversity, number and relatedness of the *PRT1* copies in these populations were investigated and the resulting data was analysed in order to determine whether there was regulation of *PRT1* expression.

UCS/MSG data indicated that simple populations of rat-derived *Pneumocystis* could be obtained with the low dose inoculation method. The *PRT1* assay identified two predominant *PRT1* sequence types, the genomic majority sequence type and the cDNA majority sequence type. Southern analysis using specific hybridisation probes showed these sequence types to be subgroups of *PRT1* alleles. The cDNA majority sequence type was the predominant expressed *PRT1I* gene in three populations, EA15, EA17 and EA19. In these three populations, expression of *PRT1* genes was not random therefore some means of regulation of *PRT1* gene expression appears to occur.

When a second group of rat-derived *Pneumocystis* simple populations were analysed the results showed that passaging of rat-derived *Pneumocystis* simple populations was possible. Airborne transmission of rat-derived *Pneumocystis* simple populations occurred and the cDNA majority sequence type was present in the second group of rat-derived *Pneumocystis* simple populations.

PRT1 genes are most likely to be regulated by transcription factors acting on promoters or silencing elements. 5' RACE successfully generated DNA sequence upstream of *PRT1* but there was little evidence to suggest that *PRT1* has a single expression site equivalent to the UCS.

To provide convincing evidence that *PRT1* gene expression is regulated there needs to be a better understanding of the level of clonality of the simple populations. There are three possible scenarios for the clonality of the simple populations; either all the organism types are identical, alternatively, there is a majority organism type but minor organism types exist too, or finally, all the organism types are different.

Assumptions about the clonality of the simple populations are based upon UCS/MSG data. The UCS/MSG junction is present in the *P. carinii* genome as a single copy locus; therefore, it can be used as a clonality indicator, assuming that two different *Pneumocystis* types do not have the same *MSG* gene at the UCS locus. From the UCS/MSG assay data, population EA21 appears to be clonal while the other four populations are simple. As well as the heterogeneity of the population a further factor that could influence the UCS/MSG data is the switching rate of the *MSG* genes to the UCS locus. Assessing *MSG* cDNA copies by RT-PCR amplification would provide more evidence about the clonality of the populations. This was attempted but was not successful (data not shown). It is not possible to assess the level of complexity of the populations from the *PRT1* genomic clones. If there was one sole organism type, there would be approximately 30 *PRT1* genomic alleles. These cannot be characterised from 12 genomic clones alone and it is arduous to do a complete characterisation. Before *PRT1* cDNA clones could be used as a clonality indicator, the assumption would have to be made about whether *PRT1* genes were regulated or not. It is possible that the simple populations are more heterogeneous than the UCS/MSG data suggests (section 6.2.1) but the interpretation of the expression of *PRT1* genes can only be done if the UCS/MSG data is used as a reliable indicator of clonality.

Thus, assuming from the UCS/MSG data that population EA21 is clonal and that the other four populations are simple, there are three possible scenarios for the regulation of *PRT1* expression within these populations. Firstly, expression of *PRT1* genes could be tightly regulated where *PRT1* gene expression is under the control of a single expression site. If there were one single organism type, as the *MSG* gene evidence suggests for population EA21, and *PRT1* expression was tightly regulated, one *PRT1* cDNA allele would be identified from the *PRT1* assay. This was not observed. Alternatively, if there were a majority organism type, which seems to be true for the other four populations, there would

be a majority *PRT1* cDNA allele type. This is true of EA15, EA17 and EA19 but not EA24. So tight regulation of *PRT1* genes is possible for EA15, EA17 and EA19.

In the second scenario, *PRT1* gene expression could be loosely regulated where *PRT1* gene expression is under the control of several expression sites. Again, if there was one sole organism type, as for EA21, and *PRT1* gene expression was loosely regulated, there would be few cDNA majority sequence types. Again this was not observed. For the other four populations, if there was a majority organism type and *PRT1* gene expression was loosely regulated, there would be a hint of a majority sequence type. This is true for EA24 where clusters of sequences were observed. The third scenario is that there may be no regulation of *PRT1* gene expression. There would be no cDNA majority sequence type regardless of population clonality. This could possibly be true for EA21, yet, cannot be true of EA15, EA17 and EA19. Furthermore, only a subset of *PRT1* copies was analysed from each population, therefore a comprehensive picture of the *PRT1* gene expression in each population cannot be achieved, only a representation is possible.

Based on the assumption that the data from the UCS/MSG assay gives an accurate indication of clonality, it is most likely from the *PRT1* assay, that in the simple populations EA15, EA17 and EA19, *PRT1* gene expression is tightly regulated while *PRT1* gene expression is loosely regulated in EA24 and not regulated in EA21. Many factors and assumptions have been made and lead to the conclusion that *PRT1* gene expression is almost certainly regulated but to what extent is not clear. Biologically, it is most likely that one means of regulation of *PRT1* gene expression occurs for all five populations yet this means of regulation is not apparent from the *PRT1* assay.

Thus, the above hypotheses are a simple interpretation of the data, based on several assumptions. Firstly, there is no latent infection of rat-derived *Pneumocystis,* and that the populations are free of contamination. This cannot be taken for granted with either set of simple populations. The clonality of the populations is based on the UCS/MSG assay data, which is thought to be correct. Yet, two different types of *Pneumocystis* may have the same *MSG* gene at the UCS locus. Switching of *MSG* and *PRT1* alleles could occur potentially complicating the data. As a single expression site has yet to be identified for the *PRT1* genes, it is likely there is more than one expression site and the organisms can switch between many different types of *PRT1* genes. This is true for *Trypanosoma brucei*, which has 20 similar telomeric-expression sites for variant surface glycoprotein genes (Chaves *et al.,* 1999). In the same manner that *MSR* genes are related to the *MSG* gene family there might be other forms of *PRT1* genes, which would add another level of diversity to the multigene families within *Pneumocystis.*

8.2.2. Co-expression of *PRT1* genes and *MSG* genes.

Expression of *PRT1* and *MSG* genes could be co-regulated because of their colinearity within the genome. The Sanger centre has sequenced *Pneumocystis* telomeric cosmid clones that consisted of unidirectional *MSG* and *PRT1* gene arrays. It is possible that a polycistronic message is generated, starting with the UCS followed by an *MSG* and *PRT1* gene. Alternatively, there could be a universal upstream promoter, governing the regulation of both genes via a collection of transcription factors.

Table 8.1 describes a comparison of the *MSG* and the *PRT1* predominantly expressed alleles.

Population	Majority *MSG* type	Majority *PRT1* type
EA15	*MSG*-A	*PRT1*-B
EA17	*MSG*-Y	*PRT1*-B
EA19	*MSG*-A	*PRT1*-B
EA21	*MSG*-A	*PRT1*-x
EA24	*MSG*-Y	*PRT1*-x

Table 8.1. Majority *MSG* and *PRT1* types within each of the simple populations

In table 8.1, each majority type has been assigned a letter. *PRT1*-B refers to the cDNA majority sequence type and *PRT1*-x implies there is no majority type. Both EA15 and EA19 express *MSG*-A and *PRT1*-B. From this it could be inferred that when the allele *MSG*-A is expressed so is *PRT1*-B. However, EA17 opposes this by having the genotype, *MSG*-Y and *PRT1*-B. Therefore, the two loci appear to be independent. Alternatively, regulation of their expression could be linked but switching has caused different alleles of *MSG* genes and *PRT1* genes to be present at the expression site within the individual populations. A majority *PRT1* type was not detected in populations EA21 and EA24 while a majority *MSG* gene type was. Assuming these two populations are simple, it appears that many *PRT1* genes are expressed while only one *MSG* gene is expressed, which negates the idea of a polycistronic message.

8.2.3. *PRT1* Genes may be Antigenically Variable

From the SNAP analysis, most of the *PRT1* sequences in the simple populations had a ds/dn ratio of less than one. This implies that PRT1 is under some form of selective pressure from the external environment, which in the rat lung would be the immune system of the rat. Antibodies specific to PRT1 have been detected in both rat and human sera (Lugli *et al.*, 1999b) showing that PRT1 is an important immunodominant *Pneumocystis* protein. Also, even though the rats are immunosuppressed, rarely is the immune system fully depleted. The simple populations were generated in nude rats. Nude rats are deficient in T cell activity;

however, they do have some B cell activity, which would not be completely depleted by the dexamethasone treatment. The immune system could select against various protein copies of PRT1. To evade the immune system the rat-derived *Pneumocystis* population could express a different copy of a *PRT1* gene. As an antibody against one predominant subgroup of PRT1 proteins is developed, a *Pneumocystis* organism expresses a different copy of a *PRT1* gene. This organism then divides into a population and thrives until another antibody is developed against the new isoform of PRT1. PRT1 would be undergoing antigenic variation. The assumption is being made that one or a few copies of *PRT1* genes are being expressed per organism and that there is a specific mechanism of *PRT1* gene expression.

SNAP analysis provides another piece of data supporting the idea that *PRT1* genes are the basis of antigenic variation. Some regions of the catalytic domain of *PRT1* genes have a predominance of non-synonymous changes (Figure 6.11). These regions could be exposed and also be highly variable epitopes to help evade the rat immune system. As previously described, most of the simple populations subdivided into two groups, one that appeared to be under selective pressure ($ds/dn < 1$) while the other was under selective constraint ($ds/dn > 1$). A hypothesis for this data is that the group under selective constraint represents the initial inoculum, which has proliferated into a *Pneumocystis* infection. However, the host then exerts selective pressure in the form of an immune response so subgroups that are expressing an antigenically different allele of PRT1 thrive. These subgroup alleles are under selective pressure so have lower ds/dn ratios. Yet, this subgroup phenomenon is seen in all populations and with both genomic and cDNA clones, implying it is most likely the result of the actual *PRT1* sequence, rather than a biological occurrence. There is no evidence of rapid switching of *PRT1* genes, which would be necessary for antigenic variation. It is more likely that the variation characterised within the *PRT1* alleles is the result of antigenic drift and

reflects the selective pressure of the host immune system on the variability of the *PRT1* multigene family over many generations of *Pneumocystis* populations.

The simple populations represent the best clonal stocks of *Pneumocystis* that has been generated so far and the low-dose methodology has the potential to produce prolific reproducible clonal stocks of rat-derived *Pneumocystis* organisms. The results of contamination need to be better monitored. Once this occurs, a continuous source of clonal organisms could be generated for use in genetic manipulation experiments, such as the switching of multicopy telomeric genes.

To obtain a clearer picture of the genomic context of *PRT1* genes and of *PRT1* gene expression, DNA hybridisation probes were designed that were specific to the cDNA majority sequence type and the genomic majority sequence type and were demonstrated to bind to a subset of *PRT1* alleles from genomic rat-derived *Pneumocystis* DNA. A high level of specificity was reproducibly achieved with these probes. The method could be fully exploited so that specific probes could be designed that bound to every expressed *PRT1* and *MSG* allele. A series of hybridisation experiments comparing different clonal populations could be used to determine if expressed *PRT1* and *MSG* alleles were always present on the same chromosomes or if they switched to other chromosomes. Ultimately, each *PRT1* allele could be tracked across clonal populations of rat-derived *Pneumocystis*.

The *Pneumocystis* genome sequencing project has the potential to clarify many issues about the organism and rapidly advance research into *Pneumocystis*. Firstly, the progenitor kexin-like protease might be identified along with the *PRT1* equivalent of the UCS locus. The gene arrays that contain copies of the *Pneumocystis* multigene families would be fully characterised. The number, frequency and chromosomal location of each *MSG* and *PRT1*

copy would then be known. The genes encoding the proteins in the sexual stages of the *Pneumocystis* life cycle would be identified and characterised and far more would be known about this stage in the *Pneumocystis* life cycle. The same is true about the genes involved in the metabolism of *Pneumocystis*.

Ultimately, with the combination of clonal populations and the *Pneumocystis* genome sequencing project, the field of *Pneumocystis* research is about to break the limitations imposed by the lack of an effective culture system.

References

Aliouat, E.M., Mazars, E., Dei-Cas, E., Cesbron, J.Y. and Camus, D., Intranasal inoculation of mouse, rat or rabbit-derived *Pneumocystis*in SCID mice. *J. Protozool. Res.* 1993. **3**: 94-98.

Aliouat, E.M., Mazars, E., Dei-Cas, E., Delcourt, P., Billaut, P. and Camus, D., Pneumocystis cross infection experiments using SCID mice and nude rats as recipient host, showed strong host-species specificity. *J Eukaryot Microbiol.* 1994. **41**(5): 71S.

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

Arend, S.M., Kroon, F.P. and van't Wout, J.W., *Pneumocystis carinii* pneumonia in patients without AIDS, 1980 through 1993. An analysis of 78 cases. *Arch Intern Med.* 1995. **155**(22): 2436-41.

Atzori, C., Angeli, E., Mainini, A., Agostoni, F., Micheli, V. and Cargnel, A., In vitro activity of human immunodeficiency virus protease inhibitors against *Pneumocystis carinii*. *J Infect Dis.* 2000. **181**(5): 1629-34.

Atzori, C., Agostoni, F., Angeli, E., Mainini, A., Micheli, V. and Cargnel, A., P.carinii host specificity: attempt of cross infections with human derived strains in rats. *J Eukaryot Microbiol.* 1999. **46**(5): 112S.

Banerji, S., Wakefield, A.E., Allen, A.G., Maskell, D.J., Peters, S.E. and Hopkin, J.M., The cloning and characterization of the arom gene of *Pneumocystis carinii*. *J Gen Microbiol.* 1993. **139**(Pt 12): 2901-14.

Bartlett, M.S., Vermund, S.H., Jacobs, R., Durant, P.J., Shaw, M.M., Smith, J.W., Tang, X., Lu, J.J., Li, B., Jin, S. and Lee, C.H., Detection of *Pneumocystis carinii* DNA in air samples: likely environmental risk to susceptible persons. *J Clin Microbiol.* 1997. **35**(10): 2511-3.

Bartlett, M.S., Lee, C.H., Lu, J.J., Bauer, N.L., Bettz, J.F., McLaughlin, G.L. and Smith, J.W., *Pneumocystis carinii* detected in air. *J Eukaryot Microbiol.* 1994. **41**(5): 75S.

Baughman, R.P., Dohn, M.N., Shipley, R., Buchsbaum, J.A. and Frame, P.T., Increased *Pneumocystis carinii* recovery from the upper lobes in *Pneumocystis* pneumonia. The effect of aerosol pentamidine prophylaxis. *Chest.* 1993. **103**(2): 426-32.

Bazaz, G.R., Manfredi, O.L., Howard, R.G. and Claps, A.A., *Pneumocystis carinii* pneumonia in three full-term siblings. *J. Pediatr.* 1970. **76**(5): 767-9.

Beach, R.S., Garcia, E.R., Sosa, R. and Good, R.A., *Pneumocystis carinii* pneumonia in a human immunodeficiency virus 1- infected neonate with meconium aspiration. *Pediatr Infect Dis J.* 1991. **10**(12): 953-5.

Beard, C.B., Carter, J.L., Keely, S.P., Huang, L., Pieniazek, N.J., Moura, I.N., Roberts, J.M., Hightower, A.W., Bens, M.S., Freeman, A.R., Lee, S., Stringer, J.R., Duchin, J.S., del Rio,

C., Rimland, D., Genetic variation in *Pneumocystis carinii* isolates from different geographic regions: implications for transmission. *Emerg Infect Dis*. 2000. **6**(3): 265-72.

Benfield, T.L. and Lundgren, J.D., The *Pneumocystis carinii* major surface glycoprotein (MSG): its potential involvement in the pathophysiology of pneumocystosis. *FEMS Immunol Med Microbiol*. 1998. **22**(1-2): 129-34.

Boylan, C.J. and Current, W.L., Improved rat model of *Pneumocystis carinii* pneumonia: induced laboratory infections in *Pneumocystis*-free animals. *Infect Immun*. 1992. **60**(4): 1589-97.

Brenner, C., Bevan, A. and Fuller, R.S., Biochemical and genetic methods for analyzing specificity and activity of a precursor-processing enzyme: yeast Kex2 protease, kexin. *Methods Enzymol*. 1994. **244**: 152-67.

Brooks, D.R., Wang, P., Read, M., Watkins, W.M., Sims, P.F. and Hyde, J.E., Sequence variation of the hydroxymethyldihydropterin pyrophosphokinase: dihydropteroate synthase gene in lines of the human malaria parasite, *Plasmodium falciparum*, with differing resistance to sulfadoxine. *Eur J Biochem*. 1994. **224**(2): 397-405.

Calderon, E.J., Regordan, C., Medrano, F.J., Ollero, M. and Varela, J.M., *Pneumocystis carinii* infection in patients with chronic bronchial disease. *Lancet*. 1996. **347**(9006): 977.

Cere, N., Drouet-Viard, F., Dei-Cas, E., Chanteloup, N. and Coudert, P., In utero transmission of *Pneumocystis carinii sp. f. oryctolagi*. *Parasite*. 1997. **4**(4): 325-30.

Chaves, I., Rudenko, G., Dirks-Mulder, A., Cross, M. and Borst, P., Control of variant surface glycoprotein gene-expression sites in *Trypanosoma brucei*. *Embo J*. 1999. **18**(17): 4846-55.

Chin, K., Luttrell, T.D., Roe, J.D., Shadzi, S., Wyder, M.A. and Kaneshiro, E.S., Putative Pneumocystis dormant forms outside the mammalian host, and long-term culture derived from them: initial characterizations. *J Eukaryot Microbiol*. 1999. **46**(5): 95S-99S.

Connor, E., Bagarazzi, M., McSherry, G., Holland, B., Boland, M., Denny, T. and Oleske, J., Clinical and laboratory correlates of *Pneumocystis carinii* pneumonia in children infected with HIV. *JAMA*. 1991. **265**(13): 1693-7.

Cushion, M.T., Linke, M.J., Collins, M., Keely, S.P. and Stringer, J.R., The minimum number of *Pneumocystis carinii f. sp. carinii* organisms required to establish infections is very low. *J Eukaryot Microbiol*. 1999. **46**(5): 111S.

Cushion, M.T., Kaselis, M., Stringer, S.L. and Stringer, J.R., Genetic stability and diversity of *Pneumocystis carinii* infecting rat colonies. *Infect Immun*. 1993a. **61**(11): 4801-13.

Cushion, M.T., Zhang, J., Kaselis, M., Giuntoli, D., Stringer, S.L. and Stringer, J.R., Evidence for two genetic variants of *Pneumocystis carinii* coinfecting laboratory rats. *J Clin Microbiol*. 1993b. **31**(5): 1217-23.

Cushion, M.T., Ruffolo, J.J. and Walzer, P.D., Analysis of the developmental stages of *Pneumocystis carinii*, in vitro. *Lab Invest*. 1988. **58**(3): 324-31.

Day, J.H., Williams, I.G., Zumla, A. and Miller, R.F., Protection from *Pneumocystis carinii* pneumonia. *Sex Transm Infect*. 1999. **75**(3): 202-3.

Dei-Cas, E., *Pneumocystis* infections: the iceberg? *Med Mycol*. 2000. **38**(Suppl 1): 23-32.

Demanche, C., Berthelemy, M., Petit, T., Polack, B., Wakefield, A.E., Dei-Cas, E. and Guillot, J., Phylogeny of *Pneumocystis carinii* from 18 primate species confirms host specificity and suggests coevolution. *J Clin Microbiol*. 2001. **39**(6): 2126-33.

Denis, C.M., Mazars, E., Guyot, K., Odberg-Ferragut, C., Viscogliosi, E., Dei-Cas, E. and Wakefield, A.E., Genetic divergence at the SODA locus of six different *formae speciales* of *Pneumocystis carinii*. *Med Mycol*. 2000. **38**(4): 289-300.

Dohn, M.N., White, M.L., Vigdorth, E.M., Ralph Buncher, C., Hertzberg, V.S., Baughman, R.P., George Smulian, A. and Walzer, P.D., Geographic clustering of *Pneumocystis carinii* pneumonia in patients with HIV infection. *Am J Respir Crit Care Med*. 2000. **162**(5): 1617-21.

Dumoulin, A., Mazars, E., Seguy, N., Gargallo-Viola, D., Vargas, S., Cailliez, J.C., Aliouat, E.M., Wakefield, A.E. and Dei-Cas, E., Transmission of *Pneumocystis carinii* disease from immunocompetent contacts of infected hosts to susceptible hosts. *Eur J Clin Microbiol Infect Dis*. 2000. **19**(9): 671-8.

Durand-Joly, I., Aliouat el, M., Recourt, C., Guyot, K., Francois, N., Wauquier, M., Camus, D. and Dei-Cas, E., *Pneumocystis carinii f. sp. hominis* is not infectious for SCID mice. *J Clin Microbiol*. 2002. **40**(5): 1862-5.

Edman, J.C., Kovacs, J.A., Masur, H., Santi, D.V., Elwood, H.J. and Sogin, M.L., Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. *Nature*. 1988. **334**(6182): 519-22.

Edman, U., Edman, J.C., Lundgren, B. and Santi, D.V., Isolation and expression of the *Pneumocystis carinii* thymidylate synthase gene. *Proc Natl Acad Sci U S A*. 1989. **86**(17): 6503-7.

English, K., Peters, S.E., Maskell, D.J. and Collins, M.E., DNA analysis of *Pneumocystis* infecting a Cavalier King Charles spaniel. *J Eukaryot Microbiol*. 2001. .

Feavers, I.M., Gray, S.J., Urwin, R., Russell, J.E., Bygraves, J.A., Kaczmarski, E.B. and Maiden, M.C., Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. *J Clin Microbiol*. 1999. **37**(12): 3883-7.

Fermer, C., Kristiansen, B.E., Skold, O. and Swedberg, G., Sulfonamide resistance in *Neisseria meningitidis* as defined by site- directed mutagenesis could have its origin in other species. *J Bacteriol*. 1995. **177**(16): 4669-75.

Furuta, T., Fujita, M., Mukai, R., Sakakibara, I., Sata, T., Miki, K., Hayami, M., Kojima, S. and Yoshikawa, Y., Severe pulmonary pneumocystosis in simian acquired immunodeficiency syndrome induced by simian immunodeficiency virus: its characterization by the polymerase-chain-reaction method and failure of experimental transmission to immunodeficient animals. *Parasitol Res*. 1993. **79**(8): 624-8.

Frenkel, J.K., *Pneumocystis* pneumonia, an immunodeficiency-dependent disease (IDD): a critical historical overview. *J Eukaryot Microbiol.* 1999. **46**(5): 89S-92S.

Frenkel, J.K., Good, J.T. and Shultz, J.A., Latent *Pneumocystis* infection of rats, relapse, and chemotherapy. *Lab Invest.* 1966. **15**(10): 1559-77.

Fuller, R.S., Brake, A. and Thorner, J., Yeast prohormone processing enzyme (KEX2 gene product) is a Ca2+- dependent serine protease. *Proc Natl Acad Sci U S A.* 1989. **86**(5): 1434-8.

Garbe, T.R. and Stringer, J.R., Molecular characterization of clustered variants of genes encoding major surface antigens of human *Pneumocystis carinii. Infect Immun.* 1994. **62**(8): 3092-101.

Gibb, D.M., Davison, C.F., Holland, F.J., Walters, S., Novelli, V. and Mok, J., *Pneumocystis carinii* pneumonia in vertically acquired HIV infection in the British Isles. *Arch Dis Child.* 1994. **70**(3): 241-4.

Gigliotti, F. and Haidaris, C.G., Antigenic characterization of *Pneumocystis carinii. Semin Respir Infect.* 1998. **13**(4): 313-22.

Gigliotti, F. and McCool, T., Glycoprotein A is the immunodominant antigen of *Pneumocystis carinii* in mice following immunization. *Parasitol Res.* 1996. **82**(1): 90-1.

Gigliotti, F., Harmsen, A.G., Haidaris, C.G. and Haidaris, P.J., *Pneumocystis carinii* is not universally transmissible between mammalian species. *Infect Immun.* 1993. **61**(7): 2886-90.

Gigliotti, F., Host species-specific antigenic variation of a mannosylated surface glycoprotein of *Pneumocystis carinii. J Infect Dis.* 1992. **165**(2): 329-36.

Giuntoli, D., Stringer, S.L. and Stringer, J.R., Extraordinarily low number of ribosomal RNA genes in *P. carinii. J Eukaryot Microbiol.* 1994. **41**(5): 88S.

Grebski, E., Russi, E.W., Speich, R., Opravil, M., Kuster, H. and Wust, J., The role of two-segment bronchoalveolar lavage in the diagnosis of pulmonary infections. *Chest.* 1994. **106**(2): 414-20.

Guadiz, G., Haidaris, C.G., Maine, G.N. and Simpson-Haidaris, P.J., The carboxyl terminus of *Pneumocystis carinii* glycoprotein A encodes a functional glycosylphosphatidylinositol signal sequence. *J Biol Chem.* 1998. **273**(40): 26202-9.

Guarro, J., GeneJ and Stchigel, A.M., Developments in fungal taxonomy. *Clin Microbiol Rev.* 1999. **12**(3): 454-500.

Guillot, J., Demanche, C., Hugot, J.P., Berthelemy, M., Wakefield, A.E., Dei-Cas, E. and Chermette, R., Parallel phylogenies of *Pneumocystis* species and their mammalian hosts. *J Eukaryot Microbiol.* 2001. : 113S-115S.

Haidaris, P.J., Wright, T.W., Gigliotti, F. and Haidaris, C.G., Expression and characterization of a cDNA clone encoding an immunodominant surface glycoprotein of *Pneumocystis carinii. J Infect Dis.* 1992. **166**(5): 1113-23.

Hauser, P.M., Blanc, D.S., Sudre, P., Senggen Manoloff, E., Nahimana, A., Bille, J., Weber, R. and Francioli, P., Genetic diversity of *Pneumocystis carinii* in HIV-positive and -negative patients as revealed by PCR-SSCP typing. *AIDS*. 2001. **15**(4): 461-6.

Hauser, P.M., Blanc, D.S., Bille, J., Nahimana, A. and Francioli, P., Carriage of *Pneumocystis carinii* by immunosuppressed patients and molecular typing of the organisms. *AIDS*. 2000. **14**(4): 461-3.

Helweg-Larsen, J., Skov Jensen, J., Dohn, B., Benfield, T.L. and Lundgren, B., Detection of *Pneumocystis* DNA in samples from patients suspected of bacterial pneumonia - a case-control study. *BMC Infect Dis*. 2002. **2**(1): 28.

Helweg-Larsen, J., Lundgren, B. and Lundgren, J.D., Heterogeneity and compartmentalization of *Pneumocystis carinii f. sp. hominis* genotypes in autopsy lungs. *J Clin Microbiol*. 2001a. **39**(10): 3789-92.

Helweg-Larsen, J., Lee, C.H., Jin, S., Hsueh, J.Y., Benfield, T.L., Hansen, J., Lundgren, J.D. and Lundgren, B., Clinical correlation of variations in the internal transcribed spacer regions of rRNA genes in *Pneumocystis carinii f.sp. hominis*. *AIDS*. 2001b. **15**(4): 451-9.

Helweg-Larsen, J., Benfield, T.L., Eugen-Olsen, J., Lundgren, J.D. and Lundgren, B., Effects of mutations in *Pneumocystis carinii* dihydropteroate synthase gene on outcome of AIDS-associated *P. carinii* pneumonia. *Lancet*. 1999. **354**(9187): 1347-51.

Helweg-Larsen, J., Tsolaki, A.G., Miller, R.F., Lundgren, B. and Wakefield, A.E., Clusters of *Pneumocystis carinii* pneumonia: analysis of person-to-person transmission by genotyping. *QJM*. 1998. **91**(12): 813-20.

Hong, S.T., Lee, M., Seo, M., Choo, D.H., Moon, H.R. and Lee, S.H., Immunoblot analysis for serum antibodies to *Pneumocystis carinii* by age and intensity of infection in rats. *Korean J Parasitol*. 1995. **33**(3): 187-94.

Hosoya, N., Takahashi, T., Wada, M., Endo, T., Nakamura, T., Sakashita, H., Kimura, K., Ohnishi, K., Nakamura, Y., Mizuochi, T. and Iwamoto, A., Genotyping of *Pneumocystis carinii f. sp. hominis* isolates in Japan based on nucleotide sequence variations in internal transcribed spacer regions of rRNA genes. *Microbiol Immunol*. 2000. **44**(7): 591-6.

Huang, L., Beard, C.B., Creasman, J., Levy, D., Duchin, J.S., Lee, S., Pieniazek, N., Carter, J.L., del Rio, C., Rimland, D. and Navin, T.R., Sulfa or sulfone prophylaxis and geographic region predict mutations in the *Pneumocystis carinii* dihydropteroate synthase gene. *J Infect Dis*. 2000. **182**(4): 1192-8.

Huang, S.N., Fischer, S.H., E, O.S., Gill, V.J., Masur, H. and Kovacs, J.A., Development of a PCR assay for diagnosis of *Pneumocystis carinii* pneumonia based on amplification of the multicopy major surface glycoprotein gene family. *Diagn Microbiol Infect Dis*. 1999a. **35**(1): 27-32.

Huang, S.N., Angus, C.W., Turner, R.E., Sorial, V. and Kovacs, J.A., Identification and characterization of novel variant major surface glycoprotein gene families in rat *Pneumocystis carinii*. *J Infect Dis*. 1999b. **179**(1): 192-200.

Hughes, W.T., Current issues in the epidemiology, transmission, and reactivation of *Pneumocystis carinii*. *Semin Respir Infect*. 1998a. **13**(4): 283-8.

Hughes, W.T., Prologue to AIDS. The recognition of infectious opportunists. *Medicine (Baltimore)*. 1998b. **77**(4): 227-32.

Hughes, W.T., *Pneumocystis* in infants and children. *N Engl J Med*. 1995a. **333**(5): 320-1.

Hughes, W.T., LaFon, S.W., Scott, J.D. and Masur, H., Adverse events associated with trimethoprim-sulfamethoxazole and atovaquone during the treatment of AIDS-related *Pneumocystis carinii* pneumonia. *J Infect Dis*. 1995b. **171**(5): 1295-301.

Hughes, W.T., Bartley, D.L. and Smith, B.M., A natural source of infection due to *Pneumocystis carinii*. *J Infect Dis*. 1983. **147**(3): 595.

Hughes, W.T., Natural mode of acquisition for de novo infection with *Pneumocystis carinii*. *J Infect Dis*. 1982. **145**(6): 842-8.

Hunter, J.A. and Wakefield, A.E., Genetic divergence at the mitochondrial small subunit ribosomal RNA gene among isolates of *Pneumocystis carinii* from five mammalian host species. *J Eukaryot Microbiol*. 1996. **43**(5): 24S-25S.

Icenhour, C.R., Rebholz, S.L., Collins, M.S. and Cushion, M.T., Early Acquisition of *Pneumocystis carinii* in Neonatal Rats using Targeted PCR and Oral Swabs. *J. Eukaryot. Microbiol*. 2002. **49**(2): 135S-136S.

Jensen, T.K., Boye, M. and Bille-Hansen, V., Application of fluorescent in situ hybridization for specific diagnosis of *Pneumocystis carinii* pneumonia in foals and pigs. *Vet Pathol*. 2001. **38**(3): 269-74.

Jules-Elysee, K.M., Stover, D.E., Zaman, M.B., Bernard, E.M. and White, D.A., Aerosolized pentamidine: effect on diagnosis and presentation of *Pneumocystis carinii* pneumonia. *Ann Intern Med*. 1990. **112**(10): 750-7.

Kaiser, K., Rabodonirina, M. and Picot, S., Real time quantitative PCR and RT--PCR for analysis of *Pneumocystis carinii hominis*. *J Microbiol Methods*. 2001 Jun. **45**(2): 113-8.

Kaneshiro, E.S., Are cytochrome b gene mutations the only cause of atovaquone resistance in *Pneumocystis*? *Drug Resist Updat*. 2001. **4**(5): 322-9.

Karos, M., Chang, Y.C., McClelland, C.M., Clarke, D.L., Fu, J., Wickes, B.L. and Kwon-Chung, K.J., Mapping of the Cryptococcus neoformans MATalpha locus: presence of mating type-specific mitogen-activated protein kinase cascade homologs. *J Bacteriol*. 2000. **182**(21): 6222-7.

Katz, M.H., Hessol, N.A., Buchbinder, S.P., Hirozawa, A., O'Malley, P. and Holmberg, S.D., Temporal trends of opportunistic infections and malignancies in homosexual men with AIDS. *J Infect Dis*. 1994. **170**(1): 198-202.

Kazanjian, P., Armstrong, W., Hossler, P.A., Burman, W., Richardson, J., Lee, C.H., Crane, L., Katz, J. and Meshnick, S.R., *Pneumocystis carinii* mutations are associated with duration of sulfa or sulfone prophylaxis exposure in AIDS patients. *J Infect Dis*. 2000. **182**(2): 551-7.

Kazanjian, P., Locke, A.B., Hossler, P.A., Lane, B.R., Bartlett, M.S., Smith, J.W., Cannon, M. and Meshnick, S.R., *Pneumocystis carinii* mutations associated with sulfa and sulfone prophylaxis failures in AIDS patients. *AIDS*. 1998. **12**(8): 873-8.

Keely, S.P., Cushion, M.T. and Stringer, J.R., Diversity at the Locus Associated with Transcription of a Variable Surface Antigen of *Pneumocystis carinii* as an Index of Population Structure and Dynamics in Infected Rats. *Infect Immun*. 2003. **71**(1): 47-60.

Keely, S.P., Wakefield, A.E., Cushion, M.T., Smulian, A.G., Hall, N., Barrell, B.G. and Stringer, J.R., Detailed structure of *Pneumocystis carinii* Chromosome Ends. *J Eukaryot Microbiol*. 2001.128S-130S.

Keely, S.P., Cushion, M.T. and Stringer, J.R., Determination of the maximum frequency of genetic rearrangements associated with *Pneumocystis carinii* surface antigen variation. *J Eukaryot Microbiol*. 1999. **46**(5): 128S.

Keely, S.P., Baughman, R.P., Smulian, A.G., Dohn, M.N. and Stringer, J.R., Source of *Pneumocystis carinii* in recurrent episodes of pneumonia in AIDS patients. *AIDS*. 1996. **10**(8): 881-8.

Keely, S.P., Stringer, J.R., Baughman, R.P., Linke, M.J., Walzer, P.D. and Smulian, A.G., Genetic variation among *Pneumocystis carinii hominis* isolates in recurrent pneumocystosis. *J Infect Dis*. 1995. **172**(2): 595-8.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M., MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*. 2001. **17**(12): 1244-5.

Kutty, G. and Kovacs, J.A., A Single-Copy Gene Encodes Kex1, a Serine Endoprotease of *Pneumocystis jiroveci*. *Infect Immun*. 2003. **71**(1): 571-4.

Kutty, G., Ma, L. and Kovacs, J.A., Characterization of the expression site of the major surface glycoprotein of human-derived *Pneumocystis carinii*. *Mol Microbiol*. 2001. **42**(1): 183-93.

Kyes, S., Horrocks, P. and Newbold, C., Antigenic variation at the infected red cell surface in malaria. *Annu Rev Microbiol*. 2001. **55**: 673-707.

Kyes, S., Pinches, R. and Newbold, C., A simple RNA analysis method shows var and rif multigene family expression patterns in *Plasmodium falciparum*. *Mol Biochem Parasitol*. 2000. **105**(2): 311-5.

Laakkonen, J., Fisher, R.N. and Case, T.J., Pneumocystosis in wild small mammals from California. *J Wildl Dis*. 2001. **37**(2): 408-12.

Lane, B.R., Ast, J.C., Hossler, P.A., Mindell, D.P., Bartlett, M.S., Smith, J.W. and Meshnick, S.R., Dihydropteroate synthase polymorphisms in *Pneumocystis carinii*. *J Infect Dis*. 1997. **175**(2): 482-5.

Larsen, H.H., Kovacs, J.A., Stock, F., Vestereng, V.H., Lundgren, B., Fischer, S.H. and Gill, V.J., Development of a rapid real-time PCR assay for quantitation of *Pneumocystis carinii f. sp. carinii*. *J Clin Microbiol*. 2002. **40**(8): 2989-93.

Latouche, S., Totet, A., Lacube, P., Bolognini, J., Nevez, G. and Roux, P., Development of an RT-PCR on the Heat Shock Protein 70 Gene for Viability Detection of *Pneumocystis carinii f. sp. hominis* in Patients with Pneumocystosis and in Air Sample. *J Eukaryot Microbiol*. 2001. **48**(2): 176S-177S.

Latouche, S., Olsson, M., Polack, B., Brun-Pascaud, M., Bernard, C. and Roux, P., Detection of *Pneumocystis carinii* f. sp. in air samples collected in animal rooms. *J Eukaryot Microbiol*. 1997a **44**(6): 46S-47S.

Latouche, S., Poirot, J.L., Bernard, C. and Roux, P., Study of internal transcribed spacer and mitochondrial large-subunit genes of *Pneumocystis carinii hominis* isolated by repeated bronchoalveolar lavage from human immunodeficiency virus-infected patients during one or several episodes of pneumonia. *J Clin Microbiol*. 1997b **35**(7): 1687-90.

Latouche, S., Ortona, E., Mazars, E., Margutti, P., Tamburrini, E., Siracusano, A., Guyot, K., Nigou, M. and Roux, P., Biodiversity of *Pneumocystis carinii hominis*: typing with different DNA regions. *J Clin Microbiol*. 1997c **35**(2): 383-7.

Lee, C.H., Helweg-Larsen, J., Tang, X., Jin, S., Li, B., Bartlett, M.S., Lu, J.J., Lundgren, B., Lundgren, J.D., Olsson, M., Lucas, S.B., Roux, P., Cargnel, A., Atzori, C., Matos, O., Update on *Pneumocystis carinii f. sp. hominis* typing based on nucleotide sequence variations in internal transcribed spacer regions of rRNA genes. *J Clin Microbiol*. 1998. **36**(3): 734-41.

Lee, L.H., Gigliotti, F., Wright, T.W., Simpson-Haidaris, P.J., Weinberg, G.A. and Haidaris, C.G., Molecular characterization of KEX1, a kexin-like protease in mouse *Pneumocystis carinii*. *Gene*. 2000. **242**(1-2): 141-50.

Leigh, T.R., Millett, M.J., Jameson, B. and Collins, J.V., Serum titres of *Pneumocystis carinii* antibody in health care workers caring for patients with AIDS. *Thorax*. 1993. **48**(6): 619-21.

Limper, A.H., Thomas, C.F., Jr., Anders, R.A. and Leof, E.B., Interactions of parasite and host epithelial cell cycle regulation during *Pneumocystis carinii* pneumonia. *J Lab Clin Med*. 1997. **130**(2): 132-8.

Lu, J.J., Chen, C.H., Bartlett, M.S., Smith, J.W. and Lee, C.H., Comparison of six different PCR methods for detection of *Pneumocystis carinii*. *J Clin Microbiol*. 1995. **33**(10): 2785-8.

Lu, J.J., Bartlett, M.S., Shaw, M.M., Queener, S.F., Smith, J.W., Ortiz-Rivera, M., Leibowitz, M.J. and Lee, C.H., Typing of *Pneumocystis carinii* strains that infect humans based on nucleotide sequence variations of internal transcribed spacers of rRNA genes. *J Clin Microbiol*. 1994. **32**(12): 2904-12.

Lugli, E.B. analysis of a Multi-Gene Family, *PRT1* encoding subtilisin-like serine proteases in *Pneumocystis carinii*. 1999a. Oxford.

Lugli, E.B., Bampton, E.T., Ferguson, D.J. and Wakefield, A.E., Cell surface protease PRT1 identified in the fungal pathogen *Pneumocystis carinii*. *Mol Microbiol*. 1999b. **31**(6): 1723-33.

Lugli, E.B., Allen, A.G. and Wakefield, A.E., A *Pneumocystis carinii* multi-gene family with homology to subtilisin- like serine proteases. *Microbiology*. 1997. **143**(Pt 7): 2223-36.

Lundgren, B., Elvin, K., Rothman, L.P., Ljungstrom, I., Lidman, C. and Lundgren, J.D., Transmission of *Pneumocystis carinii* from patients to hospital staff. *Thorax*. 1997. **52**(5): 422-4.

Ma, L., Kutty, G., Jia, Q., Imamichi, H., Huang, L., Atzori, C., Beckers, P., Groner, G., Beard, C.B. and Kovacs, J.A., Analysis of Variation in Tandem Repeats in the Intron of the Major Surface Glycoprotein Expression Site of the Human Form of *Pneumocystis carinii*. *J Infect Dis*. 2002. **186**(11): 1647-54.

Ma, L. and Kovacs, J.A., Genetic analysis of multiple loci suggests that mutations in the *Pneumocystis carinii f. sp. hominis* dihydropteroate synthase gene arose independently in multiple strains. *Antimicrob Agents Chemother*. 2001. **45**(11): 3213-5.

Ma, L., Borio, L., Masur, H. and Kovacs, J.A., *Pneumocystis carinii* dihydropteroate synthase but not dihydrofolate reductase gene mutations correlate with prior trimethoprim-sulfamethoxazole or dapsone use. *J Infect Dis*. 1999. **180**(6): 1969-78.

Maher, N.H., Vermund, S.H., Welsh, D.A., Dillon, H.K., Awooda, A. and Unnasch, T.R., Development and characterization of a molecular viability assay for *Pneumocystis carinii f sp hominis*. *J Infect Dis*. 2001. **183**(12): 1825-7.

Maher, N.H., Vermund, S., Lasbury, M., Lee, C., Bartlett, M. and Unnasch, T.R., Development and evaluation of a molecular viability assay for *Pneumocystis carinii*. *J Clin Microbiol*. 2000. **38**(5): 1947-52.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. and Spratt, B.G., Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A*. 1998. **95**(6): 3140-5.

Mazars, E., Odberg-Ferragut, C., Dei-Cas, E., Fourmaux, M.N., Aliouat, E.M., Brun-Pascaud, M., Mougeot, G. and Camus, D., Polymorphism of the thymidylate synthase gene of *Pneumocystis carinii* from different host species. *J Eukaryot Microbiol*. 1995. **42**(1): 26-32.

Mei, Q., Turner, R.E., Sorial, V., Klivington, D., Angus, C.W. and Kovacs, J.A., Characterization of major surface glycoprotein genes of human *Pneumocystis carinii* and high-level expression of a conserved region. *Infect Immun*. 1998. **66**(9): 4268-73.

Merali, S., Frevert, U., Williams, J.H., Chin, K., Bryan, R. and Clarkson, A.B., Jr., Continuous axenic cultivation of *Pneumocystis carinii*. *Proc Natl Acad Sci U S A*. 1999. **96**(5): 2402-7.

Meyerhans, A., Vartanian, J.P. and Wain-Hobson, S., DNA recombination during PCR. *Nucleic Acids Res.* 1990. **18**(7): 1687-91.

Miller, R.F. and Wakefield, A.E., *Pneumocystis carinii* genotypes and severity of pneumonia. *Lancet.* 1999a. **353**(9169): 2039-40.

Miller, R.F., *Pneumocystis carinii* infection in non-AIDS patients. *Current Opinion in Infectious Diseases.* 1999b. **12**: 371-377.

Miller, R.F., Kocjan, G., Buckland, J., Holton, J., Malin, A. and Semple, S.J., Sputum induction for the diagnosis of pulmonary disease in HIV positive patients. *J Infect.* 1991. **23**(1): 5-15.

Morris, A., Wachter, R.M., Luce, J., Turner, J. and Huang, L., Improved survival with highly active antiretroviral therapy in HIV- infected patients with severe *Pneumocystis carinii* pneumonia. *AIDS.* 2003. **17**(1): 73-80.

Morris, A., Beard, C.B. and Huang, L., Update on the epidemiology and transmission of *Pneumocystis carinii. Microbes Infect.* 2002. **4**(1): 95-103.

Morris, A.M., Swanson, M., Ha, H. and Huang, L., Geographic distribution of human immunodeficiency virus-associated *Pneumocystis carinii* pneumonia in San Francisco. *Am J Respir Crit Care Med.* 2000. **162**(5): 1622-6.

Mortier, E., Pouchot, J., Bossi, P. and Molinie, V., Maternal-fetal transmission of *Pneumocystis carinii* in human immunodeficiency virus infection. *N Engl J Med.* 1995. **332**(12): 825.

Nahimana, A., Francioli, P., Blanc, D.S., Bille, J., Wakefield, A.E. and Hauser, P.M., Determination of the copy number of the nuclear rDNA and beta-tubulin genes of *Pneumocystis carinii f. sp. hominis* using PCR multicompetitors. *J Eukaryot Microbiol.* 2000. **47**(4): 368-72.

Nakayama, K., Furin: a mammalian subtilisin/Kex2p-like endoprotease involved in processing of a wide variety of precursor proteins. *Biochem J.* 1997. **327**(Pt 3): 625-35.

Navin, T.R., Beard, C.B., Huang, L., del Rio, C., Lee, S., Pieniazek, N.J., Carter, J.L., Le, T., Hightower, A. and Rimland, D., Effect of mutations in *Pneumocystis carinii* dihydropteroate synthase gene on outcome of *P carinii* pneumonia in patients with HIV-1: a prospective study. *Lancet.* 2001. **358**(9281): 545-9.

Nei, M. and Gojobori, T., Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 1986. **3**(5): 418-26.

Nevez, G., Totet, A., Jounieaux, V., Schmit, J.L., Dei-Cas, E. and Raccurt, C., *Pneumocystis jiroveci* Internal Transcribed Spacer Types in Patients Colonized by the Fungus and in Patients with Pneumocystosis from the Same French Geographic Region. *J Clin Microbiol.* 2003. **41**(1): 181-186.

Nevez, G., Raccurt, C., Jounieaux, V., Dei-Cas, E. and Mazars, E., Pneumocystosis versus pulmonary *Pneumocystis carinii* colonization in HIV-negative and HIV-positive patients. *AIDS*. 1999a. **13**(4): 535-6.

Nevez, G., Raccurt, C., Vincent, P., Jounieaux, V. and Dei-Cas, E., Pulmonary colonization with *Pneumocystis carinii* in human immunodeficiency virus-negative patients: assessing risk with blood CD4+ T cell counts. *Clin Infect Dis*. 1999b. **29**(5): 1331-2.

Newbold, C.I., Antigenic variation in *Plasmodium falciparum*: mechanisms and consequences. *Curr Opin Microbiol*. 1999. **2**(4): 420-5.

Nimri, L.F., Moura, I.N., Huang, L., del Rio, C., Rimland, D., Duchin, J.S., Dotson, E.M. and Beard, C.B., Genetic diversity of *Pneumocystis carinii f. sp. hominis* based on variations in nucleotide sequences of internal transcribed spacers of rRNA genes. *J Clin Microbiol*. 2002. **40**(4): 1146-51.

Olsson, M., Sukura, A., Lindberg, L.A. and Linder, E., Detection of *Pneumocystis carinii* DNA by filtration of air. *Scand J Infect Dis*. 1996. **28**(3): 279-82.

Oz, H.S. and Hughes, W.T., Search for *Pneumocystis carinii* DNA in upper and lower respiratory tract of humans. *Diagn Microbiol Infect Dis*. 2000. **37**(3): 161-4.

Paabo, S., Irwin, D.M. and Wilson, A.C., DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem*. 1990. **265**(8): 4718-21.

Page, R. and Holmes, E., Molecular Evolution: A Phylogenetic Approach. *Blackwell Science Ltd*. 1998.

Page, R.D., TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*. 1996. **12**(4): 357-8.

Palladino, S., Kay, I., Fonte, R. and Flexman, J., Use of real-time PCR and the LightCycler system for the rapid detection of *Pneumocystis carinii* in respiratory specimens. *Diagn Microbiol Infect Dis*. 2001. **39**(4): 233-6.

Palmer, R.J. and Wakefield, A.E., Functional glycosylphosphatidylinositol anchor signal sequences in the *Pneumocystis carinii* PRT1 protease family. *Am J Respir Cell Mol Biol*. 2001. **25**(4): 466-73.

Palmer, R.J. Population Structure and the *PRT1* multigene family in rat-derived *Pneumocystis carinii*. 2000a. Oxford.

Palmer, R.J., Settnes, O.P., Lodal, J. and Wakefield, A.E., Population structure of rat-derived *Pneumocystis carinii* in Danish wild rats. *Appl Environ Microbiol*. 2000b. **66**(11): 4954-61.

Palmer, R.J., Cushion, M.T. and Wakefield, A.E., Discrimination of rat-derived *Pneumocystis carinii f. sp. Carinii and Pneumocystis carinii f. sp. Ratti* using the polymerase chain reaction. *Mol Cell Probes*. 1999. **13**(2): 147-55.

Peglow, S.L., Smulian, A.G., Linke, M.J., Pogue, C.L., Nurre, S., Crisler, J., Phair, J., Gold, J.W., Armstrong, D. and Walzer, P.D., Serologic responses to *Pneumocystis carinii* antigens in health and disease. *J Infect Dis*. 1990. **161**(2): 296-306.

Peters, S.E., Wakefield, A.E., Whitwell, K.E. and Hopkin, J.M., *Pneumocystis carinii* pneumonia in thoroughbred foals: identification of a genetically distinct organism by DNA amplification. *J Clin Microbiol*. 1994. **32**(1): 213-6.

Peters, S.E., Wakefield, A.E., Banerji, S. and Hopkin, J.M., Quantification of the detection of *Pneumocystis carinii* by DNA amplification. *Mol Cell Probes*. 1992a. **6**(2): 115-7.

Peters, S.E., Wakefield, A.E., Sinclair, K., Millard, P.R. and Hopkin, J.M., A search for *Pneumocystis carinii* in post-mortem lungs by DNA amplification. *J Pathol*. 1992b. **166**(2): 195-8.

Pifer, L.L., Hughes, W.T., Stagno, S. and Woods, D., *Pneumocystis carinii* infection: evidence for high prevalence in normal and immunosuppressed children. *Pediatrics*. 1978. **61**(1): 35-41.

Pixley, F.J., Wakefield, A.E., Banerji, S. and Hopkin, J.M., Mitochondrial gene sequences show fungal homology for *Pneumocystis carinii*. *Mol Microbiol*. 1991. **5**(6): 1347-51.

Pneumocystis Workshop, Revised nomenclature for *Pneumocytsis carinii*. *J Eukaryot Microbiol*. 1994. **41**(1): 121S-122S.

Redding, K., Holcomb, C. and Fuller, R.S., Immunolocalization of Kex2 protease identifies a putative late Golgi compartment in the yeast *Saccharomyces cerevisiae*. *J Cell Biol*. 1991. **113**(3): 527-38.

Rudenko, G., Genes involved in phenotypic and antigenic variation in African trypanosomes and malaria. *Curr Opin Microbiol*. 1999. **2**(6): 651-6.

Russian, D.A., Andrawis-Sorial, V., Goheen, M.P., Edman, J.C., Vogel, P., Turner, R.E., Klivington, D.L., Angus, C.W. and Kovacs, J.A., Characterization of a multicopy family of genes encoding a surface- expressed serine endoprotease in rat *Pneumocystis carinii*. *Proc Assoc Am Physicians*. 1999. **111**(4): 347-56.

Saitou, N. and Nei, M., The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol*. 1987. **4**(4): 406-25.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning, a Laboratory Manual, Cold Spring Harbor Laboratory Press.

Schaffzin, J.K. and Stringer, J.R., The major surface glycoprotein expression sites of two special forms of rat *Pneumocystis carinii* differ in structure. *J Infect Dis*. 2000. **181**(5): 1729-39.

Schaffzin, J.K., Sunkin, S.M. and Stringer, J.R., A new family of *Pneumocystis carinii* genes related to those encoding the major surface glycoprotein. *Curr Genet*. 1999. **35**(2): 134-43.

Scherf, A., Hernandez-Rivas, R., Buffet, P., Bottius, E., Benatar, C., Pouvelle, B., Gysin, J. and Lanzer, M., Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *Embo J.* 1998. **17**(18): 5418-26.

Schneider, M.M., Hoepelman, A.I., Eeftinck Schattenkerk, J.K., Nielsen, T.L., van der Graaf, Y., Frissen, J.P., van der Ende, I.M., Kolsters, A.F. and Borleffs, J.C., A controlled trial of aerosolized pentamidine or trimethoprim- sulfamethoxazole as primary prophylaxis against *Pneumocystis carinii* pneumonia in patients with human immunodeficiency virus infection. The Dutch AIDS Treatment Group. *N Engl J Med.* 1992. **327**(26): 1836-41.

Scott, G.B., Hutto, C., Makuch, R.W., Mastrucci, M.T., O'Connor, T., Mitchell, C.D., Trapido, E.J. and Parks, W.P., Survival in children with perinatally acquired human immunodeficiency virus type 1 infection. *N Engl J Med.* 1989. **321**(26): 1791-6.

Simonds, R.J., Oxtoby, M.J., Caldwell, M.B., Gwinn, M.L. and Rogers, M.F., *Pneumocystis carinii* pneumonia among US children with perinatally acquired HIV infection. *JAMA.* 1993. **270**(4): 470-3.

Sing, A., Roggenkamp, A., Autenrieth, I.B. and Heesemann, J., *Pneumocystis carinii* carriage in immunocompetent patients with primary pulmonary disorders as detected by single or nested PCR. *J Clin Microbiol.* 1999. **37**(10): 3409-10.

Sloand, E., Laughon, B., Armstrong, M., Bartlett, M.S., Blumenfeld, W., Cushion, M., Kalica, A., Kovacs, J.A., Martin, W., Pitt, E. and et al., The challenge of *Pneumocystis carinii* culture. *J Eukaryot Microbiol.* 1993. **40**(2): 188-95.

Smulian, A.G., *Pneumocystis carinii*: genetic diversity and cell biology. *Fungal Genet Biol.* 2001a. **34**(3): 145-54.

Smulian, A.G., Sesterhenn, T., Tanaka, R. and Cushion, M.T., The ste3 pheromone receptor gene of *Pneumocystis carinii* is surrounded by a cluster of signal transduction genes. *Genetics.* 2001b. **157**(3): 991-1002.

Soulez, B., Dei-Cas, E. and Camus, D., The rabbit, experimental host of *Pneumocystis carinii. Ann Parasitol Hum Comp.* 1988. **63**(1): 5-15.

Steiner, D.F., The proprotein convertases. *Curr Opin Chem Biol.* 1998. **2**(1): 31-9.

Stringer, J.R., Beard, C.B., Miller, R.F. and Wakefield, A.E., A New Name (*Pneumocystis jiroveci)* for Pneumocystis from Humans. *Emerg Infect Dis.* 2002. **8**(9): 891-6.

Stringer, J.R., Cushion, M.T. and Wakefield, A.E., New nomenclature for the Genus *Pneumocystis*. *J Eukaryot Microbiol.* 2001a. 184S-189S.

Stringer, J.R. and Keely, S.P., Genetics of surface antigen expression in *Pneumocystis carinii. Infect Immun.* 2001b. **69**(2): 627-39.

Stringer, J.R., *Pneumocystis carinii*: What is it, exactly? *Clin Microbiol Rev.* 1996a. **9**(4): 489-98.

Stringer, J.R. and Walzer, P.D., Molecular biology and epidemiology of *Pneumocystis carinii* infection in AIDS. *AIDS.* 1996b. **10**(6): 561-71.

Stringer, J.R., Stringer, S.L., Zhang, J., Baughman, R., Smulian, A.G. and Cushion, M.T., Molecular genetic distinction of *Pneumocystis carinii* from rats and humans. *J Eukaryot Microbiol.* 1993a. **40**(6): 733-41.

Stringer, J.R., The identity of *Pneumocystis carinii*: not a single protozoan, but a diverse group of exotic fungi. *Infect Agents Dis.* 1993b. **2**(3): 109-17.

Stringer, S.L., Stringer, J.R., Blase, M.A., Walzer, P.D. and Cushion, M.T., *Pneumocystis carinii*: sequence from ribosomal RNA implies a close relationship with fungi. *Exp Parasitol.* 1989. **68**(4): 450-61.

Sunkin, S.M., Linke, M.J., McCormack, F.X., Walzer, P.D. and Stringer, J.R., Identification of a putative precursor to the major surface glycoprotein of *Pneumocystis carinii. Infect Immun.* 1998. **66**(2): 741-6.

Sunkin, S.M. and Stringer, J.R., Residence at the expression site is necessary and sufficient for the transcription of surface antigen genes of *Pneumocystis carinii. Mol Microbiol.* 1997. **25**(1): 147-60.

Sunkin, S.M. and Stringer, J.R., Translocation of surface antigen genes to a unique telomeric expression site in *Pneumocystis carinii. Mol Microbiol.* 1996. **19**(2): 283-95.

Tanabe, K., Fuchimoto, M., Egawa, K. and Nakamura, Y., Use of *Pneumocystis carinii* genomic DNA clones for DNA hybridization analysis of infected human lungs. *J Infect Dis.* 1988. **157**(3): 593-6.

Tang, X., Bartlett, M.S., Smith, J.W., Lu, J.J. and Lee, C.H., Determination of copy number of rRNA genes in *Pneumocystis carinii f. sp. hominis. J Clin Microbiol.* 1998. **36**(9): 2491-4.

Theus, S.A. and Walzer, P.D., Adoptive transfer of specific lymphocyte populations sensitized to the major surface glycoprotein of *Pneumocystis carinii* decreases organism burden while increasing survival rate in the rat. *J Eukaryot Microbiol.* 1997. **44**(6): 23S-24S.

Theus, S.A., Linke, M.J., Andrews, R.P. and Walzer, P.D., Proliferative and cytokine responses to a major surface glycoprotein of *Pneumocystis carinii. Infect Immun.* 1993. **61**(11): 4703-9.

Tsolaki, A.G., Miller, R.F. and Wakefield, A.E., Oropharyngeal samples for genotyping and monitoring response to treatment in AIDS patients with *Pneumocystis carinii* pneumonia. *J Med Microbiol.* 1999. **48**(10): 897-905.

Tsolaki, A.G., Beckers, P. and Wakefield, A.E., Pre-AIDS era isolates of *Pneumocystis carinii f. sp. hominis*: high genotype similarity with contemporary isolates. *J Clin Microbiol.* 1998. **36**(1): 90-3.

Tsolaki, A.G., Miller, R.F., Underwood, A.P., Banerji, S. and Wakefield, A.E., Genetic diversity at the internal transcribed spacer regions of the rRNA operon among isolates of *Pneumocystis carinii* from AIDS patients with recurrent pneumonia. *J Infect Dis.* 1996. **174**(1): 141-56.

Underwood, A.P., Louis, E.J., Borts, R.H., Stringer, J.R. and Wakefield, A.E., *Pneumocystis carinii* telomere repeats are composed of TTAGGG and the subtelomeric sequence contains a gene encoding the major surface glycoprotein. *Mol Microbiol.* 1996. **19**(2): 273-81.

Vargas, S.L., Hughes, W.T., Santolaya, M.E., Ulloa, A.V., Ponce, C.A., Cabrera, C.E., Cumsille, F. and Gigliotti, F., Search for primary infection by *Pneumocystis carinii* in a cohort of normal, healthy infants. *Clin Infect Dis.* 2001. **32**(6): 855-61.

Vargas, S.L., Ponce, C.A., Gigliotti, F., Ulloa, A.V., Prieto, S., Munoz, M.P. and Hughes, W.T., Transmission of *Pneumocystis carinii* DNA from a patient with *P. carinii* pneumonia to immunocompetent contact health care workers. *J Clin Microbiol.* 2000. **38**(4): 1536-8.

Vargas, S.L., Ponce, C.A., Hughes, W.T., Wakefield, A.E., Weitz, J.C., Donoso, S., Ulloa, A.V., Madrid, P., Gould, S., Latorre, J.J., Avila, R., Benveniste, S., Gallo, M., Belletti, J. and Lopez, R., Association of primary *Pneumocystis carinii* infection and sudden infant death syndrome. *Clin Infect Dis.* 1999. **29**(6): 1489-93.

Vargas, S.L., Hughes, W.T., Wakefield, A.E. and Oz, H.S., Limited persistence in and subsequent elimination of *Pneumocystis carinii* from the lungs after *P. carinii* pneumonia. *J Infect Dis.* 1995. **172**(2): 506-10.

Wada, M. and Nakamura, Y., cDNA cloning and overexpression of cell surface subtilisin-like proteases (SSP) of *Pneumocystis carinii. J Eukaryot Microbiol.* 1997. **44**(6): 54S.

Wada, M., Sunkin, S.M., Stringer, J.R. and Nakamura, Y., Antigenic variation by positional control of major surface glycoprotein gene expression in *Pneumocystis carinii. J Infect Dis.* 1995. **171**(6): 1563-8.

Wada, M. and Nakamura, Y., MSG gene cluster encoding major cell surface glycoproteins of rat *Pneumocystis carinii. DNA Res.* 1994. **1**(4): 163-8.

Wada, M., Kitada, K., Saito, M., Egawa, K. and Nakamura, Y., cDNA sequence diversity and genomic clusters of major surface glycoprotein genes of *Pneumocystis carinii. J Infect Dis.* 1993. **168**(4): 979-85.

Wakefield, A.E., Lindley, A.R., Ambrose, H.E., Denis, C.M. and Miller, R.F., Limited Asymptomatic Carriage of *Pneumocystis. jiroveci* in Human Immunodeficiency Virus-Infected Patients. *J Infect Dis.* 2003. **187**: 901-8.

Wakefield, A.E., *Pneumocystis carinii. Br Med Bull.* 2002. **61**: 175-88.

Wakefield, A.E., Genetic heterogeneity in *Pneumocystis carinii*: an introduction. *FEMS Immunol Med Microbiol.* 1998. **22**(1-2): 5-13.

Wakefield, A.E., Keely, S.P., Stringer, J.R., Christensen, C.B., Ahrens, P., Peters, S.E., Bille-Hansen, V., Henriksen, S.A., Jorsal, S.E. and Settnes, O.P., Identification of porcine *Pneumocystis carinii* as a genetically distinct organism by DNA amplification. *APMIS.* 1997. **105**(4): 317-21.

Wakefield, A.E., DNA sequences identical to *Pneumocystis carinii f. sp. carinii* and *Pneumocystis carinii f. sp. hominis* in samples of air spora. *J Clin Microbiol.* 1996. **34**(7): 1754-9.

Wakefield, A.E., Fritscher, C.C., Malin, A.S., Gwanzura, L., Hughes, W.T. and Miller, R.F., Genetic diversity in human-derived *Pneumocystis carinii* isolates from four geographical locations shown by analysis of mitochondrial rRNA gene sequences. *J Clin Microbiol.* 1994. **32**(12): 2959-61.

Wakefield, A.E., Miller, R.F., Guiver, L.A. and Hopkin, J.M., Oropharyngeal samples for detection of *Pneumocystis carinii* by DNA amplification. *Q J Med.* 1993. **86**(6): 401-6.

Wakefield, A.E., Guiver, L., Miller, R.F. and Hopkin, J.M., DNA amplification on induced sputum samples for diagnosis of *Pneumocystis carinii* pneumonia. *Lancet.* 1991. **337**(8754): 1378-9.

Wakefield, A.E., Pixley, F.J., Banerji, S., Sinclair, K., Miller, R.F., Moxon, E.R. and Hopkin, J.M., Detection of *Pneumocystis carinii* with DNA amplification. *Lancet.* 1990. **336**(8713): 451-3.

Walzer, P.D., Ashbaugh, A., Collins, M. and Cushion, M.T., Anti-human immunodeficiency virus drugs are ineffective against *Pneumocystis carinii* in vitro and in vivo. *J Infect Dis.* 2001. **184**(10): 1355-7.

Walzer, P.D. and Linke, M.J., A comparison of the antigenic characteristics of rat and human *Pneumocystis carinii* by immunoblotting. *J Immunol.* 1987. **138**(7): 2257-65.

Walzer, P.D., Experimental models of *Pneumocystis carinii* infections. L.S. Young (ed.), *Pneumocystis carinii* pneumonia, *Marcel Dekker, New York, N.Y.* 1984. (1st ed): 7-76.

Williams, A.J., Duong, T., McNally, L.M., Tookey, P.A., Masters, J., Miller, R., Lyall, E.G. and Gibb, D.M., *Pneumocystis carinii* pneumonia and cytomegalovirus infection in children with vertically acquired HIV infection. *AIDS.* 2001. **15**(3): 335-9.

Yale, S.H. and Limper, A.H., *Pneumocystis carinii* pneumonia in patients without acquired immunodeficiency syndrome: associated illness and prior corticosteroid therapy. *Mayo Clin Proc.* 1996. **71**(1): 5-13.

Ypma-Wong, M.F., Fonzi, W.A. and Sypherd, P.S., Fungus-specific translation elongation factor 3 gene present in *Pneumocystis carinii. Infect Immun.* 1992. **60**(10): 4140-5.

Zhu, X., Lamango, N.S. and Lindberg, I., Involvement of a polyproline helix-like structure in the interaction of 7B2 with prohormone convertase 2. *J Biol Chem.* 1996. **271**(38): 23582-7.